Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Viral evolution: beyond drift and shift

Benjamin D Greenbaum[1] and Elodie Ghedin[2]

Technological advances have allowed aspects of viral evolution to be explored at unprecedented scales. As a consequence, new quantitative approaches are needed to investigate features of viral evolution that fall outside traditional areas of study, such as antigenic evolution. We examine three areas of viral evolution where tools from disciplines such as statistical physics, topology, and information theory have been used recently as quantitative frameworks for large-scale studies and, in some cases, suggest a novel theoretical approach to a problem. Ongoing interaction among these disciplines with biology is necessary so that experimental researchers can determine which quantitative tools are right for them and quantitative researchers can learn which aspects of viral evolution can be understood and advanced with their approaches.

Addresses
[1] Tisch Cancer Institute, Departments of Medicine and Pathology, 1190 5th Ave, New York, NY 10029, United States
[2] Center for Genomics & Systems Biology, Department of Biology, and Global Institute of Public Health, New York University, 100 Washington Place, 1009 Silver Center, New York, NY 10003, United States

Corresponding author: Greenbaum, Benjamin D (benjamin.greenbaum@mssm.edu)

Current Opinion in Microbiology 2015, **26**:109–115

This review comes from a themed issue on **Host-microbe interactions: viruses**

Edited by **Ivan Marazzi** and **Adolfo Garcia-Sastre**

For a complete overview see the Issue and the Editorial

Available online 17th July 2015

http://dx.doi.org/10.1016/j.mib.2015.06.015

Introduction

High mutation rates of RNA viruses, caused by an error-prone RNA-dependent RNA polymerase, make them a veritable goldmine for researchers interested in observing evolutionary novelty and developing new approaches to study evolution in action. The last several years have seen a growth of methods for observing and analyzing viral evolution, driven by datasets at a scale that would have been unimaginable a decade ago. The combination of large-scale sequencing efforts, spatiotemporal surveillance, and detailed characterization of virus interactions with host mechanisms, both from patient data and *in vitro*, have contributed to rich, varied datasets, yielding a correspondingly prodigious opportunity to test quantitative hypotheses about virus evolution and use that information to learn about host biology.

In the wake of these rich datasets, a wide range of methods have emerged to make sense of the information they contain and to quantify aspects of viral evolution that previously were difficult to address directly. Often, these methods extend beyond the confines of traditional statistical analyses or population genetics-based approaches, taking inspiration from branches of mathematics, theoretical physics, and/or computer science. Consequently, a biologist assessing quantitative approaches whose origins span so many disciplines might have trouble determining whether a given method was applicable to his or her own work. Recently, such interdisciplinary methods have been implemented to infer information about virus evolution from large-sequence datasets, particularly focused on influenza and HIV. In the case of influenza, the first aspects of viral evolution to be probed were for variation in viral epitopes from mutation and reassortment — also called antigenic drift and shift — and several new tools have been directed towards their study. Here, however, we have chosen to highlight some examples that go beyond studying drift and shift, to emphasize aspects of viral evolution that we see as needing further quantification.

Virus diversity and fitness landscapes

RNA virus populations within infected hosts are genetically diverse, comprising a mix of *de novo* mutations and distinct variants. This diversity is important as it allows the virus to adapt to changing environments and overcome multiple selective pressures, but it is difficult to capture and quantify accurately. Assessing the composition of a viral community, including major and minor members, and characterizing within-host diversity has been greatly facilitated by next-generation sequencing platforms and by novel laboratory methods. Most of what we know to date about RNA virus evolution within hosts has come from studies of chronic infections such as HCV and HIV [1,2]. Now, deep sequencing is giving us a detailed look beyond the consensus sequence to characterize intra-host population dynamics of acute RNA viruses as well [3••,4•].

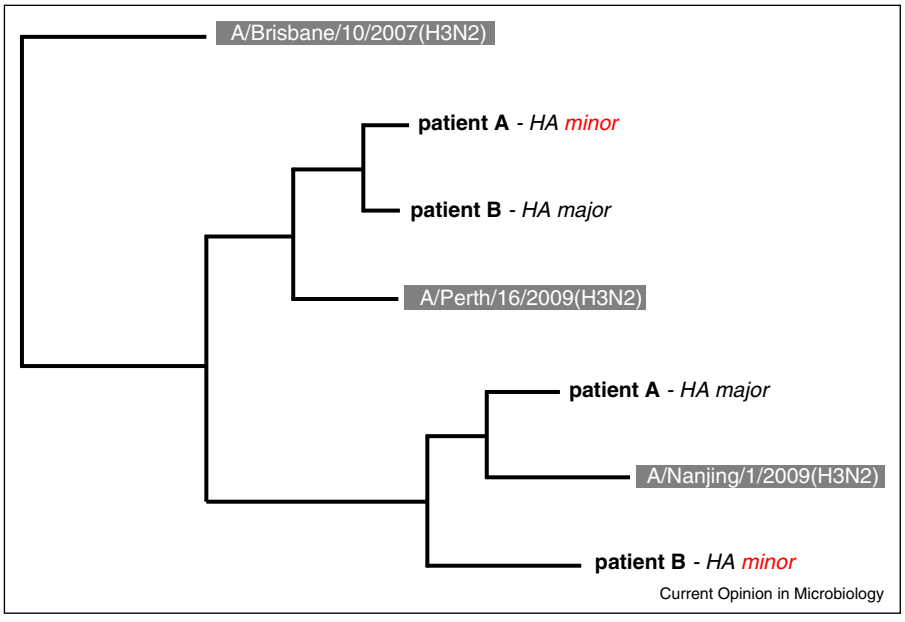
These next-generation sequencing platforms are unquestionably powerful in the data they produce, yet they contain intrinsic errors that often exceed the RNA virus mutation rate, limiting a true characterization of the within-host evolutionary rate. Stringent analytical methods are needed to evaluate this rate accurately. A recent study proposes an auto-correction method to overcome

some of the random errors introduced by amplification and sequencing [3••,5]. In this method genomic RNA fragments are circularized and reverse-transcribed multiple times by a 'rolling circle,' generating a cDNA molecule with physically linked tandem repeats. This built-in redundancy confers strong confidence in mutations identified by deep sequencing when they are present on all copies of the repeats. The power of this approach to capture very low frequencies of mutations (two orders of magnitude lower than the average mutation frequency of many RNA viruses), provided evidence for directionality of evolution in the poliovirus. More specifically, it showed: (i) inequality in the types of transitions that occurred with the ability to move directionally in sequence space, and (ii) that synonymous mutations could be under strong selection with an impact on fitness. This method, developed using cultured viruses, has however important limitations when dealing with low levels of genomic viral RNA seen in clinical samples. It is also a difficult method to apply because of the low efficiency of circularization. Thus, for most studies error correction will have to be implemented at the data analysis step.

Viruses are often described as a quasi-species moving through fitness landscapes where specific members of the viral community will be favored depending on the selective pressure they encounter [6]. Approaches such as evolutionary antigenic cartography quantify classical measures of antigenic change [7,8]. Antigenic evolution has itself recently undergone extensive novel quantitative analysis, with the goal to improve statistical vaccine prediction [9•,10]. But when focusing on genetic variants that are not in putative antigenic sites, assessing quantitatively both the impact of the changes on virus community composition and on virus fitness for transmission will require other approaches. Moreover, the heterogeneous and highly time-dependent nature of the host response can lead to a breakdown in fitness landscape approaches which assume that the topography of such landscapes are static or slowly varying relative to the evolution of a virus. In the case of time-dependent 'seascapes' the dynamics between the host and viral community will be non-linear and more difficult to quantify [11].

How can we quantify the fitness of each member of the community, or determine overall fitness of the population as a whole in such a dynamic host environment? We have seen in our own studies on influenza virus how minor variants can persist within infected hosts and across chains of transmission [12]. We also observe, particularly in patients carrying two different antigenic strains of the virus, that when the minority strains do become dominant in the population they often simply push the major variant down to a minority status; they do not necessarily eliminate it (Figure 1). This phenomenon is reminiscent of the 'mutational coupling' among variants one would expect in a quasispecies population [13] where selection favors a mix of genotypes because the ensemble has a replicative advantage [14]. The concept of quasispecies in this case includes viral strains that co-circulate in the population. Defective RNA virus genomes — or sub-genomic RNA — have been shown to be transmitted in

Figure 1



Neighbor-joining tree of reconstructed haplotypes found in two patients located in the same city and infected with H3N2 in 2009. Each patient is infected with two different antigenic strains (A/Nanjing/1/2009-like and A/Perth/16/2009-like), with the minor variant at approximately 20% frequency, based on deep-sequence data (data not shown).