Cell PRESS

Capturing escape in infectious disease dynamics

Sarah Cobey[1] and Katia Koelle[2]

[1] Department of Ecology and Evolutionary Biology, University of Michigan, 830 North University Avenue, Ann Arbor, MI 48109, USA
[2] Department of Biology, Duke University, Box 90338, Durham, NC 27708, USA

Identifying the causes of interannual variability in disease dynamics is important for understanding and managing epidemics. Traditionally, these causes have been classified as intrinsic (e.g. immunity fluctuations) or extrinsic (e.g. climate forcing); ecologists determine the relative contributions of these factors by applying statistical models to time series of cases. Here we address the problem of isolating the drivers of pathogen dynamics that are influenced by antigenic evolution. Recent findings indicate that many pathogens escape immunity in a punctuated manner; for them, we argue that time series of cases alone will be insufficient to isolate causal drivers. We detail observations that can reveal the presence of punctuated immune escape, and which can be used in new statistical approaches to identify extrinsic and intrinsic regulators of disease.

Drivers of disease

Many of the most prevalent infectious diseases are caused by RNA viruses that can escape immunity through rapid antigenic evolution. Their evolution allows them to reinfect previously immune individuals (e.g. influenza, norovirus) or to persist as chronic infections (e.g. HIV, hepatitis C virus). Control of these pathogens is currently hindered by our limited understanding of how these pathogens evolve, how their evolutionary and epidemiological dynamics interact [1] and how other (extrinsic) factors influence the size of their outbreaks.

Here we first provide a brief overview of the statistical approaches that have been used to understand the factors regulating the disease dynamics of antigenically stable pathogens and describe why, except in a very limited number of circumstances, we expect these approaches to fail when applied to pathogens that rapidly evolve to escape immunity. We argue that for these pathogens, successfully capturing their intrinsic dynamics will require determining the appropriate phenotypic resolution and accurately identifying the timing of phenotypic change. Recent observations have shown that punctuated immune escape can arise from *de novo* mutations, recombination and reassortment in RNA viruses. We use them as an example to outline several observable patterns that allow us to detect this type of evolutionary dynamic. Theoretical principles underlying these patterns will need to be incorporated into the next generation of statistical models to better grasp the dynamics of rapidly evolving pathogens.

Statistical models for infectious diseases

Many statistical models for infectious diseases have been developed to address one of the oldest questions in ecology: the roles of intrinsic (density-dependent) versus extrinsic (density-independent) factors in regulating population dynamics. For host–pathogen systems, the dominant intrinsic factor is immunity generated by previous exposure to the pathogen. Extrinsic factors include changes in contact patterns, birth rates, vaccination policy and climatically driven transmission rates [2,3].

Statistical approaches attempting to distinguish intrinsic from extrinsic factors usually first postulate a model for intrinsic dynamics and then determine whether incorporat-

---

Glossary

dN/dS: The ratio of nonsynonymous substitution rates (dN) to synonymous substitution rates (dS). Purifying selection manifests as dN/dS < 1, neutral evolution as dN/dS = 1 and positive selection as dN/dS > 1.

Epistasis: The dependence of the phenotypic effects of a genetic element on the presence or absence of other genetic elements. Most commonly, epistasis refers to interactions between genes, but it can also refer to interactions among individual amino acids.

Epitope: An antigenic site on a pathogen. Epitopes can be located on surface proteins, where they might be targeted by antibodies. They can also be in the primary structure of proteins that are presented by infected cells for elimination by T cells.

Epochal evolution: Evolution that appears governed by periods of phenotypic stasis and sudden innovations. Mutations accumulating during periods of phenotypic stasis facilitate the process of innovation by changing the genetic background in which future mutations occur.

Hamming distance: The Hamming distance between two sequences is the number of positions by which they differ.

Hemagglutination inhibition (HI) assay: A measure of the ability of antibodies to bind to viruses, such as influenza, that can agglutinate red blood cells. The better an antibody matches a particular virus, the more effectively it prevents hemagglutination.

Lagged incidence: Incidence values at previous time steps from which projections are made about current incidence. For example, a lag of two time steps corresponds to I_{t-2}.

Mechanistic models: Models that explain observations by postulating components and interactions that are hypothesized to be at play.

Neutral mutations: Mutations that have no effect on fitness. They can nonetheless result in amino acid substitutions and affect phenotype.

Neutralization assay: A measure of the ability of antibodies to block viral infection of host cells.

Non-mechanistic models: Models that explain observations phenomenologically by postulating functional forms to describe relationships among the data. Examples include autoregressive models and neural network models.

Punctuated immune escape: Immune escape that occurs abruptly as a result of antigenic change in the pathogen. Antigenic change can arise from a number of different genetic processes, including *de novo* mutations, recombination, reassortment and allele switching.

Selective sweep: The rise to fixation of one genotype in a population.

SIR(S): A class of epidemiological models in which hosts belong to one of three classes, compartments or states: susceptible (S), infected (I) or recovered (R). In the SIR model, immunity is permanent. In the SIRS model, hosts lose their immunity and transition from the recovered to the susceptible class.
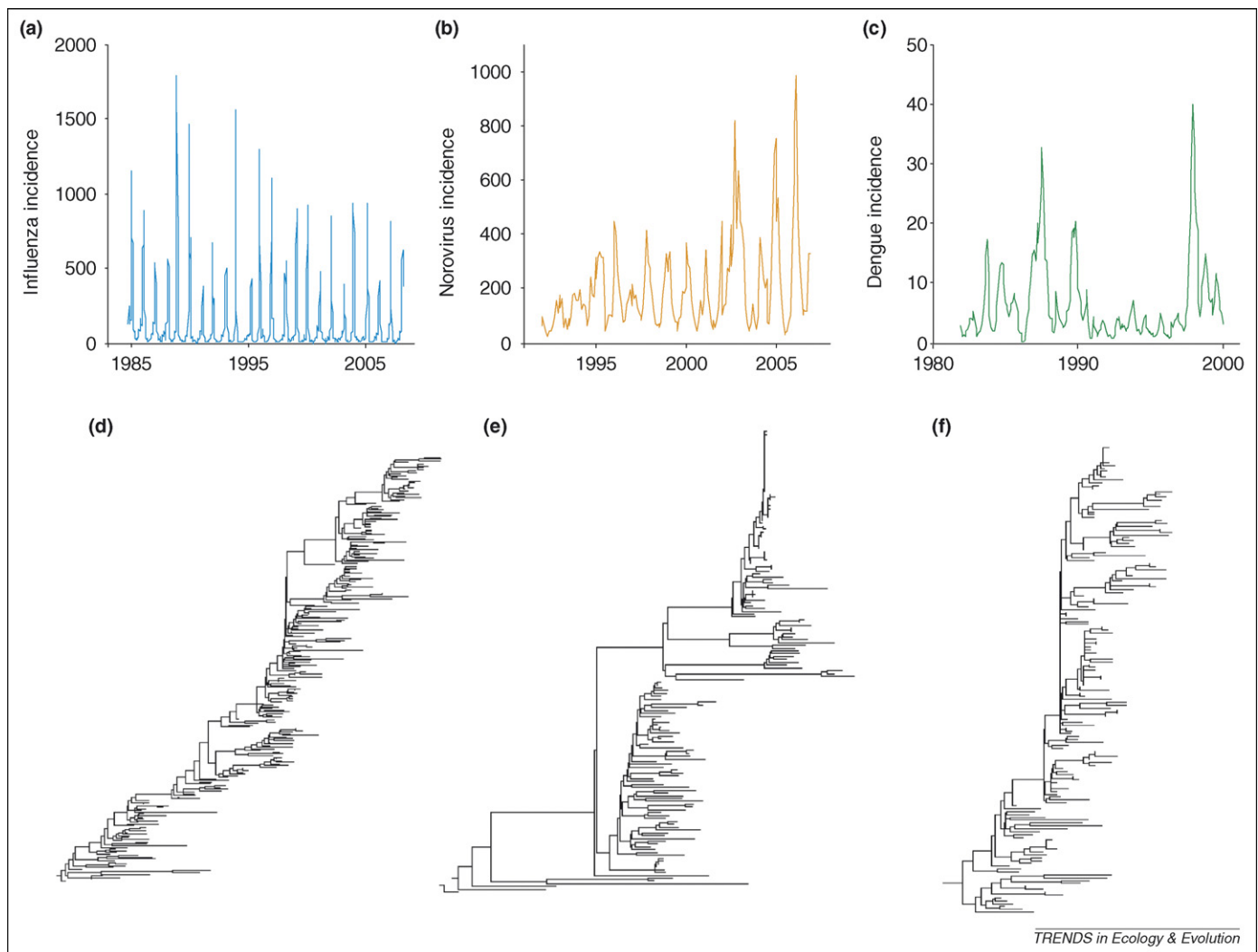
Figure 1. Ecological and evolutionary dynamics of three RNA viruses. **(a–c)** Disease dynamics of several rapidly evolving RNA viruses that show interannual variability in incidence. (a) Influenza dynamics from France over 1984–2008, reported weekly, per 100 000 (from http://www.sentiweb.org). (b) Norovirus dynamics from the United States, over 1992–2006, reported monthly (from http://www.hpa.org.uk/infections/topics_az/norovirus/data_ew_month.htm). (c) Dengue hemorrhagic fever dynamics from Bangkok, Thailand over 1981–1999, reported monthly, per 100 000 (from the Thai Ministry of Public Health). Both intrinsic and extrinsic factors have been hypothesized to play a role in generating observed patterns of interannual variability in all three of the examples shown. **(d–f)** Phylogenies of viral pathogens. (d) Phylogeny of influenza A's (H3N2) hemagglutinin glycoprotein, inferred from sequences used in Ref. [51]. Included sequences were isolated over the time period 1980–2000. (e) Phylogeny of norovirus's capsid protein, inferred from sequences used in Ref. [52]. Included sequences were isolated over the time period 1987–2006. (f) Phylogeny of dengue virus's envelope glycoprotein, inferred from sequences used in Refs [53–56]. Included sequences belonged to the Asian I genotype of serotype 2 and were isolated over the time period 1958–2001. Viral phylogenies illustrate the ladderlike topology discussed in the main text, which can result from repeated occurrences of punctuated immune escape.

ing extrinsic variability significantly improves the fit of the model to incidence data [4]. Non-mechanistic, mechanistic and semi-mechanistic statistical models alike follow this approach. Non-mechanistic models use lagged incidence to capture the intrinsic state of the system phenomenologically (e.g. [5,6]). Mechanistic models explicitly describe the hypothesized dynamical interactions, usually by separating the host population into susceptible (S), infected (I) and recovered (R) states and defining rates of transitioning between these states. Temporal variation in the number of susceptible hosts is then reconstructed from time series of infected individuals (e.g. [7,8]). In both of these types of models, additional (unexplained) changes in incidence can be attributed to extrinsic factors. Semi-mechanistic models directly combine aspects of non-mechanistic models with mechanistic ones, frequently to allow for additional variation in a parameter; this variation can be ascribed to either intrinsic [9] or extrinsic [8,10] factors.

These types of models have provided convincing evidence for the role of birth rates in generating annual versus biennial epidemics of measles [11] and the El Niño-Southern Oscillation (ENSO) in driving outbreaks of cholera [6,10]. However, these success stories share a common protagonist: a pathogen that does not rapidly evolve with respect to its antigenic phenotype.

Like their antigenically stable brethren, many rapidly evolving RNA viruses show dynamics that vary interannually (Figure 1a–c). Unfortunately, statistical approaches for antigenically stable pathogens are not easily extended to rapidly evolving pathogens. These approaches predict disease incidence using time series of infected individuals, whose strain phenotypes are implicitly assumed to be identical. In the case of rapidly evolving pathogens, cocirculating strains can vary antigenically, with each strain 'seeing' different levels of immunity in the population. A complete description of this system requires specifying, at

minimum, the number of hosts susceptible, immune and recovered to each strain, thus greatly increasing the number of states, or dimensionality, necessary to describe the intrinsic dynamics. A single time series of infected individuals aggregates over these phenotypic variants: it is a low-dimensional projection of a much more complicated system.

Under a very restricted set of evolutionary assumptions, a low-dimensional time-series projection might be appropriate. For example, if the relationship between the number of amino acid substitutions and the probability of reinfection is linear, if only one antigenic variant circulates at any one time and if amino acid fixations result from random drift (instead of selective processes), then immunity to reinfection can be modeled as a function of time since last infection [12]. In this case, the dynamics can be described by a low-dimensional susceptible-infected-recovered-(re)susceptible (SIRS) system, in which recovered individuals lose their immunity and become susceptible again from rapid evolution of the pathogen. Statistical methods developed for antigenically stable pathogens (which rely on a single time series of infected individuals) can in this case be applied (e.g. [9,13]).

The evolutionary assumptions required to justify the use of SIRS models are particularly worrisome in light of accumulating theory. Both observations and models show that antigenic variants can arise rapidly from standing genetic variation, possibly through recombination of pre-existing alleles [14,15] or by reassortment between genes [16]. A second mechanism of antigenic evolution, not exclusive of the first and well documented at the within-host level, is variable expression of genes for surface proteins (e.g. in infections by *Trypanosoma brucei*, which causes sleeping sickness, and *Borrelia* spp., including Lyme disease [17]). Finally, several recent observations have shown that many viral pathogens escape immunity through *de novo* point mutations that have variable effects on their antigenic phenotypes. For these pathogens, genetic changes can occur without discernable effect on antigenic phenotype, and large changes in antigenicity can result suddenly from small events (e.g. a single amino acid replacement). Many of these pathogens therefore appear to evolve in a phenotypically punctuated manner.

Punctuated immune escape has been most clearly documented in flu [18] and norovirus [19,20]. It has also been suggested for HIV within hosts [21], and clade replacements in dengue also point toward punctuated immune escape [22–24]. Although the mechanisms behind the punctuated emergence of antigenically novel variants have not been fully elucidated, one explanation lies in what has been termed 'epochal evolution' [25]. Under epochal evolution of viral pathogens [26], neutral mutations accumulate and alter the genetic background of antigenic proteins. Through epistasis, a single amino acid substitution is then able to substantially alter epitope conformation, effectively reducing preexisting immunity at the level of the population.

How we specify the intrinsic dynamics of rapidly evolving pathogens will lead to qualitatively different inferences about the factors driving their patterns of incidence. We outline here a set of observations that comprise general tests for punctuated immune escape.

Detection of punctuated immune escape: host patterns

One signal that an antigenically novel immune escape variant has emerged is an increase in the number of infected individuals. This pattern has been observed in influenza [27] and norovirus [19] during periods of antigenic change. However, an increase in cases is an unreliable indicator, because it can result from other processes, including variability in extrinsic factors (e.g. temporary increases in transmission rates) and intrinsic factors unrelated to immune escape (e.g. a threshold of susceptible hosts that has been reached via a replenishment by births).

A second suggestive pattern is a temporary increase in the mean age of infected individuals. Because an immunity-conferring disease primarily infects younger cohorts, and because the appearance of a novel variant causes some older individuals to regain susceptibility, the average age of infected individuals is expected to rise when an antigenic variant invades. Importantly, the increase in mean age is only transient: as the new variant becomes endemic, the average age of infection declines to an equilibrium value. However, changes in the average age of infection can be confounded by age-related differences in contact rates.

Detection of punctuated immune escape: pathogen patterns

Epidemiological processes, including the duration and strength of cross-immunity between strains, drive evolutionary dynamics and shape viral diversity [1]. Accordingly, observations of the viral population can be used to infer patterns of competition among strains. We outline below antigenic and genetic signatures that are characteristic of punctuated immune escape.

Antigenic phenotypes

The most convincing evidence of punctuated immune escape comes from studies that demonstrate discrete changes in antigenic phenotype. These include serological studies (e.g. hemagglutination inhibition assays of influenza [18] and neutralization assays of poliovirus's *VP1* protein [28] and HIV's *gp120* protein [29]) and structural studies of relevant epitopes (e.g. [20]). Structural studies are slightly less powerful than serological studies, in that they traditionally do not examine the effects of substitutions on the binding affinity to antibodies [30] and frequently assume that the correct locations of epitopes are known.

Because new variants of a pathogen have a frequency-dependent selective advantage, and will thus grow rapidly in the host population, strong evidence for punctuated immune escape comes from observing how the frequencies of the antigenic phenotypes change over time. Rates of increase must be shown to be higher than the neutral, stochastic expectation [31].

Evolutionary rates

Evolutionary rates inferred from genetic sequences provide two kinds of evidence for punctuated immune escape. First, comparisons of nonsynonymous to synonymous substitution rates (dN and dS) can suggest when the pathogen might be evolving owing to positive selection, and where in the genome the changes might be occurring. The clearest illustration of such episodic selection comes from the

hemagglutinin protein of influenza, where dN/dS measurements have shown strong positive selection over long time periods [32] and either purifying selection or neutral evolution over short time periods [33]. This result is consistent with the repeated emergence and replacement dynamics of major antigenic variants, or clusters, found in subtype H3N2 [18]. Interestingly, similar patterns have been observed within hosts for the *env* protein of HIV [34].

Although informative, this approach faces several challenges. First, it is difficult to determine whether observed substitution patterns are related to changes in antigenicity, as opposed to other components of fitness. Second, because multiple substitutions are included in dN/dS, the individual effects of specific substitutions cannot be pried apart. A third challenge stems from the observation that current tests for temporal variation in dN/dS rely on some kind of directed search. A common method of detecting episodic selection using sequences and phylogenies, the branch-site test of positive selection [35], requires specifying *a priori* which part of the gene is under positive selection and on which branch or branches of the phylogeny selection might be occurring. This approach compares the likelihood of a model in which the dN/dS ratio in an area possibly undergoing positive selection is constrained to one, corresponding to neutral evolution, to a model in which dN/dS is fitted. If the fitted value is greater than one, and there is a significant difference in likelihoods, positive selection might be occurring. The low divergence in many RNA virus phylogenies can require that branches and partitions be aggregated to gain statistical power, sacrificing resolution. To demonstrate that the driving evolutionary force is punctuated immune escape, one faces the cumbersome task of showing that positive selection is significantly weaker or absent on all other branches and parts of the genome.

A second possible approach is to use classic predictions in population genetics to infer the presence of neutral [36] and episodic [37] selection from substitution rates. Neutral evolution can be modeled as a Poisson or overdispersed process, and periods of episodic selection should be accompanied by bursts of substitution events, as has been observed for HIV [38]. One major challenge in applying these theories will be to control for disease dynamics, especially severe bottlenecks in pathogen populations, and to account for changes in selective constraints [39]. As in the case of dN/dS tests, another challenge of this approach is to determine whether the observed substitution patterns are related to changes in antigenicity *per se*.

Genetic diversity

The occurrence of punctuated immune escape can also be gauged from the genetic diversity of the pathogen, sampled over time. If the resident phenotype has accumulated neutral mutations, so that many genetic variants are present at low frequencies, a rapid expansion of an immune escape mutant will cause a temporary reduction in pathogen diversity. This pattern was well described in early research on epochal evolution *in silico* [40] and recently used to postulate selective sweeps in influenza [26]. Even if a selective sweep does not occur, a temporary reduction in diversity should be observed following punctuated immune escape as the new variant increases in

frequency in the gene pool. Similar to the age-related patterns, the reduction in diversity is transient, disappearing as the antigenic variant mutates.

Analyses of the relatedness of sampled genetic sequences can also suggest relevant phenotypes. Positively selected strains should produce many descendents with high sequence similarity. Simple measures of Hamming distance [41] and more complex analyses of nucleotide co-occurrence networks [42] have been able to accurately infer populations corresponding to the major antigenic phenotypes of influenza. Exceptions (e.g. abrupt changes in the relatedness of genes outside hemagglutinin [42]) are useful for generating alternative hypotheses about other genes that might be under selection and might drive pathogen population dynamics. A major danger in this analysis is not accommodating for spatial or temporal sampling bias in reconstructing pathogen population diversity. For this reason, analyses based on genetic ancestry—phylogenetics—can be more revealing.

Phylogenies

Neutral evolution causes a characteristic, balanced branching pattern [43]. By contrast, directional selection leads to variability in lineage extinction rates, and can result in a skewed or ladderlike topology. Punctuated immune escape via epochal evolution has recently been shown to generate a tree combining aspects of both processes: clades whose members belong to the same phenotype will show neutral evolution, whereas over longer timescales (represented by the trunk of the tree), the topology will be ladderlike. This pattern was observed in influenza [26] (Figure 1d) and norovirus [19,20] (Figure 1e), and similar tree topologies appear in dengue [44] (Figure 1f) and HIV within hosts [34,45]. The most strongly selected substitutions over time are those that occur along the trunk [46,47]. To give evidence of punctuated immune escape (as opposed to punctuated selection for, e.g. antiviral resistance), the trunk substitutions need to occur in antigenic sites.

More sophisticated methods that take advantage of phylogenies and precise sampling dates are rapidly evolving. For example, the survival times of lineages can be used to reveal periods of phenotypic stasis and rapid strain turnover. In application to influenza, this type of analysis, with survival times correlating positively with numbers of non-synonymous substitutions in epitopes, has suggested episodic, immune-mediated selection [48]. Another example is the recent integration of coalescent theory with population dynamic models to produce 'skyline plots' [49,50]. This method uses branch lengths and dates of isolation to reconstruct past population diversity from phylogenies and could conceivably be used to test for punctuated immune escape. One challenge in this method is finding an accurate model of disease dynamics [16]. As in other methods based on genetic data, a second challenge is accurately inferring that selection is for changes in antigenic properties.

Recombination and allele switching will complicate phylogenetic inference. Depending on the relatedness of recombined genes or expressed alleles, directional selection might not produce a skewed tree. These factors also create challenges for coalescent inference. Patterns in lineage turnover, however, will still hold.

Table 1. Observable patterns to detect punctuated immune escape

Source of data	Prediction under punctuated immune escape	Confounding or limiting factors
Hosts		
Magnitude of cases	An increase in the number of cases should occur as an immune escape mutant invades.	Variability in extrinsic factors; intrinsic dynamics other than those associated with immune escape
Ages of infected individuals	Average age should increase transiently.	Differences in contact rates by age
Pathogens		
Antigenic phenotype, such as HI inhibition assays or structure	Many genotypes should map into one phenotype (for epochal evolution); phenotypes should be distinct; novel phenotypes should increase more rapidly in frequency than expected under neutral evolution.	Must be able to infer antigenic phenotype accurately.
dN and dS	The nonsynonymous-to-synonymous substitution rate ratio will episodically be greater than one.	Observed patterns might not be related to antigenicity *per se*; challenges associated with computing dN/dS on isolated branches.
Substitution rates	Episodic selection resulting from punctuated immune escape should precipitate a burst of substitutions.	Observed patterns might not be related to antigenicity *per se*; disease dynamics and changes in selective constraints.
Genetic diversity	Genetic diversity should transiently decrease as a new antigenic variant emerges; genetic clustering of sequences should occur during periods of antigenic stasis.	Observed patterns might not be related to antigenicity *per se*.
Phylogeny (topology)	Balanced branching pattern in the short run, in some cases combined with a skewed topology in the long run	Observed patterns might not be related to antigenicity *per se*.
Phylogeny (dated)	Periods of rapid lineage turnover; survival times should correlate positively with nonsynonymous substitution rates in epitopes.	Observed patterns might not be related to antigenicity *per se*.
Phylogeny (dated)	Fluctuations in diversity over time revealed by skyline plots	Observed patterns might not be related to antigenicity *per se*; method sensitive to population dynamic model assumed; skyline plots complicated by recombination and allele switching.

Box 1. Outstanding questions

○ For which pathogens does immune escape dominate their evolution? Which of these pathogens evolve by punctuated immune escape? Processes other than immune escape might dominate pathogen evolution. Such processes include selection on phenotypic traits such as virulence or the nonselective effects of spatial structure [1].
○ Which kinds of immunity (e.g. humoral, cellular) are likely to drive punctuated immune escape? Evasion from herd immunity through mutations at T cell epitopes and antibody epitopes has been well documented for many pathogens. It would be useful to determine the relative importance of each kind of immune escape for different pathogens.
○ What are other patterns for detecting punctuated immune escape by pathogens? Within-host data could provide evidence of immune escape. For example, information on the dynamics of host immune cells and immunoglobulins might complement observations of the pathogen population.
○ How differently are individual hosts affected by new antigenic variants that escape immunity at the level of the entire host population? An antigenic variant might not be novel to all hosts if hosts differ in which epitopes their immune systems target. Adults, with broader antibody repertoires, might perceive fewer novel variants than children and be less susceptible or infectious. These differences could alter the epidemiological signature of punctuated immune escape.
○ When does punctuated immune escape cause a selective sweep? The emergence of a new antigenic variant might not result in replacement of older variants. Understanding how cross-immunity, host lifespan and other factors affect pathogen diversity when an immune escape mutant emerges is a ripe area of future research.
○ What are open problems associated with integrating multiple sources of data for statistical inference? Methods and approaches need to be developed to guide us in choosing the most effective combinations of observations. Many of the proposed approaches rely on the same data (see Table 1) and are therefore not independent. A challenge will be to decide appropriate weights to give different kinds of data, including sequences, phenotypes and disease dynamics.

Capturing escape

The host and pathogen patterns described above provide a diverse set of ways to identify antigenic variants and the timing of punctuated immune escape. However, each pattern has confounding or limiting factors (summarized in Table 1). Because most observations have different limitations, successful statistical models of pathogens exhibiting punctuated immune escape will require integrating multiple sources of data. For example, increases in incidence, together with a transient increase in the mean age of infected individuals and a drop in pathogen diversity, strongly point toward immune escape. More difficult scenarios arise when signals from the different observations conflict.

Our ability to resolve these disputes depends critically on having models of the appropriate dimensionality. Finding the appropriate dimensionality requires both empirical and theoretical progress (Box 1). A broadening of our observations and methods to encompass punctuated immune escape will be an essential strategy for understanding the intrinsic and extrinsic factors regulating the dynamics of many common diseases.

References

1 Grenfell, B.T. *et al.* (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303, 327–332
2 Grenfell, B. and Bjornstad, O. (2005) Sexually transmitted diseases—epidemic cycling and immunity. *Nature* 433, 366–367

3 Rogers, D.J. *et al.* (2002) Satellite imagery in the study and forecast of malaria. *Nature* 415, 710–715

4 Pascual, M. and Ellner, S.P. (2000) Linking ecological patterns to environmental forcing via nonlinear time series models. *Ecology* 81, 2767–2780

5 Zhou, G. *et al.* (2004) Association between climate variability and malaria epidemics in the East African highlands. *Proc. Natl. Acad. Sci. U. S. A.* 101, 2375–2380

6 Pascual, M. *et al.* (2000) Cholera dynamics and El Niño-Southern Oscillation. *Science* 289, 1766–1769

7 Bobashev, G.V. *et al.* (2000) Reconstruction of susceptible and recruitment dynamics from measles epidemic data. *Math. Popul. Stud.* 8, 1–29

8 Koelle, K. and Pascual, M. (2004) Disentangling extrinsic from intrinsic factors in disease dynamics: a nonlinear time series approach with an application to cholera. *Am. Nat.* 163, 901–913

9 Xia, Y.C. *et al.* (2005) Semiparametric estimation of the duration of immunity from infectious disease time series: influenza as a case-study. *J. R. Stat. Soc. Ser. C Appl. Stat.* 54, 659–672

10 Koelle, K. *et al.* (2005) Refractory periods and climate forcing in cholera dynamics. *Nature* 436, 696–700

11 Finkenstädt, B.F. and Grenfell, B.T. (2000) Time series modelling of childhood diseases: a dynamical systems approach. *J. R. Stat. Soc. Ser. C Appl. Stat.* 49, 187–205

12 Pease, C.M. (1987) An evolutionary epidemiologic mechanism, with applications to type A influenza. *Theor. Popul. Biol.* 31, 422–452

13 Dushoff, J. *et al.* (2004) Dynamical resonance can account for seasonality of influenza epidemics. *Proc. Natl. Acad. Sci. U. S. A.* 101, 16915–16916

14 Gupta, S. *et al.* (1998) Chaos, persistence, and evolution of strain structure in antigenically diverse infectious agents. *Science* 280, 912–915

15 Recker, M. *et al.* (2007) The generation of influenza outbreaks by a network of host immune responses against a limited set of antigenic types. *Proc. Natl. Acad. Sci. U. S. A.* 104, 7711–7716

16 Rambaut, A. *et al.* (2008) The genomic and epidemiological dynamics of human influenza A virus. *Nature* 453, 615–619

17 Frank, S.A. and Barbour, A.G. (2006) Within-host dynamics of antigenic variation. *Infect. Genet. Evol.* 6, 141–146

18 Smith, D.J. *et al.* (2004) Mapping the antigenic and genetic evolution of influenza virus. *Science* 305, 371–376

19 Siebenga, J.J. *et al.* (2007) Epochal evolution of GGII.4 norovirus capsid proteins from 1995 to 2006. *J. Virol.* 81, 9932–9941

20 Lindesmith, L.C. *et al.* (2008) Mechanisms of GII.4 norovirus persistence in human populations. *PLoS Med.* 5, e31

21 McKeating, J.A. *et al.* (1989) Characterization of HIV-1 neutralization escape mutants. *AIDS* 3, 777–784

22 Zhang, C. *et al.* (2005) Clade replacements in dengue virus serotypes 1 and 3 are associated with changing serotype prevalence. *J. Virol.* 79, 15123–15130

23 Bennett, S.N. *et al.* (2003) Selection-driven evolution of emergent dengue virus. *Mol. Biol. Evol.* 20, 1650–1658

24 Guzman, M.G. *et al.* (2000) Do escape mutants explain rapid increases in dengue case-fatality rates within epidemics? *Lancet* 355, 1902–1903

25 van Nimwegen, E. *et al.* (1997) Finite populations induce metastability in evolutionary search. *Phys. Lett. A* 229, 144–150

26 Koelle, K. *et al.* (2006) Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* 314, 1898–1903

27 Greene, S.K. *et al.* (2006) Patterns of influenza-associated mortality among US elderly by geographic region and virus subtype, 1968–1998. *Am. J. Epidemiol.* 163, 316–326

28 Huovilainen, A. *et al.* (1987) Evolution of poliovirus during an outbreak: sequential type 3 poliovirus isolates from several persons show shifts of neutralization determinants. *J. Gen. Virol.* 68, 1373–1378

29 Wolfs, T.F. *et al.* (1991) Naturally occurring mutations within HIV-1 V3 genomic RNA lead to antigenic variation dependent on a single amino acid substitution. *Virology* 185, 195–205

30 Frank, S.A. (2002) *Immunology and Evolution of Infectious Disease,* Princeton University Press

31 Fraser, C. *et al.* (2005) Neutral microepidemic evolution of bacterial pathogens. *Proc. Natl. Acad. Sci. U. S. A.* 102, 1968–1973

32 Suzuki, Y. (2006) Natural selection on the influenza virus genome. *Mol. Biol. Evol.* 23, 1902–1911

33 Nelson, M.I. *et al.* (2006) Stochastic processes are key determinants of the short-term evolution of influenza A virus. *PLoS Pathog.* 2, 1144–1151

34 Yamaguchi, Y. and Gojobori, T. (1997) Evolutionary mechanisms and population dynamics of the third variable envelope region of HIV within single hosts. *Proc. Natl. Acad. Sci. U. S. A.* 94, 1264–1269

35 Zhang, J.Z. *et al.* (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* 22, 2472–2479

36 Kimura, M. (1983) *The Neutral Theory of Molecular Evolution,* Cambridge University Press

37 Gillespie, J.H. (1984) The molecular clock may be an episodic clock. *Proc. Natl. Acad. Sci. U. S. A.* 81, 8009–8013

38 Bazykin, G.A. *et al.* (2006) Bursts of nonsynonymous substitutions in HIV-1 evolution reveal instances of positive selection at conservative protein sites. *Proc. Natl. Acad. Sci. U. S. A.* 103, 19396–19401

39 Takahata, N. (1987) On the overdispersed molecular clock. *Genetics* 116, 169–179

40 Schuster, P. (2003) Molecular insights into evolution of phenotypes. In *Evolutionary Dynamics: Exploring the Interplay of Selection, Accident, Neutrality, and Function* (Crutchfield, J.P. and Schuster, P., eds), pp. 163–215, Oxford University Press

41 Plotkin, J.B. *et al.* (2002) Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* 99, 6263–6268

42 Du, X. *et al.* (2008) Networks of genomic co-occurrence capture characteristics of human influenza A (H3N2) evolution. *Genome Res.* 18, 178–187

43 Heard, S.B. (1992) Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. *Evolution Int. J. Org. Evolution* 46, 1818–1826

44 Adams, B. *et al.* (2006) Cross-protective immunity can account for the alternating epidemic pattern of dengue virus serotypes circulating in Bangkok. *Proc. Natl. Acad. Sci. U. S. A.* 103, 14234–14239

45 Shankarappa, R. *et al.* (1999) Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J. Virol.* 73, 10489–10502

46 Bush, R.M. *et al.* (1999) Predicting the evolution of human influenza A. *Science* 286, 1921–1925

47 Lemey, P. *et al.* (2007) Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. *PLOS Comput. Biol.* 3, 282–291

48 Wolf, Y.I. *et al.* (2006) Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. *Biol. Direct* 1, 34

49 Pybus, O.G. *et al.* (2000) An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 155, 1429–1437

50 Drummond, A.J. *et al.* (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22, 1185–1192

51 Fitch, W.M. *et al.* (1997) Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc. Natl. Acad. Sci. U. S. A.* 94, 7712–7718

52 Lindesmith, L.C. *et al.* (2008) Mechanisms of GII.4 norovirus persistence in human populations. *PLoS Med.* 5, 269–290

53 Shiu, S.Y.W. *et al.* (1992) Envelope protein sequences of dengue virus isolates TH-36 and TH-Sman, and identification of a type-specific genetic marker for dengue and tick-borne flaviviruses. *J. Gen. Virol.* 73, 207–212

54 Zhang, C. *et al.* (2006) Structure and age of genetic diversity of dengue virus type 2 in Thailand. *J. Gen. Virol.* 87, 873–883

55 Blok, J. *et al.* (1989) Variation of the nucleotide and encoded amino acid sequences of the envelope gene from eight dengue-2 viruses. *Arch. Virol.* 105, 39–53

56 Leitmeyer, K.C. *et al.* (1999) Dengue virus structural differences that correlate with pathogenesis. *J. Virol.* 73, 4738–4747