Ecological and evolutionary dynamics of multi-strain RNA viruses

Ecological and evolutionary dynamics of multi-strain RNA viruses

Dennis N. Makau[1], Samantha Lycett[2], Matthew Michalska-Smith[3], Igor A. D. Paploski[1], Maxim C.-J. Cheeran[1], Meggan E. Craft[3], Rowland R. Kao[2], Declan C. Schroeder [1,4], Andrea Doeschl-Wilson[2], Kimberly VanderWaal[1*]

[1] Department of Veterinary Population Medicine, University of Minnesota, USA
[2] Roslin Institute, University of Edinburgh, UK
[3] Department of Ecology, Evolution, and Behavior, University of Minnesota, USA
[4] University of Reading, Reading, UK
*Corresponding author, kvw@umn.edu

Abstract

Potential interactions amongst co-circulating viral strains in host populations are often overlooked in the study of virus transmission. However, these interactions likely shape transmission dynamics by influencing host immune responses or altering the relative fitness amongst co-circulating strains. In this Review, we describe multi-strain dynamics from ecological and evolutionary perspectives, outline scales in which multi-strain dynamics occur, and summarize important immunological, phylogenetic, and mathematical modeling approaches used to quantify interactions amongst strains. We also discuss how host-pathogen interactions influence the co-circulation of pathogens. Finally, we highlight outstanding questions and knowledge gaps in the current theory and study of ecological and evolutionary dynamics of multi-strain viruses.

Introduction

The existence of multiple co-circulating strains or phylogenetic lineages is common for many pathogens, particularly for rapidly evolving RNA viruses. As viruses evolve, immune responses generated against a past variant may become less effective, which creates a complex system, with different antigenic variants interacting through the cross-immunity that is generated within hosts [1,2]. In the past decade, the increasing ubiquity of viral genetic data has created opportunities to interrogate how ecological processes, such as competition for susceptible hosts [3], shape both the epidemiological and evolutionary dynamics of many viruses. In multi-strain dynamics, epidemics occur when a novel viral variant evolves and evades host immunity created by its predecessors [4], or when the fitness of an existing variant is modulated by the changing immunity of the population independent of the ability of the virus to mutate [5].

Potential ecological and evolutionary interactions amongst co-circulating viral strains are rarely investigated, particularly in animal populations, even though these interactions likely drive transmission dynamics both through immune-mediated competition and natural selection. These processes may ultimately shape the temporal and spatial distribution of viral genetic diversity across multiple scales, particularly when there are underlying spatiotemporal heterogeneities in the susceptibility of hosts as a result of previous patterns of viral circulation. The challenges in controlling Influenza A in swine due to vaccine inefficacies and emergence of distinct divergent

viral communities attributed to viral 'mixing' in pigs are a good example of the potential benefits of understanding multi-strain dynamics of viruses [6–9].

While definitions of "strain" vary widely and are often pathogen-specific, here we broadly define strain as when a pathogen occurs in identifiable phylogenetic lineages or clades that that also differ phenotypically [10]. Immunogenic or antigenic phenotype variation may alter the fitness of a genetic variant in terms of its ability to compete with other variants. Phenotypic variation in virulence, transmissibility, or other infection attributes may also confer fitness advantages (or disadvantages) and could be considered the basis for strain structure. While phylogenetic structure can be useful for reconstructing transmission history and patterns of dispersal, we would not consider the existence of phylogenetic structure to constitute multi-strain dynamics in the absence of phenotypic variation amongst lineages. We also do not consider the evolution of distinct phylogenetic clades based on geographic isolation to be representative of multi-strain dynamics unless those clades co-occur in the same host population.

While recent reviews of multi-strain dynamics of pathogens have focused on mathematical modeling frameworks for investigating strain-host interactions [2,11,12], we synthesize immunological, ecological, and evolutionary drivers and implications of multi-strain dynamics in rapidly evolving viruses. We first contrast conceptual differences and similarities between multi-strain dynamics from ecological versus evolutionary perspectives, then outline scales in which multi-strain dynamics occur, and summarize immunological, phylogenetic, and mathematical modeling approaches used to quantify interactions amongst strains. While multi-strain dynamics may occur across a range of pathogens, we focus our discussion on multi-strain viruses. RNA virus-host systems are particularly likely to exhibit multi-strain dynamics because their high mutation rate allows for ecological and evolutionary processes to occur on the same time scale.

Ecological versus evolutionary dynamics

Although the difference between the ecological and evolutionary perspectives on dynamics of multi-strain pathogens is somewhat arbitrary given that both processes occur simultaneously, this conceptual division is useful in summarizing key theories and methodological approaches surrounding multi-strain dynamics. In both perspectives, past infection by one variant results in only partial cross-immunity to a related strain, and such partial cross-protection is expected to result in a change in the susceptibility, infectivity, and/or clinical signs in the partially immune host. Ecological multi-strain dynamics generally encapsulate situations where a discrete number of antigenic alternatives or strains exist in the population and strains are assumed not to evolve phenotypically (i.e., only neutral or nearly neutral evolution occurs on the time-scale of interest). Cross-immunity amongst strains is variable and fitness is frequency-dependent based on residual immunity in the population developed against previous strains, as has been suggested for human influenza[13]. Questions of interest focus on how and why the relative frequency of different strains changes through space and time. In contrast, evolutionary multi-strain dynamics focuses more on how competition and natural selection amongst genetic variants can drive genetic change, allowing for the emergence of new genetic variants or strains through time (Figure 1). "Immune escape" occurs when a novel antigenic variant evolves that is

no longer controlled by individual/herd-level immunity [14,15]. In some instances, small mutations (resulting in minimal genetic change) may result in considerable antigenic changes if substitutions occur in immunogenic sites. In such cases, genetic distance may not be a useful measure of the extent of cross-immunity amongst strains.

Common to both perspectives are a) the existence of variable levels of cross-immunity between strains or variants, and b) viral variants that are more effective at evading host immunity (induced by previous exposure to a related strain) have higher fitness, and thus can out-compete other variants either within an individual or at the population level. Depending on the nature of cross-immunity, this can lead to fitness advantages for strains and variants occurring at low frequencies as compared to more common strains or variants towards which host immunity is already strong. Theory predicts that due to imperfect cross-immunity and frequency-dependent fitness amongst co-circulating strains, rare strains bearing novel antigenic mutations are expected to be able to spread more widely in the host population but then subsequently decline as herd immunity rises [16]. Cyclic or chaotic changes in the frequency of different strains occur in host-pathogen systems with intermediate levels of immune selection. These changes can complicate and restrict our ability to interpret and predict the outcome of interventions, including vaccination[1,13,17] or selection for disease resistance traits in hosts, which is increasingly implemented in animal-based agriculture [18].

Figure 1. Influence of partial cross-immunity on evolutionary dynamics of multi-strain pathogens. a) Antigenic drift/shift: After infection by a specific variant at t_0, the virus begins to accrue genetic mutations as it replicates, creating a viral cloud (t_1). Partial cross-immunity can exert evolutionary selection pressures by which more divergent variants are likely to propagate through time ($t_1 \rightarrow t_2$, either within or between hosts) due to their ability to evade host immunity. b) This process can result in a shift in the antigenic phenotype of viral populations through time. c) Pathogens characterized by antigenic drift/shift often exhibit ladder-like phylogenetic trees wherein older strains go extinct and are replaced by newer strains, as suggested for influenza viruses[19]. d) Ecological antigenic shift: Immunity in the population creates ecological pressure for antigenically divergent strains to increase in frequency through time ($t_1 \rightarrow t_2$), resulting in e) shifts in the antigenic phenotype as a new dominant strain in the population takes over. f) Pathogens characterized by ecological antigenic shifts likely exhibit more symmetrical/balanced phylogenetic trees with longer branches, as hypothesized for sub-lineages within porcine reproductive and respiratory syndrome virus type 2[19–21].

Scales of action and impact of multi-strain dynamics

Multi-strain dynamics can be quantified across multiple scales, from within-host processes to host-to-host transmission within a single population or between populations. Both ecological and evolutionary processes can occur at each of these scales. Different scales are visualized in Figure 2, where greater similarity in color indicates hosts with higher levels of cross-immunity to each other's viruses (based on past by exposure to more similar antigenic viral variants) as compared to two hosts with more divergent colors. Viral populations replicating

within an individual host (Figure 2a) form a viral cloud of highly-related genetic variants, sometimes referred to as quasi-species [22]. Some genetic mutations may alter a variant's antigenic phenotype allowing immune escape to occur. Due to their ability to evade host immunity, the relative frequency of escape mutants may increase within the host and thus increase the likelihood that they are transmitted (however, see [23] for a discussion of how within-host adaptation of viral populations may be detrimental to between-host transmission). Despite diminishing viral populations as infection progresses, surviving variants are likely to have mutations favored in the immune-mediated selection process, as observed for porcine reproductive and respiratory syndrome virus [24,25]. An escape mutant that emerges from within-host evolutionary processes has the potential to propagate within the population due to its antigenic novelty against which the population has limited cross-immunity (light blue individual in Figure 2b).

Successful transmission of different variants between hosts can be influenced by bottlenecks in host susceptibility and infectivity. Reduced infectivity could be expected if the host's immunity (e.g., based on the history of exposure or host genetics) towards the viral variant is sufficient to reduce viral replication and shedding, making transmission less likely. Susceptibility bottlenecks can occur in between-host infection chains, e.g., if past exposure to a similar variant influences a host's susceptibility to a new variant. Because of variability in cross-immunity, for example, a host infected by the light blue variant in Figure 2b is more likely to transmit to hosts that have immunity to more dissimilar variants (e.g., darker blue) than hosts with more similar immunologic histories (Figure 2b). Thus, the ultimate success of the light blue variant in spreading within the population is expected to be higher when the frequency of light blue is low because fewer individuals would have developed immunity against it. If we amplify this concept to consider between-population dynamics, we can expect heterogeneities in population immunity to shape the invasion success of different variants in new populations. For example, a dark blue variant may be much more likely to invade a population that has strong herd immunity towards green variants than a population that has an immunological history of blue variants (Figure 2c). Across all these scales, antigenic novelty is expected to confer some degree of fitness advantage, which will allow more divergent variants to propagate within- and between-hosts and consequently shape the invasion success of different strains across populations.

Figure 2: Scales at which multi-strain dynamics occur, from a) within-host to b) between-host to c) between populations. *Colors:* Greater similarity of colors represent higher levels of cross-immunity conferred by exposure to more similar antigenic viral variants. *Arrows*: Thickness of arrows represent the relative likelihood of spread between hosts or populations with different immunological histories. Successful transmission of the light blue variant between hosts is more likely if the recipient host has previous exposure to a more dissimilar virus, such as the darker blue variant (panel b). Similarly, the dark blue variant is more likely to be transmitted to a population largely exposed to the dissimilar green variant than a population exposed to the more similar variant (panel c) .'

Quantifying immunogenic interactions between strains

Viral entry into cells can occur by various mechanisms depending on the virus species, including fusion of the virion membrane with the cell membrane, or receptor-mediated endocytosis [26]. Cell entry of RNA viruses typically involves binding of viral surface proteins to a cellular receptor [26–31], which also triggers host immune responses [29,32–34]. These responses often include antibodies that bind to the surface antigens (proteins) on a virion to hinder binding of the viral proteins to the cellular receptors, preventing infection of the cell [35,36]. Consequently, whether infection with one strain of a virus can influence the host's susceptibility to another viral strain may depend, amongst other things, on the extent to which the host's immune system recognizes and inhibits infection with the newly infecting strain. Thus, partial immunity amongst genetically/immunogenically similar strains [37,38] can shape the fitness of different strains and influence the likelihood that multiple strains co-circulate in a population.

To quantify antigenic distance between strains, binding and cross-neutralization assays are often used to measure the cross-reactivity of immune reactions elicited by different strains. Briefly, hyperimmune serum is generated against specific viruses (hypothetical strains A, B, C) by exposing naïve animals. Different viruses are then cross-reacted with serially diluted sera and the highest neutralization titer is identified. A comparison between neutralization titers achieved by serum A on virus A (homologous titer) and the titers achieved for serum A against virus B or C (heterologous titer) can be interpreted as an indicator of antigenic difference/similarity between the strains [39]. These assays have been extensively employed in the study of influenzas, and cross-immunity profiles across a panel of different strains are often mapped through the application of antigenic cartography[40–42]. Antigenic cartography, a computational technique used for graphical visualization of antigenic distances obtained from inhibition assays [43], can be used to visualize the genetic and antigenic differences amongst co-circulating variants and identify clusters of variants with similar immune profiles [44]. Data from panels of cross-reactivity assays can be combined with genetic mapping and epidemiological data, and analyzed using machine learning, and other statistical approaches to identify specific amino acid changes that underlie antigenic phenotypes and potentially result in the emergence of different viral variants [45–47]. These tools can be used to refine the relationship between genetic and antigenic variation amongst co-circulating strains of a virus in a population.

Evolutionary processes for multi-strain pathogens

Evaluating the existence of multiple strains co-circulating in a population is a complex process because ecological and host factors may influence how evolution manifests in different strains, and interactions between strains against the backdrop of host population structure needs to be disentangled to understand the evolutionary trends of the virus [48]. However, quantifying the nature of multi-strain dynamics and drivers of co-evolution of multiple strains is of epidemiological significance when dealing with outbreaks of infectious diseases in populations, as observed for SARS-CoV-2 [49], dengue fever [3,50], and influenza [51]. Multiple co-circulating strains can also influence disease severity through antibody-dependent enhancement or shaping the evolution of virulence (Box 1). Below, we summarize different tools and approaches that can be applied to investigate evolutionary dynamics of multi-strain viruses.

Bayesian phylodynamic models provide a versatile framework for the study of pathogens over time through the inclusion of ecological or host-specific factors that may influence viral evolution in a landscape [15,21,52–55], including how host traits, population structure, and environmental characteristics impact the emergence, spread, and turnover of viral populations [56–59]. The flexibility of including discrete or continuous traits [53,56,60], estimating viral population change under structured coalescent models [61], and inclusion of structured birth-death models [62] in these analyses allow for more nuanced estimation of the prevalence of various mutations in viral populations. These methods also can be used to reconstruct viral population dynamics, identify emergence, population expansions, and extinction events of different strains, and quantify the sustained co-circulation of distinct viral populations while accounting for variation in host population structure [63]. While the overall amount of genetic diversity through time may be somewhat constant for endemic multi-strain pathogens, analyzing each lineage separately can help visualize these emergence-extinction cycles, as seen with porcine reproductive and respiratory syndrome virus [20].

Phylogenetic branching patterns can be analyzed to provide insights on multi-strain dynamics and immune-mediated selection. This analytical approach has been used extensively to describe the ladder-like phylogeny of seasonal influenza and some coronaviruses associated with immune-mediated selection [64,65]. In these examples, specific lineages of the viruses circulate over relatively short time-periods before being replaced by a new strain, creating a ladder-like tree (Figure 1c). As a result, the most recent common ancestor for contemporary variants is relatively recent. However, phylogenies may not always exhibit step-like temporal topologies since ancestral clades may continue to persist even as descendant clades expand, as observed for foot and mouth disease virus and different lineages within the same serotype of dengue virus [66,67], with immune-mediated competition dictating the fitness of different clades through time (Figure 1f). Two strains are expected to be antigenically differentiated in order to co-circulate within a host population without one going extinct. Additionally, host genetic diversity, environmental heterogeneities, and spatial structure of the host population may also contribute to diversifying evolution (increasing genetic and antigenic distance), with the impact of the latter two dependent on the pathogen's transmission mode and dispersal capabilities [68–70].

Selection pressures and resulting mutations responsible for adaptation or immune evasion are not always easily identifiable from phylogenetic trees alone [71]. Therefore, we describe four approaches that can complement phylodynamic models to evaluate rates of viral evolution depicted on phylogenetic trees: a) Tajima's D is a statistical test used to calculate the genetic deviation of a population from a neutrally evolving population [72] and can be used to identify non-random mutations, bottlenecks and selective pressure driving the evolution process [73]. Tajima's D relies on two measures of genetic difference between organisms: the mean pairwise differences in genetic sequences and the number of differentiating sites. b) Based on the tree topology, fitness models can be used to estimate the rate of population expansion and fitness of a viral variant in a population [74,75]. The local branching index (LBI), for example, is a statistical calculation to estimate the fitness of a node (an ancestor) in a phylogenetic tree by calculating the size of a node's neighborhood (number of descendants/progeny of a node) over a given period in time [75,76]. Mutations that increase viral fitness are associated with higher LBI, and LBI has been shown to correlate with other metrics of fitness [76]. Since nodes with higher LBI

are likely to be ancestors for future clades [76], LBI can be used to predict expansions of different clades in a phylogenetic tree. Co-circulation of strains may be expected in cases where several contemporary nodes have near equal LBI. c) The fixation index (F_{ST}) is a measure of changes in a population associated with the population's genetic structure. Locus-by-locus F_{ST} using analysis of molecular variance can be used to identify potential genomic regions that determine the difference in accumulation of group-specific genes by a pathogen [77–79]. By comparing locus-by-locus differences, one can distinguish between groups of genomes isolated from a host population and determine the presence of one or multiple strains. d) The rate of synonymous (dS) vs non-synonymous (dN) mutations can also elucidate dynamics of viral evolution [80]. Synonymous mutations are nucleotide substitutions that do not change the amino acid coded for by the respective codon while non-synonymous mutations result in changes in the amino acids. Synonymous mutations are generally considered neutral as they do not affect protein phenotype (though this is not always the case [81–83]), and the rate at which such "neutral" mutations occur is typically interpreted as the expectation for background rates of change. Non-synonymous mutations may impact viral fitness if they are deleterious or beneficial, and thus may experience negative or positive selection pressures [84]. Calculating the codon-level dN/dS ratios can help identify whether selective pressure in a population is driving viral evolution. Higher than expected rates of non-synonymous change, usually inferred when dN/dS >1, can be interpreted as evidence of positive or diversifying selection on that codon, suggesting that mutations resulting in amino acid changes are favored [85]. Positive selective pressure at antigenic sites is indicative of immune-mediated selection. Combining dN/dS analysis with host/ environmental factor analysis can further identify drivers of strain/variant co-circulation [86].

[Start Box 1]

Box 1: Can multi-strain dynamics drive the evolution of virulence?

While immune-mediated selection has received the majority of attention in the evolutionary processes of multi-strain pathogens, multi-strain dynamics also have important implications for the evolution of virulence and tradeoffs with transmissibility. If we assume that faster within-host multiplication by a strain allows it to outcompete other strains in a co-infected host, then the prevalence of hosts coinfected by multiple strains should favor the evolution of increased virulence (assuming low levels of cross-protection and that faster multiplication is associated with increased virulence, which is not always true)[87]. However, in the absence of competition within co-infected hosts, high virulence may be disadvantageous as it may reduce the infectious period (through infection-induced mortality) and thus decrease the ability of a pathogen to transmit within population [87]. This concept can be translated to the meta-population level: increased virulence may be favored if sub-populations are co-infected by multiple competing strains, while at the same time, a highly virulent/rapidly transmitting strain may go extinct in a sub-population before it can disperse to a new susceptible sub-population [88]. Opportunities for the pathogen to disperse between sub-populations are shaped by host (or vector) movement, which further complicates understanding of optimal virulence for multi-strain pathogens. Thus, the frequency of co-infection in hosts and host-populations by multiple strains,

and the nature of cross-protection between competing strains [89], may alter the tradeoff between virulence and transmissibility and thus influence the evolution of virulence.

[End Box 1]

Mathematical models of multi-strain pathogens

Mathematical models have been instrumental in understanding the dynamics of disease outbreaks and spread. They facilitate the estimation and prediction of changes in pathogen population size, the speed and duration of epidemics, and the impact of control measures. Despite the ubiquity of strain structure [90], models that incorporate such diversity have remained focused on a few prevalent human diseases, such as influenza [4,5], human papillomavirus [91,92], Dengue fever [3,93] and human immunodeficiency virus (especially in the context of the emergence of treatment-resistant strains)[94,95]. Due to their inherent complexity and differences in assumptions about model structure, models of multi-strain disease can exhibit a wide variety of dynamics, from globally stable equilibria to cyclic or chaotic fluctuations in the frequency of different strains. Thus, multi-strain dynamics are difficult to predict.

Multi-strain disease models can track either individuals (agent-based models [96]) or changing proportions of different infection states (compartmental models), but the underlying dynamics are similar: individuals/groups of the population (hereafter just "individuals" for simplicity) are divided into a finite set of possible classes based on their exposure history. In the simplest case, this mirrors the commonly used single-strain SIR framework where each individual is either susceptible to a pathogen, currently infectious, or recovered and no longer capable of being infected nor infecting others. Considering a pathogen with two strains, one might use an SI_1I_2R model in which individuals are delineated into one of four classes: susceptible to both diseases, infectious with each of the two potential strains, or immune to further infection from either.

The above example highlights two key considerations that arise when modelling multi-strain diseases. First, what is the optimal model structure in terms of the number and resolution of the classes? This in turn depends on how one classifies previous infections (does it matter which strains an individual has been exposed to, the order of infection, or simply how many?), and has dramatic consequences for the computational complexity of a model [97]. Additionally, as with single-strain models, one must consider whether and how to implement population structure and heterogeneity among individuals (e.g. differences in susceptibility) [48,98,99]. Second, how should cross-immunity be modelled? Cross-immunity can vary in degree (how much less likely is infection with strain B following infection with strain A?), duration (is immunity waning, or lifelong?) and implementation (does immunity affect susceptibility or infectivity?).

In the face of this complexity, multi-strain disease modelers have frequently focused on systems with only two competing strains (*e.g.,*[95,96,100–105]), and employed simplifying assumptions, such as the discretization of a finite strain space. Put another way, strains are typically modelled as a set of strains that are all categorically different from one another (but see

[106]). This is typically accomplished by assuming that infectious agents are clustered into functionally equivalent antigenic phenotypes [90]. Models of multi-strain disease are more disposed to non-stationary dynamics (*e.g.,* cycles/chaos) than their single-strain counterparts [107,108], largely driven by the degree of cross-immunity. When infection by one genetic variant provides near-complete immunity to another, stable and discrete strains emerge, whereas intermediate levels of cross-immunity lead to cyclic or chaotic fluctuations in strain prevalence [1]. Importantly, however, this effect can be overridden if strains differ too much in their epidemiological parameters [108].

The incorporation of evolution into models of multi-strain disease introduces a wide range of additional complexities, but, in general, the framework for modeling evolving pathogens consists of two linked modules: one for the epidemiology, as discussed at length above, and one for the evolution. The proximity of this linkage depends on the nature of the evolutionary module, which can range from explicitly modeling nucleotide substitutions [109], to allowing epidemiological parameter values to evolve (e.g., transmissibility)[104], or to adding a new parameter corresponding to an abstract phenotype [106,110] or genotype space [111,112]. One of the more studied areas of multi-strain dynamics is the evolution and emergence of novel variants within a treatment and resistance paradigm.

To improve fit to empirical systems, some models incorporate spatial structure, which can promote strain coexistence [113]. Cyclical patterns of strain dominance, for example, can be produced in the absence of immune interactions if host population structure is introduced. In a model of Dengue virus, for example, spatial sub-structuring of the population explained stochastic differences between neighboring areas in the prevalence of different serotypes, even in the absence of immune-mediated competition [3]. Finally, host contact networks can introduce another layer of complexity through the influence of local network structure on disease spread [114].

Population structure and stochasticity

Host population structure can have major impacts on how multi-strain dynamics manifest by impacting the frequency with which strains serially or co-infect hosts. For example, host contact networks can impact the strength of immune-mediated selection pressure by influencing how rapidly the network becomes locally saturated with immune hosts [114], and thus increase the likelihood of escape mutants to evolve. In virulence evolution (Box 2), the severity of the trade-off between competition amongst strains within co-infected sub-populations and transmissibility between sub-populations is reduced if between-population spread occurs frequently [88]. In other words, increased opportunities for viral dispersal between sub-populations may favor increased virulence[88].

At the host level, superspreader hosts or events may play a large role in the spread of a specific strain of a virus [115]. In such cases, the spreading success of a strain may be more related to host behavioral or physiological attributes than the fitness of that particular viral strain [98]. In highly structured livestock populations, for example, farms that ship high volumes of animals and

occupy central positions in animal transport networks can disproportionately contribute to spread of a particular strain regardless of the fitness displayed by that particular strain [116].

More generally, stochastic events may also be responsible for the apparent success of a given viral strain in a population [117]. Viral founder effects, population bottlenecks, and superspreading events, for example, may influence viral populations in manners not clearly related to viral fitness [118–120]. Depending on how many viral particles are transmitted between two individuals, the transmission event itself may introduce stochasticity (i.e., random founder effects) in determining which strains transmit and persist. For example, multiple introductions of SARS-CoV-2 in specific populations leads to, at least in the beginning, outbreaks of strains that just happened to be earlier introduced rather than outbreaks of particularly fit strains [121–123]. Alternatively, transmission between hosts or populations may represent a selective bottleneck wherein a variant's ability to be transmitted is mediated by characteristics of both the transmitter and recipient. Furthermore, the fitness of a particular variant is contextual and may not be the same within different hosts or populations, especially given hosts/populations vary immunologically, physiologically, behaviorally, and genetically (Box 2).

[Start of Box 2]

Box 2. Host genetic diversity

Host-level factors, such as host genetics, may influence variation in host-pathogen interactions [124]. How this genetic variation affects multi-strain pathogen dynamics is currently not understood. However, evidence from both theoretical and empirical studies point to a general pattern of greater host genetic diversity resulting in an increase in pathogen genetic diversity [125,126] and vice versa [127]. Multi-strain pathogens may thus be expected to naturally emerge in co-evolving host-pathogen systems. Indeed, numerous multi-strain modelling studies have demonstrated that host genetic diversity is an important determinant of pathogen evolution, strain emergence, and persistence [128–132]. The direction and degree of influence however depend on multiple factors, including the nature of host genetic variation (e.g. affecting host resistance or host infectivity [132]), population structure (e.g., well-mixed populations versus genetically distinct sub-populations;[130]), the genetic architecture underlying host genetic variation (e.g., single genes conferring complete or partial resistance vs polygenic effects represented by a continuous spectrum for resistance; e.g.,[128,132]), the existence and nature of trade-offs between pathogen virulence and transmissibility among different host genotypes [129–132], as well as on the within-host dynamics of the pathogen [131,132]. In particular, models with supporting empirical evidence predict that host genetic heterogeneity generally tends to increase the chance of stochastic extinction of emerging strains with low transmission potential (R_0<1) [128,132,133], thus reducing the risk of emergence and establishment of novel strains. However, Yates et al., [132] demonstrated that host heterogeneity could also lead to increased emergence and spread of novel pathogen strains, if these can adapt quickly to different host types. While this body of work highlights linkages between pathogen transmission dynamics and host genetic diversity at individual, population, or meta-population scales, more empirical studies on how multi-strain viral dynamics are modulated by genetically diverse host populations are needed.

[End of Box 2]

Outstanding questions

Numerous unresolved questions need to be addressed to understand multi-strain dynamics in different host-virus systems: a) With complex host immune responses and interaction with co-circulating strains, how does co-infection and co-evolution influence the effectiveness of disease management such as vaccination or other control strategies? b) Although we have described different phylodynamic tools useful for understanding genetic evolution of co-circulating strains, what are the best approaches to investigate and contextualize antigenic evolution in those strains? Additionally, are there distinct and measurable phylogenetic tree topologies characteristic of ecological multi-strain dynamics, and how do perturbations in host populations affect tree structure? c) Host genotypes may non-uniformly influence susceptibility to certain pathogens. How do these host differences affect multi-strain pathogen dynamics at the population level? d) Host populations may be stratified or sub-structured for many reasons (natural or artificial). Since strains theoretically evolve to balance transmissibility-virulence tradeoffs specific to a given sub-population, how do changes in host population structure affect the co-evolution/co-circulation of different strains in a population? e) How quickly and to what extent does the fitness of a particular strain vary between individual hosts and across space and time? What are the most suitable approaches to quantify and predict the role of viral fitness in the establishment of multiple strains in a population or sub-population? And can these tools be used to predict future success or invasion potential of different strains?

Concluding remarks

Although multi-strain dynamics are likely to occur in many rapidly evolving pathogens, the implications of immune-mediated competition amongst co-circulating strains for shaping spatiotemporal dynamics, maintenance of genetic diversity, and emergence of novel variants are often overlooked. However, such multi-strain dynamics are critical for predicting the invasion success of novel genetic variants and anticipating outcomes of vaccination programs. In this review, we synthesized the interacting ecological and evolutionary processes that constitute multi-strain dynamics. To predict sequential or cyclic dominance of different strains, it is essential to understand the interplay between population immunity and the emergence of novel strains, as well as to understand the ecological dynamics amongst co-circulating strains that interact via frequency-dependent fitness advantages related to partial cross-immunity. Even though the availability of sequence data has increasingly enabled studies of pathogen evolution and molecular ecology, examining the complex interactions occurring in multi-strain systems is challenging both theoretically and empirically. By highlighting the different components and scales of understanding multi-strain dynamic in viruses, we call attention to the need for more holistic studies in future. Methodological approaches are rapidly developing [117,134], with the evolution of SARS-CoV-2 variants now providing the quintessential exemplar of multi-strain dynamics (Box 3). However, there are many fundamental questions still to be answered to more fully understand the interplay between the immunology, evolution, and epidemiology of multi-strain pathogens. Whereas previous research has largely focused on human host-pathogen systems, such as influenza [1,5,109], dengue [3], and rotavirus [14], research on multi-strain dynamics in animal populations provides a rich area to further explore fundamental questions and generalizable insights for multi-strain pathogens [135]. Investigating these questions will improve our ability to anticipate the behavior of multi-strain pathogens.

[Start of Box 3]

Box 3 Multi-strain dynamics of SARS-CoV-2

The repeated emergence and spread of new variants during the SARS-CoV-2 pandemic has raised the prominence of research on multi-strain dynamic, leading to development, refinement, and integration of analytical approaches to better elucidate the interplay between immunology and evolution and their combined impact of the epidemiology of the disease. Near real-time tracking of genomic data from across the globe has revealed SARS-CoV-2 evolution and the relative frequency of different variants across different geographies [136]. In particular, the emergence of the alpha and beta variants of concern were detected in Europe and South Africa, respectively, with alpha establishing a foothold worldwide. Subsequently, the delta and omicron variants emerged and, due to changes in either transmissibility or antigenicity, successfully invaded host populations with high levels of immunity, demonstrating abilities to outcompete or evade immunity elicited by other variants on local, national, and global scales [137].

The fitness advantages of variants, for which phylogenetic clade growth is assumed to be a useful proxy, can be statistically modeled through approaches such as multinomial or logistic regression on the frequency of different variants [134], and suites of mutations have been found to correlate with clade growth [117,134]. In general, variants of concern are characterized by higher than expected numbers of mutations, particularly in the S1 domain of the spike protein – a region important for cell entry that is targeted by neutralizing antibodies [117]. In addition, the strength of selection, as measured by dN/dS ratios, increased dramatically after the first 12 months of the epidemic, likely as a result of immune-driven selection [117]. Data from within-host [138] and population levels [117,136] both show the repeated selection for certain mutations potentially associated with antibody evasion (amongst others).

Such mutations pose a concern for immune escape, motivating ongoing immunological studies to quantify the extent of cross-neutralization amongst variants and vaccines [139–142]. In parallel, mathematical models are being employed to assess the implications of the emergence of variants with phenotypic differences (i.e., transmissibility or immune escape) on projected epidemiological dynamics. For example, variants with enhanced transmissibility are likely of more concern compared to variants exhibiting partial immune escape, with the latter primarily increasing the numbers of mild breakthrough cases in vaccinated populations as opposed to enhancing epidemic severity [143]. The effectiveness of vaccination in controlling the epidemic is most limited when a variant displays both traits [143], particularly if escape mutants are allowed to evolve under immune pressure [144]. Taken together, studies of SARS-CoV-2 bring to fore the intricate interplay between host-pathogen systems and population immunity and are advancing our understanding of multi-strain dynamics.

[End of Box 3]

References

1. Gupta, S. Chaos, Persistence, and Evolution of Strain Structure in Antigenically Diverse Infectious Agents. *Science (80-.).* **280**, 912–915 (1998).
2. Kucharski, A. J., Andreasen, V. & Gog, J. R. Capturing the dynamics of pathogens with many strains. *J. Math. Biol.* **72**, 1–24 (2016).
3. Lourenço, J. & Recker, M. Natural, Persistent Oscillations in a Spatial Multi-Strain Disease System with Application to Dengue. *PLoS Comput. Biol.* **9**, (2013).
4. Gog, J. R. & Grenfell, B. T. Dynamics and selection of many-strain pathogens. *Proc. Natl. Acad. Sci. U. S. A.* **99**, (2002).
5. Recker, M., Pybus, O. G., Nee, S. & Gupta, S. 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**, (2007).
6. Jang, Y., Seo, T. & Seo, S. H. Higher virulence of swine H1N2 influenza viruses containing avian-origin HA and 2009 pandemic PA and NP in pigs and mice. *Arch. Virol.* **165**, 1141–1150 (2020).
7. Salvesen, H. A. & Whitelaw, C. B. A. Current and prospective control strategies of influenza A virus in swine. *Porc. Heal. Manag.* **7**, 23 (2021).
8. Ma, W., Kahn, R. E. & Richt, J. A. The pig as a mixing vessel for influenza viruses: Human and veterinary implications. *J. Mol. Genet. Med.* **03**, 158 (2009).
9. Mancera Gracia, J. C., Pearce, D. S., Masic, A. & Balasch, M. Influenza A Virus in Swine: Epidemiology, Challenges and Vaccination Strategies. *Front. Vet. Sci.* **7**, 647 (2020).
10. Van Regenmortel, M. H. V. Virus species and virus identification: Past and current controversies. *Infect. Genet. Evol.* **7**, 133–144 (2007).
11. Lazebnik, T., Bunimovich-Mendrazitsky, S. & Consortium, with the L. I. Generic Approach For Mathematical Model of Multi-Strain Pandemics. *bioRxiv* 2021.11.16.468823 (2021). doi:10.1101/2021.11.16.468823
12. Lazebnik, T. & Bunimovich-Mendrazitsky, S. Generic approach for mathematical model of multi-strain pandemics. *PLoS One* **17**, e0260683 (2022).
13. Wikramaratna, P. S., Sandeman, M., Recker, M. & Gupta, S. The antigenic evolution of influenza: Drift or thrift? *Philos. Trans. R. Soc. B Biol. Sci.* **368**, (2013).
14. Pitzer, V. E. *et al.* Modeling rotavirus strain dynamics in developed countries to understand the potential impact of vaccination on genotype distributions. *Proc. Natl. Acad. Sci. U. S. A.* **108**, (2011).
15. Grenfell, B. T. *et al.* Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. *Science* **303**, (2004).
16. Paploski, I. A. D. *et al.* Temporal Dynamics of Co-circulating Lineages of Porcine Reproductive and Respiratory Syndrome Virus. *Front. Microbiol.* **10**, 1–13 (2019).

17. Ferguson, N. M., Galvani, A. P. & Bush, R. M. Ecological and immunological determinants of influenza evolution. *Nature* **422**, 428–433 (2003).

18. Bishop, S. C., Axford, R. F. E., Nicholas, F. W. & Owen, J. B. *Breeding for disease resistance in farm animals*. *Breeding for disease resistance in farm animals: Third Edition* (CABI, 2010). doi:10.1079/9781845935559.0000

19. Volz, E. M., Koelle, K. & Bedford, T. Viral Phylodynamics. *PLoS Comput. Biol.* **9**, e1002947 (2013).

20. Paploski, I. A. D. *et al.* Phylogenetic Structure and Sequential Dominance of Sub-Lineages of PRRSV Type-2 Lineage 1 in the United States. *Vaccines* **9**, 608 (2021).

21. Poon, A. F. Y. *et al.* Mapping the shapes of phylogenetic trees from human and zoonotic RNA viruses. *PLoS One* **8**, (2013).

22. Domingo, E. & Schuster, P. What Is a Quasispecies? Historical Origins and Current Scope. in *Quasispecies: From Theory to Experimental Systems* (eds. Domingo, E. & Schuster, P.) 1–22 (Springer International Publishing, 2015). doi:10.1007/82_2015_453

23. Lythgoe, K. A., Gardner, A., Pybus, O. G. & Grove, J. Short-Sighted Virus Evolution and a Germline Hypothesis for Chronic Viral Infections. *Trends in Microbiology* **25**, (2017).

24. Chen, N., Trible, B. R., Kerrigan, M. A., Tian, K. & Rowland, R. R. R. ORF5 of porcine reproductive and respiratory syndrome virus (PRRSV) is a target of diversifying selection as infection progresses from acute infection to virus rebound. *Infect. Genet. Evol.* **40**, 167–175 (2016).

25. Carpenter, S. *Title: Identification of Genetic Mutations that Confer Escape from Innate or Adaptive Host Immune Responses During PRRSV Infection In Vivo-NPB #12-173 Investigator*. (2014).

26. Dimitrov, D. S. Virus entry: molecular mechanisms and biomedical applications. *Nat. Rev. Microbiol.* **2**, (2004).

27. Dou, D., Revol, R., Östbye, H., Wang, H. & Daniels, R. Influenza A virus cell entry, replication, virion assembly and movement. *Frontiers in Immunology* **9**, (2018).

28. Hamilton, B. S., Whittaker, G. R. & Daniel, S. Influenza virus-mediated membrane fusion: Determinants of hemagglutinin fusogenic activity and experimental approaches for assessing virus fusion. *Viruses* **4**, (2012).

29. Li, K. *et al.* Virus–Host Interactions in Foot-and-Mouth Disease Virus Infection. *Front. Immunol.* **12**, (2021).

30. Millet, J. K., Jaimes, J. A. & Whittaker, G. R. Molecular diversity of coronavirus host cell entry receptors. *FEMS Microbiol. Rev.* (2020). doi:10.1093/femsre/fuaa057

31. Wang, G., Wang, Y., Shang, Y., Zhang, Z. & Liu, X. How foot-and-mouth disease virus receptor mediates foot-and-mouth disease virus infection. *Virology Journal* **12**, (2015).

32. Sokol, C. L. & Luster, A. D. The chemokine system in innate immunity. *Cold Spring Harb. Perspect. Biol.* **7**, (2015).

33. Takeuchi, O. & Akira, S. Innate immunity to virus infection. *Immunological Reviews* **227**, (2009).

34. Theofilopoulos, A., Baccala, R., Beutler, B. & Kono, D. Type I interferons (alpha/beta) in immunity and autoimmunity. *Annu Rev Immunol* **23**, (2005).

35. Mueller, S. N. & Rouse, B. T. Immune responses to viruses. in *Clinical Immunology* (Elsevier, 2008). doi:10.1016/B978-0-323-04404-2.10027-2

36. Chen, X. *et al.* Host immune response to influenza A virus infection. *Frontiers in Immunology* **9**, (2018).

37. Agrawal, B. Heterologous Immunity: Role in Natural and Vaccine-Induced Resistance to Infections. *Frontiers in Immunology* **10**, (2019).

38. Sharma, S. & Thomas, P. G. The two faces of heterologous immunity: protection or immunopathology. *J. Leukoc. Biol.* **95**, (2014).

39. Spackman, E. & Sitaras, I. *Animal Influenza Virus*. **2123**, (Springer US, 2020).

40. Anderson, C. S., McCall, P. R., Stern, H. A., Yang, H. & Topham, D. J. Antigenic cartography of H1N1 influenza viruses using sequence-based antigenic distance calculation. *BMC Bioinformatics* **19**, (2018).

41. Cai, Z., Zhang, T. & Wan, X.-F. Concepts and applications for influenza antigenic cartography. *Influenza Other Respi. Viruses* **5 Suppl 1**, (2011).

42. Wang, P. *et al.* Predicting influenza antigenicity by matrix completion with antigen and antiserum similarity. *Front. Microbiol.* **9**, (2018).

43. Hirst, G. K. Studies of antigenic differences among strains of influenza by means of red cell agglutination. *J. Exp. Med.* **78**, (1943).

44. Kendra, J. A., Tohma, K., Ford-Siltz, L. A., Lepore, C. J. & Parra, G. I. Antigenic cartography reveals complexities of genetic determinants that lead to antigenic differences among pandemic GII.4 noroviruses. *Proc. Natl. Acad. Sci.* **118**, (2021).

45. Bell, S. M., Katzelnick, L. & Bedford, T. Dengue genetic divergence generates within-serotype antigenic variation, but serotypes dominate evolutionary dynamics. *Elife* **8**, (2019).

46. Yao, Y. *et al.* Predicting influenza antigenicity from Hemagglutintin sequence data based on a joint random forest method. **7**, (2017).

47. Zeller, M. A. *et al.* Machine Learning Prediction and Experimental Validation of Antigenic Drift in H3 Influenza A Viruses in Swine. *mSphere* **6**, (2021).

48. Wikramaratna, P. S. *et al.* Five challenges in modelling interacting strain dynamics. *Epidemics* **10**, (2015).

49. Elliott, P. *et al.* Exponential growth, high prevalence of SARS-CoV-2, and vaccine effectiveness associated with the Delta variant. *Science (80-.).* **374**, (2021).

50. Bianco, S., Shaw, L. B. & Schwartz, I. B. Epidemics with multistrain interactions: The

interplay between cross immunity and antibody-dependent enhancement. *Chaos* **19**, (2009).

51. Nickbakhsh, S. *et al.* Virus-virus interactions impact the population dynamics of influenza and the common cold. *Proc. Natl. Acad. Sci. U. S. A.* **116**, (2019).

52. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).

53. Lemey, P. *et al.* Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2. *PLoS Pathog.* **10**, e1003932 (2014).

54. Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian Phylogeography Finds Its Roots. *PLoS Comput. Biol.* **5**, e1000520 (2009).

55. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).

56. Suchard, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, 1–5 (2018).

57. Gill, M. S. *et al.* Improving bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013).

58. Kingman, J. F. C. On the genealogy of large populations. *J. Appl. Probab.* **19**, 27–43 (1982).

59. Griffiths, R. C. & Tavare, S. Ancestral Inference in Population Genetics. *Stat. Sci.* **9**, 307–319 (1994).

60. Magee, D., Suchard, M. A. & Scotch, M. Bayesian phylogeography of influenza A/H3N2 for the 2014-15 season in the United States using three frameworks of ancestral state reconstruction. *PLOS Comput. Biol.* **13**, e1005389 (2017).

61. Müller, N. F., Rasmussen, D. & Stadler, T. MASCOT: Parameter and state inference under the marginal structured coalescent approximation. *Bioinformatics* **34**, 3843–3848 (2018).

62. Kühnert, D., Stadler, T., Vaughan, T. G. & Drummond, A. J. Phylodynamics with Migration: A Computational Framework to Quantify Population Structure from Genomic Data. *Mol. Biol. Evol.* **33**, 2102–2116 (2016).

63. Yan, L., Neher, R. A. & Shraiman, B. I. Phylodynamic theory of persistence, extinction and speciation of rapidly adapting pathogens. *Elife* **8**, (2019).

64. Kistler, K. E. & Bedford, T. Evidence for adaptive evolution in the receptor-binding domain of seasonal coronaviruses OC43 and 229E. *Elife* **10**, (2021).

65. Bedford, T. *et al.* Integrating influenza antigenic dynamics with molecular evolution. *Elife* **2014**, (2014).

66. de Carvalho Ferreira, H. C. *et al.* An Integrative Analysis of Foot-and-Mouth Disease Virus Carriers in Vietnam Achieved Through Targeted Surveillance and Molecular

Epidemiology. *Transbound. Emerg. Dis.* **64**, (2017).

67. Huang, J. H. *et al.* Molecular characterization and phylogenetic analysis of dengue viruses imported into Taiwan during 2008-2010. *Am. J. Trop. Med. Hyg.* **87**, (2012).

68. Höckerstedt, L. M., Siren, J. P. & Laine, A.-L. Effect of spatial connectivity on host resistance in a highly fragmented natural pathosystem. *J. Evol. Biol.* **31**, 844–852 (2018).

69. Papaïx, J., Burdon, J. J., Lannou, C. & Thrall, P. H. Evolution of Pathogen Specialisation in a Host Metapopulation: Joint Effects of Host and Pathogen Dispersal. *PLoS Comput. Biol.* **10**, e1003633 (2014).

70. Tack, A. J. M., Hakala, J., Petäjä, T., Kulmala, M. & Laine, A.-L. Genotype and spatial structure shape pathogen dispersal and disease dynamics at small spatial scales. *Ecology* **95**, 703–714 (2014).

71. Smith, D. J. *et al.* Mapping the antigenic and genetic evolution of influenza virus. *Science (80-.).* **305**, (2004).

72. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, (1989).

73. Korneliussen, T. S., Moltke, I., Albrechtsen, A. & Nielsen, R. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics* **14**, (2013).

74. Wargo, A. R. & Kurath, G. Viral fitness: Definitions, measurement, and current insights. *Current Opinion in Virology* **2**, (2012).

75. Dayarian, A. & Shraiman, B. I. How to infer relative fitness from a sample of genomic sequences. *Genetics* **197**, (2014).

76. Neher, R. A., Russell, C. A. & Shraiman, B. I. Predicting evolution from the shape of genealogical trees. *Elife* **3**, (2014).

77. Doumayrou, J., Thébaud, G., Vuillaume, F., Peterschmitt, M. & Urbino, C. Mapping genetic determinants of viral traits with FST and quantitative trait locus (QTL) approaches. *Virology* **484**, 346–353 (2015).

78. Nagylaki, T. Fixation Indices in Subdivided Populations. *Genetics* **148**, 1325–1332 (1998).

79. Nei, M. & Chesser, R. K. Estimation of fixation indices and gene diversities. *Ann. Hum. Genet.* **47**, 253–259 (1983).

80. Yang, Z. & Nielsen, R. Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. *Mol. Biol. Evol.* **17**, 32–43 (2000).

81. Tubiana, L., Božič, A. L., Micheletti, C. & Podgornik, R. Synonymous Mutations Reduce Genome Compactness in Icosahedral ssRNA Viruses. *Biophys. J.* **108**, 194 (2015).

82. Van, A. *et al.* Impact of Synonymous Genome Recoding on the HIV Life Cycle. (2021). doi:10.3389/fmicb.2021.606087

83. Cuevas, J. M., Domingo-Calap, P. & Sanjuán, R. The Fitness Effects of Synonymous

Mutations in DNA and RNA Viruses. *Mol. Biol. Evol.* **29**, 17–20 (2012).

84. Kryazhimskiy, S. & Plotkin, J. B. The Population Genetics of dN/dS. *PLoS Genet.* **4**, (2008).

85. Kosakovsky Pond, S. L. & Frost, S. D. W. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Mol. Biol. Evol.* **22**, (2005).

86. Su, Y. C. F. *et al.* Phylodynamics of H1N1/2009 influenza reveals the transition from host adaptation to immune-driven selection. *Nat. Commun.* **6**, (2015).

87. Alizon, S., Hurford, A., Mideo, N. & Van Baalen, M. Virulence evolution and the trade-off hypothesis: History, current state of affairs and the future. *Journal of Evolutionary Biology* **22**, (2009).

88. Eshelman, C. M. *et al.* Unrestricted migration favours virulent pathogens in experimental metapopulations: evolutionary genetics of a rapacious life history. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 2503–2513 (2010).

89. Clay, P. A. & Rudolf, V. H. W. How parasite interaction strategies alter virulence evolution in multi-parasite communities. *Evolution (N. Y).* **73**, 2189–2203 (2019).

90. Kryazhimskiy, S., Dieckmann, U., Levin, S. A. & Dushoff, J. On state-space reduction in multi-strain pathogen models, with an application to antigenic drift in influenza A. *PLoS Comput. Biol.* **3**, (2007).

91. Peralta, R., Vargas-De-León, C., Cabrera, A. & Miramontes, P. Dynamics of high-risk nonvaccine human papillomavirus types after actual vaccination scheme. *Comput. Math. Methods Med.* **2014**, (2014).

92. Ranjeva, S. L. *et al.* Recurring infection with ecologically distinct HPV types can explain high prevalence and diversity. *Proc. Natl. Acad. Sci. U. S. A.* **114**, (2017).

93. Aguiar, M., Stollenwerk, N. & Kooi, B. W. The stochastic multi-strain dengue model: Analysis of the dynamics. in *AIP Conference Proceedings* **1389**, (2011).

94. Blower, S. M., Aschenbach, A. N., Gershengorn, H. B. & Kahn, J. O. Predicting the unpredictable: Transmission of drug-resistant HIV. *Nat. Med.* **7**, (2001).

95. Sharomi, O. & Gumel, A. B. Dynamical analysis of a multi-strain model of HIV in the presence of anti-retroviral drugs. *J. Biol. Dyn.* **2**, (2008).

96. Roche, B., Drake, J. M. & Rohani, P. An Agent-Based Model to study the epidemiological and evolutionary dynamics of Influenza viruses. *BMC Bioinformatics* **12**, (2011).

97. Sofonea, M. T., Alizon, S. & Michalakis, Y. From within-host interactions to epidemiological competition: a general model for multiple infections. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20140303 (2015).

98. VanderWaal, K. L. & Ezenwa, V. O. Heterogeneity in pathogen transmission: mechanisms and methodology. *Funct. Ecol.* **30**, 1606–1622 (2016).

99. Cobey, S. & Pascual, M. Consequences of host heterogeneity, epitope immunodominance, and immune breadth for strain competition. *J. Theor. Biol.* **270**,

(2011).

100. Aguiar, M., Ballesteros, S., Kooi, B. W. & Stollenwerk, N. The role of seasonality and import in a minimalistic multi-strain dengue model capturing differences between primary and secondary infections: Complex dynamics and its implications for data analysis. *J. Theor. Biol.* **289**, (2011).

101. Breban, R., Drake, J. M. & Rohani, P. A general multi-strain model with environmental transmission: Invasion conditions for the disease-free and endemic states. *J. Theor. Biol.* **264**, (2010).

102. Kamo, M. & Sasaki, A. The effect of cross-immunity and seasonal forcing in a multi-strain epidemic model. *Phys. D Nonlinear Phenom.* **165**, (2002).

103. Martcheva, M. A non-autonomous multi-strain SIS epidemic model. *J. Biol. Dyn.* **3**, (2009).

104. Pugliese, A. On the evolutionary coexistence of parasite strains. in *Mathematical Biosciences* **177**–**178**, (2002).

105. Roche, B. & Rohani, P. Environmental transmission scrambles coexistence patterns of avian influenza viruses. *Epidemics* **2**, (2010).

106. Korobeinikov, A. & Dempsey, C. A continuous phenotype space model of rna virus evolution within a host. *Math. Biosci. Eng.* **11**, (2014).

107. Castillo-Chavez, C., Hethcote, H. W., Andreasen, V., Levin, S. A. & Liu, W. M. Epidemiological models with age structure, proportionate mixing, and cross-immunity. *J. Math. Biol.* **27**, (1989).

108. Gupta, S., Swinton, J. & Anderson, R. M. Theoretical studies of the effects of heterogeneity in the parasite population on the transmission dynamics of malaria. *Proc. R. Soc. B Biol. Sci.* **256**, (1994).

109. Koelle, K., Khatri, P., Kamradt, M. & Kepler, T. B. A two-tiered model for simulating the ecological and evolutionary dynamics of rapidly evolving viruses, with an application to influenza. *J. R. Soc. Interface* **7**, (2010).

110. Lion, S. & Gandon, S. Spatial evolutionary epidemiology of spreading epidemics. *Proc. R. Soc. B Biol. Sci.* **283**, (2016).

111. Lange, A. & Ferguson, N. M. Antigenic diversity, transmission mechanisms, and the evolution of pathogens. *PLoS Comput. Biol.* **5**, (2009).

112. Pilosof, S. *et al.* Competition for hosts modulates vast antigenic diversity to generate persistent strain structure in Plasmodium falciparum. *PLOS Biol.* **17**, e3000336 (2019).

113. Lipsitch, M., Colijn, C., Cohen, T., Hanage, W. P. & Fraser, C. No coexistence for free: Neutral null models for multistrain pathogens. *Epidemics* **1**, (2009).

114. Read, J. M. & Keeling, M. J. Disease evolution on networks: the role of contact structure. *Proc. R. Soc. London. Ser. B Biol. Sci.* **270**, 699–708 (2003).

115. Adam, D. C. *et al.* Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat. Med.* **26**, 1714–1719 (2020).

116. Makau, D. N. *et al.* Integrating Animal Movements With Phylogeography to Model the Spread of PRRS Virus in the U.S. *Virus Evol.* (2021). doi:10.1093/ve/veab060

117. Kistler, K. E., Huddleston, J. & Bedford, T. Rapid and parallel adaptive mutations in spike S1 drive clade success in SARS-CoV-2. *bioRxiv* 2021.09.11.459844 (2021). doi:10.1101/2021.09.11.459844

118. Li, H. & Roossinck, M. J. Genetic Bottlenecks Reduce Population Variation in an Experimental RNA Virus Population. *J. Virol.* **78**, 10582–10587 (2004).

119. McCrone, J. T. *et al.* Stochastic processes constrain the within and between host evolution of influenza virus. *Elife* **7**, (2018).

120. Nelson, M. I. *et al.* Stochastic Processes Are Key Determinants of Short-Term Evolution in Influenza A Virus. *PLoS Pathog.* **2**, e125 (2006).

121. Deng, X. *et al.* Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California. *Science (80-.).* **369**, 582–587 (2020).

122. da Silva Filipe, A. *et al.* Genomic epidemiology reveals multiple introductions of SARS-CoV-2 from mainland Europe into Scotland. *Nat. Microbiol.* **6**, 112–122 (2021).

123. Tayoun, A. A. *et al.* Multiple early introductions of SARS-CoV-2 into a global travel hub in the Middle East. *Sci. Rep.* **10**, 17720 (2020).

124. Bishop, S. C., Doeschl-Wilson, A. B. & Woolliams, J. A. Uses and Implications of Field Disease Data for Livestock Genomic and Genetics Studies. *Front. Genet.* **3**, (2012).

125. Rodríguez-Nevado, C., Lam, T. T. Y., Holmes, E. C. & Pagán, I. The impact of host genetic diversity on virus evolution and emergence. *Ecol. Lett.* **21**, 253–263 (2018).

126. Schulte, R. D., Makus, C. & Schulenburg, H. Host-parasite coevolution favours parasite genetic diversity and horizontal gene transfer. *J. Evol. Biol.* **26**, 1836–1840 (2013).

127. Duxbury, E. M. L. *et al.* Host-pathogen coevolution increases genetic variation in susceptibility to infection. *Elife* **8**, (2019).

128. Chabas, H. *et al.* Evolutionary emergence of infectious diseases in heterogeneous host populations. *PLOS Biol.* **16**, e2006738 (2018).

129. Ganusov, V. V., Bergstrom, C. T. & Antia, R. Within-host population dynamics and the evolution of microparasites in a heterogeneous host population. *Evolution (N. Y).* **56**, 213–223 (2002).

130. González, R., Butković, A. & Elena, S. F. Role of host genetic diversity for susceptibility-to-infection in the evolution of virulence of a plant virus†. *Virus Evol.* **5**, (2019).

131. Regoes, R. R., Nowak, M. A. & Bonhoeffer, S. Evolution of virulence in a heterogeneous host population. *Evolution (N. Y).* **54**, 64–71 (2000).

132. Yates, A., Antia, R. & Regoes, R. R. How do pathogen evolution and host heterogeneity

interact in disease emergence? *Proc. R. Soc. B Biol. Sci.* **273**, 3075–3083 (2006).

133. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005).

134. Obermeyer, F. *et al.* Analysis of 2.1 million SARS-CoV-2 genomes identifies mutations associated with transmissibility. *medRxiv* 2021.09.07.21263228 (2021). doi:10.1101/2021.09.07.21263228

135. Wikramaratna, P. S., Pybus, O. G. & Gupta, S. Contact between bird species of different lifespans can promote the emergence of highly pathogenic avian influenza strains. *Proc. Natl. Acad. Sci. U. S. A.* **111**, (2014).

136. Rochman, N. D. *et al.* Ongoing global and regional adaptive evolution of SARS-CoV-2. *Proc. Natl. Acad. Sci.* **118**, e2104241118 (2021).

137. Volz, E. *et al.* Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **593**, 266–269 (2021).

138. Choi, B. *et al.* Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *N. Engl. J. Med.* **383**, 2291–2293 (2020).

139. Gidari, A. *et al.* Cross-neutralization of SARS-CoV-2 B.1.1.7 and P.1 variants in vaccinated, convalescent and P.1 infected. *J. Infect.* **83**, 467–472 (2021).

140. Changrob, S. *et al.* Cross-Neutralization of Emerging SARS-CoV-2 Variants of Concern by Antibodies Targeting Distinct Epitopes on Spike. *MBio* (2021). doi:10.1128/mBio.02975-21

141. Vidal, S. J. *et al.* Correlates of Neutralization against SARS-CoV-2 Variants of Concern by Early Pandemic Sera. *J. Virol.* **95**, (2021).

142. Muik, A. *et al.* Neutralization of SARS-CoV-2 lineage B.1.1.7 pseudovirus by BNT162b2 vaccine–elicited human sera. *Science (80-.).* **371**, (2021).

143. Bushman, M., Kahn, R., Taylor, B. P., Lipsitch, M. & Hanage, W. P. Population impact of SARS-CoV-2 variants with enhanced transmissibility and/or partial immune escape. *medRxiv* 2021.08.26.21262579 (2021). doi:10.1101/2021.08.26.21262579

144. Koopman, J. S., Simon, C. P., Getz, W. M. & Salter, R. Modeling the population effects of escape mutations in SARS-CoV-2 to guide vaccination strategies. *Epidemics* **36**, 100484 (2021).

Correspondence

All correspondence related to this manuscript shall be directed to the corresponding author Kimberly VanderWaal (kvw@umn.edu).

Author Contributions

K.V.W. and S.L. conceived the idea to elucidate antigenic evolution in multi-strain dynamics in viral-host systems, wrote different portions of the manuscript and developed conceptual figures to illustrate the concept. D.N.M. compiled all relevant literature, wrote part of the manuscript and coordinated the logical flow of the manuscript. M.M.S. summarized literature and wrote on mathematical modelling for multi-strain dynamics. I.A.D.P. and A.D.W. summarized the literature in Box2, M.C.J.C. and D.C.S provided insights on viral-host interaction and immune responses and assisted with writing portions of the manuscript. R.R.K. and M.E.C. summarized concepts on ecological and host population structures and assisted with writing portions of the manuscript. All authors were involved in the review and revision of the manuscript.

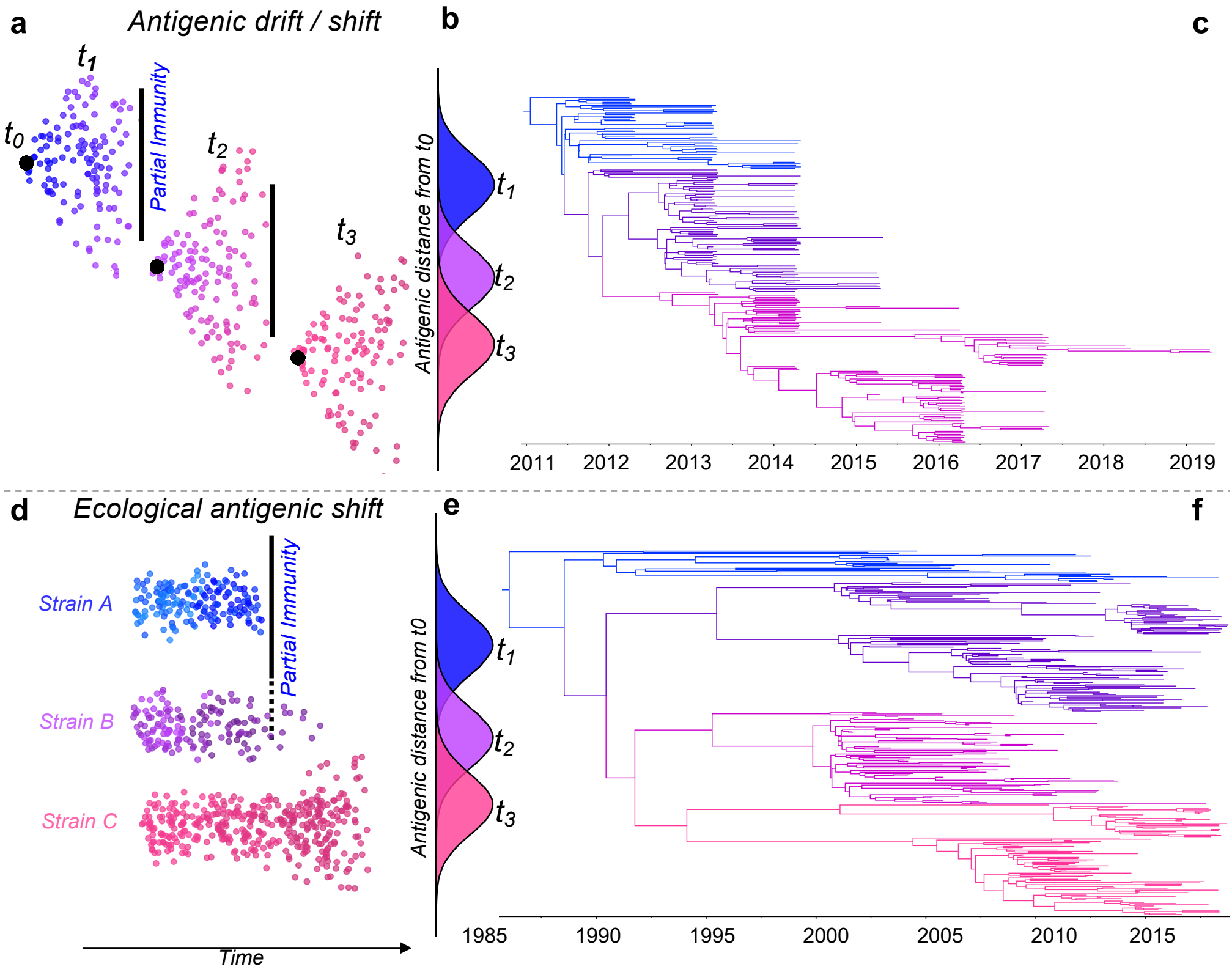
Acknowledgements

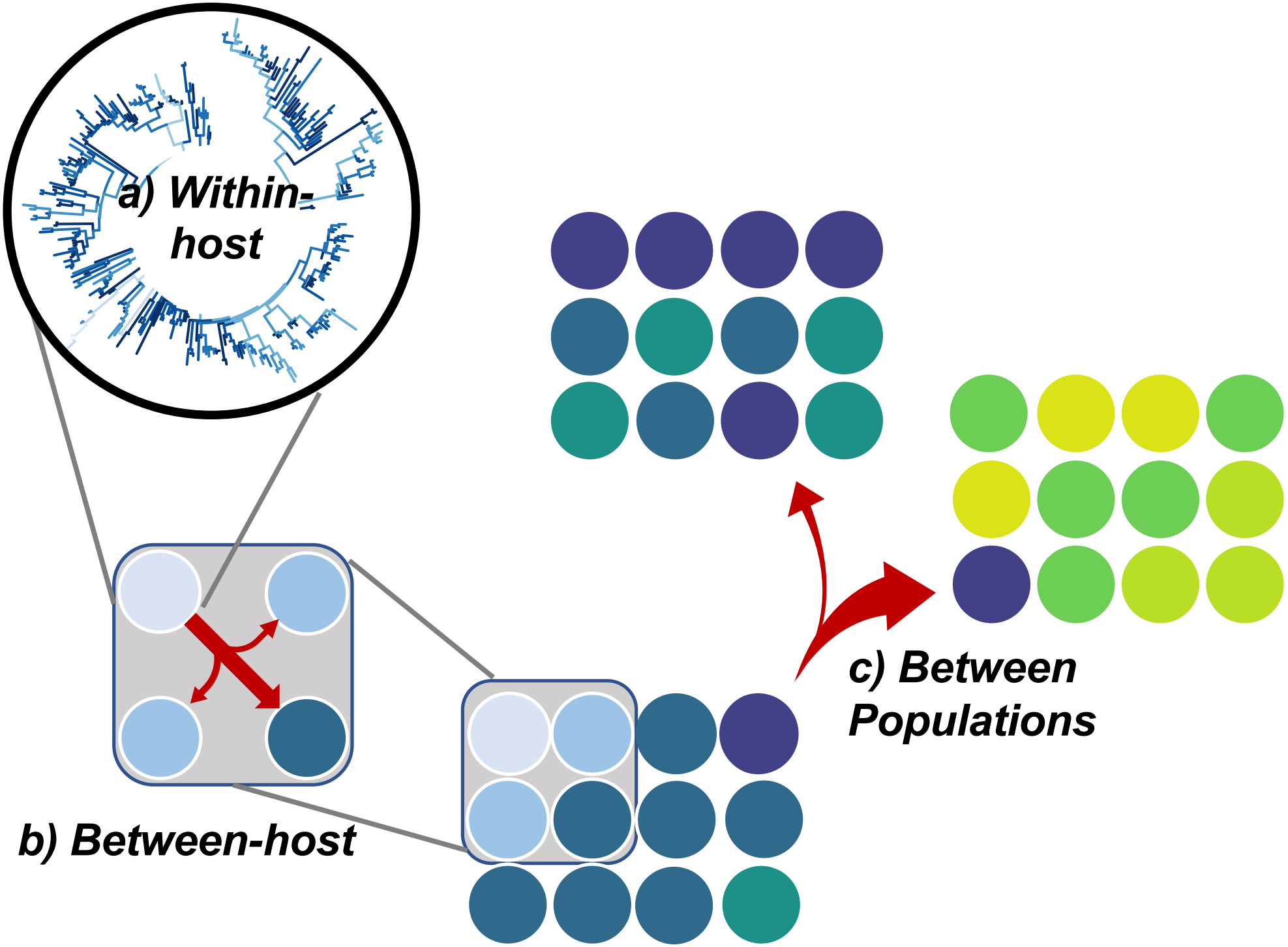
Funding was provided by the joint US-UK NIFA-NSF- NIH-BBSRC Ecology and Evolution of Infectious Disease awards 2019-67015-29918 and BB/T004401/1. This work was also supported by the USDA National Institute of Food and Agriculture, Animal Health project #MINV-62-057.

Authors also deeply thank all (current and former) members of the EEID project team for their time and constructive contributions to discussions that enriched this manuscript.

Competing Interests

The authors declare no competing interests.


a
Antigenic drift / shift
t_1
t_0
t_2
t_3
Partial Immunity
b
Antigenic distance from t0
t_1
t_2
t_3
c
2011
2012
2013
2014
2015
2016
2017
2018
2019
d
Ecological antigenic shift
Strain A
Strain B
Strain C
Partial Immunity
Time
e
Antigenic distance from t0
t_1
t_2
t_3
f
1985
1990
1995
2000
2005
2010
2015


a) Within-host
b) Between-host
c) Between Populations