

The impact of epidemiological parameters on the phylogeny of pathogen strains

Isao Kawaguchi

Department of Biology, Faculty of Science, Kyushu University, Japan

Goal

To develop a deeper understanding of how the polygenetic properties of disease strains are affected by underlying epidemiological detail.

Background and motivation

Currently, an enormous number of nucleotide sequences become available for viruses and infectious bacteria that cause major epidemiological outbreaks in humans. The phylogenetic relationships found *within* a species or a quasispecies of such pathogens vary greatly from pathogen to pathogen and might reflect their epidemiological and microevolutionary dynamics:

- The measles viruses were highly homogeneous for a long time and most subtypes have diverged only recently.
- For influenza A, new serotypes are continuously derived each year; yet, the phylogenetic tree reconstructed for the variants sampled over a few decades collapses to a single trunk (Fitch et al. 1997).
- The dengue virus quasispecies consists of four distantly related subtypes, which diverged a long time ago and have coexisted ever since (Zannoto et al. 1996, Gubler and Kuno 1997).

Little is known, at this stage, about how these striking differences in tree shapes are related to the underlying epidemiological parameters – like transmission rate, virulence, recovery rate – and to the populations' genetic setup.

Methods

I recently started to investigate these questions in collaboration with Dr. Akira Sasaki, Kyushu University. We focused on pathogen strains defined by epitope sequences of length n , and considered the competition between pathogen strains through the epidemiological dynamics with the host cross-immunity (Kawaguchi and Sasaki, manuscript). For this purpose, we have adopted an individual-based simulation model, rather than a traditional population-based SIR model, because in SIR models the number of dynamical variables (the number of immunological states of hosts) increases dramatically with the length of epitope sequences.

We consider a finite number of N host individuals and of $m=2^n$ viral serotypes, defined by binary epitope sequences of length n . The immunological status of a host with regard to each strain of the pathogen can change from S (susceptible) to I (infected and infectious), and from I to R (recovered and immune). With m serotypes of the pathogen, there are therefore 3^m different immunological host states. We also take into account the coinfection by multiple strains and the cross-immunity between pathogen strains. We allow for mutation at each site of the epitope sequence, and hence keep track of the microevolutionary change of the pathogen serotypes. We assume that the degree of cross-immunity mounted by hosts against a newly infecting strain is determined by the Hamming distance between the epitope sequence of this and the previously infecting strain.

The immune status of host individual i can be described by the vector \mathbf{s}_i which has m columns, each indicating the immune status relative to a pathogen strain. Each element in \mathbf{s}_i denotes either of three states, susceptible (0), infected (1), or recovered (2). Thus, the immune status of the individual i is specified by the vector $\mathbf{s}_i \in \{0,1,2\}^m$. The j^{th} element of that vector, \mathbf{s}_{ij} , is 0 if the individual has never been infected by viral strain j ; $\mathbf{s}_{ij} = 1$ if the individual is currently infected by the strain; and $\mathbf{s}_{ij} = 2$ if the individual is immune to the strain. The epitope sequence of strain j is denoted by a binary sequence x_j .

For viral strain j with serotype x_j infecting a host with immune status vector \mathbf{s}_i , the probability that the strain can escape the immune response acquired by a previous infection by strain l with the serotype x_l is assumed to be given by

$$\phi(x_j, x_l) = 1 - \exp[-d(x_j, x_l)^2 / 2\sigma^2] , \quad (1)$$

where $d(x_j, x_l)$ is the Hamming distance between serotypes x_j and x_l ; σ gives the characteristic distance in serotype space over which the immune response mounted by a strain is effective. This infection probability is zero when two serotypes perfectly match and approaches one as the distance between them diverges to infinity. The infection probability of a new strain should be affected by all previously infecting strains and we therefore assume that the overall effect of previous infections is given by multiplying (1) for all previously infecting strains. Thus the probability that a pathogen serotype x can evade the immune response in host i (and hence can infect the host) is

$$\eta_i(x) = \prod_{\substack{j=1 \\ \mathbf{s}_{ij}=2}}^n \phi(x, x_j) , \quad (2)$$

where the product is taken for all j with $\mathbf{s}_{ij}=2$. The transition probability that the individual i becomes infected by viral strain j is then

$$P(\mathbf{s}_{ij} = 0 \rightarrow \mathbf{s}_{ij} = 1) = \beta u_j \eta_i(x_j) , \quad (3)$$

where $u_j = \sum_j \mathbf{1}(\mathbf{s}_{ij} = 1)$ is the total density of viral strain j in the population and β is the transmission rate. The transition probability that individual i currently infected by strain j is recovered and immune to the serotype is

$$P(\mathbf{s}_{ij} = 1 \rightarrow \mathbf{s}_{ij} = 2) = \gamma_{ij} , \quad (4)$$

where γ_{ij} is the recovery rate of host i from an infection by strain j . A healthy individual dies with natural mortality d , while an infected individual in general has a higher mortality that depends on the number of current infections,

$$\text{mortality of individual } i = d + v_i \alpha_0 , \quad (5)$$

where $v_i = \sum_j \mathbf{1}(\mathbf{s}_{ij} = 1)$ is the number of currently infecting strains in the individual i , and α_0 is the virulence of the pathogen.

From this model, we generated phylogenetic trees and identified statistics that characterize the shape of these trees as well as other phylogenetic relationships. We then analyzed how these statistics are related to the epidemiological parameters and genetic setup of the studied population.

Research questions and work plan

During the summer project, I plan to extend the basic model described above in two directions. Both extensions are intended to make the basic model more realistic.

- The first direction is to allow for variation in the virulence of strains. So far, the virulence α_0 is assumed to be constant, whereas the virulence of natural diseases can evolve over relatively short periods of time t (e.g. Fenner 1983; May and Anderson 1983). I plan to first analyze the model by assuming that the virulence is determined by the epitope sequence itself. In an alternative approach, I will then assume the virulence to be determined independently of the epitope sequence and will examine the joint evolution of virulence and epitope sequence (assuming partial linkage between them). As before, I will generate phylogenetic trees and epidemiological time series through simulation of the model and will analyze their relationship with the underlying epidemiological parameters.
- A second direction of extension is to introduce demographic change of total host population size, rather than assuming a constant host population size.
- In addition, I am planning to investigate the potential for developing analytical treatments of this system in collaboration with ADN Project Coordinator Dr. Ulf Dieckmann, who supervises my YSSP project.

Relevance and link to ADN's research plan

This study ties in with ADN's research focus on disease evolution and virulence management. It focuses on the relationship between phylogenetic tree shape and epidemiological parameters. If we succeed in finding interesting tendencies in this relationship, this would open up possibilities for inferring epidemiological parameters from observed phylogenetic trees. The model will also allow for simulating how viruses evolve when host populations are treated with antibiotics or is vaccinated.

Envisaged publication

In collaboration with Ulf Dieckmann, a paper on the theoretical aspects of the developed model is planned to be submitted at the end of the YSSP period.

References

- Fenner, F. 1983, Biological control, as exemplified by smallpox eradication and myxomatosis. *Proc. R. Soc. Lond.* **B218**, 259-285.
- Fitch W.M., Bush, B.M., Bender C.A., and Cox, N.J. 1997., Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc. Natl. Acad. Sci. USA.* **94**: 7712-7718.
- Gubler D.J. and Kuno G., (eds) 1997, Dengue and Dengue Hemorrhagic Fever. CAB International.
- May R.M. and Anderson, R.M. 1983, Epidemiology and genetics in the coevolution of parasites and hosts. *Proc. R. Soc. Lond.* **B219**, 281-313
- Zanotto, P. M. d. A., Gould, E. A., Gao, G. F. & Harvey, P. H. 1996., *Proc. Natl. Acad. Sci. USA* **93**, 548-553.