communications physics

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42005-025-02406-5

Genotype networks drive oscillating endemicity and epidemic trajectories in viral evolution

Santiago Lamata-Otín [1,2], Octavian C. Rotita-Ion [1], Alex Arenas [3,4], David Soriano-Paños [2,3,6] ✉ & Jesús Gómez-Gardeñes [1,2,5,6] ✉

Rapidly evolving viruses use antigenic drift as a key mechanism to evade host immunity and persist in real populations. While traditional models of antigenic drift and epidemic spread rely on low-dimensional antigenic spaces, genomic surveillance data reveal that viral evolution produces complex antigenic genotype networks with hierarchical modular structures. In this study, we present an eco-evolutionary framework in which viral evolution and population immunity dynamics are shaped by the structure of antigenic genotype networks. Using synthetic networks, we demonstrate that network topology alone can drive transitions between stable endemic states and recurrent seasonal epidemics. Furthermore, our results show how the integration of the genotype network of the H3N2 influenza in our model allows for estimating the emergence times of various haplotypes resulting from its evolution. Our findings underscore the critical role of the topology of genotype networks in shaping epidemic behavior and, besides, provide a robust framework for integrating real-world genomic data into predictive epidemic models.

Epidemic trajectories of long-lasting viruses exhibit a wide range of behaviors and encompass diverse variant landscapes, as observed in SARS-CoV-2[1–4], influenza[5–11], dengue[12], or rabies[13] to name a few. This diversity arises from the intricate interplay between epidemic dynamics and viral evolution, which operate on compatible time scales[14]. In particular, *RNA* viruses evolve rapidly within the strain space, acquiring mutations that enable them to evade detection by the host immune system[12,15–17]. Notably, antigenic escape is the primary mechanism by which rapidly evolving viruses establish endemicity[18]. This occurs because mutations in antibody-binding regions not only allow the virus to persist within a host, evading its immune response, but also enable reinfections of previously recovered individuals[8,19–21].

Despite the high-dimensionality of antigenic spaces, genomic surveillance data revealed how some *RNA* viruses, such as Influenza A, have followed low-dimensional evolutionary trajectories[8]. Bedford et al.[22,23] showed by extensive simulations that the immune pressure existing in the population canalizes the complex evolution of the virus into a traveling wave moving across a quasi-1D path. This phenomenon greatly simplifies the analytical characterization of the evolutionary trajectories, as assuming one-dimensional antigenic spaces allows predicting both the shape and the speed of the traveling wave[24,25]. This traveling wave also appears in the evolution of non-antigenic traits with fixed selection coefficients[26,27]. Despite the success of one-dimensional approaches, these antigenic spaces cannot account for speciation events, a key aspect to understand the emergence of two Influenza B lineages or the sustained circulation of four Dengue serotypes. Along this line, recent theoretical works[28,29] show that these events typically occur in multidimensional antigenic spaces when mutations induce antigenic changes beyond the typical cross-immunity range. Therefore, the long-term behavior of epidemics not only hinges on the total number of mutations accumulated in antibody-binding regions but also on how the virus explores effectively the antigenic space.

Among the vast number of possible genomic mutations, only a small fraction of genotypes are sampled and recorded in public databases[30]. This is due to two primary factors: not all genetic sequences successfully overcome intra-host selection to emerge as viable (and thus observable) variants, and not all circulating strains are sequenced and registered. The traditional approach in epidemiology for tracking genotype evolution relies on phylogenetic trees[31,32], which infer via probabilistic models viral relationships based on large datasets of genetic sequences. There, different genotypes are associated to different lineages, all coming from the original wild type. This methodology results in a tree-like structure where each genotype comes

[1]Department of Condensed Matter Physics, University of Zaragoza, Zaragoza, Spain. [2]GOTHAM lab, Institute of Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Zaragoza, Spain. [3]Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Tarragona, Spain. [4]Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA, USA. [5]Center for Computational Social Science, University of Kobe, Kobe, Japan. [6]These authors jointly supervised this work: David Soriano-Paños, Jesús Gómez-Gardeñes. ✉e-mail: sorianopanos@gmail.com; gardenes@unizar.es

from a unique genotype and leads to a number of offspring genotypes. While phylogenetic methods provide valuable insights into viral evolution, they present significant limitations when modeling antigenic space. Specifically, phylogenetic reconstructions do not allow loops and thus do not account for the possibility of multiple evolutionary pathways leading to the same strain. This issue has been typically overlooked, as these loops are hardly observed due to the large dimensionality of the underlying genetic space[8]. Indeed, recent empirical works show that the probability of observing parallel evolutionary pathways decreases nearly exponentially with the loop length[33]. This has led previous works, such as Rouzine and Rozhnova[24], to consider tree-like antigenic topologies that explicitly reflect the absence of loops. Despite the complexity of these antigenic spaces, their simulations also retrieve the aforementioned canalization of the antigenic trajectories in quasi-1D paths.

As an alternative, genotype networks[34–38] offer a framework that captures these concurrent evolutionary trajectories, bridging the gap between real-world genomic diversity and its representation in antigenic models. In this approach, complex networks[39,40] are used to describe genetic sequences as nodes, with links connecting sequences that differ by a single point mutation. Through a detailed modeling process, Williams et al.[41] constructed a genotype network for the highly antigenic HA protein of influenza A (H3N2). Their analysis revealed the emergence of non-trivial topological properties consistent with generative models based on linear preferential attachment, highlighting fundamental differences between real genotype networks and models considering simple antigenic space embeddings. Moreover, the emergence of hierarchical modular genotype networks following viral evolution has also been recently reported for viruses affecting other hosts, such as $Q\beta$ bacteriophages[33]. The inherent complexity of antigenic space calls for the incorporation of genotype networks in eco-evolutionary frameworks to improve our understanding of how antigenic drift shapes epidemic trajectories.

Here we present a minimal multi-strain epidemiological model that capture contagion, immune dynamics, and mutation through antigenic genotype networks. This model builds upon previous multi-strain frameworks[24,28,29,42–50] but reduces the unnecessary proliferation of epidemiological states by mapping immunity acquisition for each strain in the shortening of their associated infectious period. In this framework, the genotype network governs how the virus mutates across the genotype space and determines how hosts infected by one strain build partial immunity to other variants with closely related antigenic properties. Furthermore, our model accounts for the infection history of the population by introducing a memory term. This feature represents a significant improvement over previous Markovian models[41,48], considering genotype networks, as these models assume that the immune response is only determined by the last variant contracted by the host.

With our evolutionary epidemic model in place, we: (i) recover antigenic escape dynamics consistent with previous studies, (ii) demonstrate that the topology of the genotype network alone can determine whether endemic trajectories consist of persistent seasonal waves or evolve into steady dynamics, (iii) uncover the complementary roles of mutant swarms and cross-immunity in sustaining infections, and (iv) provide a fair reconstruction of epidemic trajectories on real-world genotype networks. We round off the manuscript by discussing the implications of the former results and the research avenues that can be addressed in the future.

Results

Evolutionary epidemic model

Modeling the interplay between contagion, immune response, and mutation dynamics presents a significant challenge, as these processes operate at distinct yet compatible spatial and temporal scales. Particularly, contagion arises from interactions between individuals in a population, immune responses occur within the host through virus-antibody interactions, and mutation is governed by the internal dynamics of viral populations.

Here, we introduce the Susceptible - Infectious - Mutation - Susceptible (SIMS) model, a compartmental framework that integrates all three processes into a unified approach. In the SIMS model, each individual belongs to either the Susceptible (S) compartment or one of n distinct Infectious (I_i) compartments, where each I_i represents an infection with a strain $s_i \in \mathcal{N}$. Our model considers that intra-host viral dynamics occur on a much faster timescale than epidemiological spread. Consequently, the model neglects intra-host viral diversity and assumes that the infectious state of each individual is defined by a single strain.

The structure of the SIMS model is illustrated in Fig. 1, and its main processes are described as follows:

Contagion. As shown in Fig. 1a, susceptible (S) individuals contract an infection upon direct contact with Infectious (I_i) individuals infected with strain s_i, which has an associated infectivity rate β. Upon infection, an S individual transitions to the I_i compartment, acquiring the viral genotype s_i of the infecting host. We also introduce a recovery rate $\mu_i(t)$ governing how individuals infected by strain i transit to the Susceptible state at time t.

Immune response. Unlike other epidemic frameworks, the recovery rate μ_i is not constant over time in the SIMS model. Instead, the recovery rates $\{\mu_i(t)\}$ store the infection history of the population. In particular, we assume μ_i to represent the immune response built in the population as a response to the proliferation of strain i (see Fig. 1b). Namely, there is a disease-free value μ_0 corresponding to the baseline immune response existing in the population. Infection by strain i enhances host immunity against that strain at a given rate α. Moreover, individuals gain partial cross-immunity to genetically similar strains in the genotype network, with a characteristic cross-immunity length Δ. Acquired immunity wanes over time with a characteristic decay rate γ.

Mutation. Viral evolution (see Fig. 1c) is modeled as a diffusion process through the genotype network. Specifically, an infected individual associated with strain $s_i \in \mathcal{N}$ can mutate to a neighboring strain s_j at a rate D_x. When this mutation happens, the individual transitions from I_i to I_j in the compartmental dynamics (see Fig. 1a), creating a feedback loop between mutation and contagion dynamics.

In summary, the SIMS model consists of $n+1$ compartments and is governed by six epidemiological parameters (see Supplementary Table I in Supplementary Note 1), capturing the interplay between contagion, immune dynamics, and mutation in a genetically diverse viral population. The dynamics of the SIMS model can be captured by the set of coupled differential Eqs. (7)–(8), described in Methods. These equations yield the temporal evolution of the fraction of population infected by each strain i, $\rho_i(t)$, and the global prevalence of the disease, $I(t)$ (defined in Eq. (10)), hereinafter used as the principal metrics to characterize our epidemic trajectories.

Supplementary Fig. 1.1a in Supplementary Note 1 shows that the SIMS model for a single strain represents a versatile framework that allows reproducing the epidemic trajectories generated by multiple standard compartmental models. Namely, SIS-like dynamics are generated when neglecting the stimulation of the immune response, i.e., $\alpha = 0$, whereas SIR-like (SIRS-like) dynamics occur when the disease confers long-lasting (temporal) immune memory, i.e., $\gamma = 0$ ($\gamma \neq 0$) with $\alpha \neq 0$. Subsequently, in Supplementary Fig. 1.1b, c, we extend the analysis to multi-strain dynamics connected through a linear chain in the antigenic space. Considering long-lasting immune memory, i.e., $\gamma = 0$, the model retrieves well-known phenomenology reported by previous eco-evolutionary frameworks assuming low-dimensional antigenic spaces[24]. Namely, we observe SIRS-like trajectories generated by the antigenic drift of the virus through a traveling-wave solution across the antigenic space. To round off the description of the SIMS model, in the Supplementary Note 1, we derive analytically the expression for both the stationary prevalence (further analyzed in Supplementary Fig. 1.2) and the epidemic threshold. Likewise, in Supplementary Figs. 2 and 3 (in Supplementary Notes 2 and 3, respectively) we explore regions of the parameters space yielding unexpected dynamical regimes such as genotype re-emergence and highly fluctuating and aperiodic transitory dynamics.

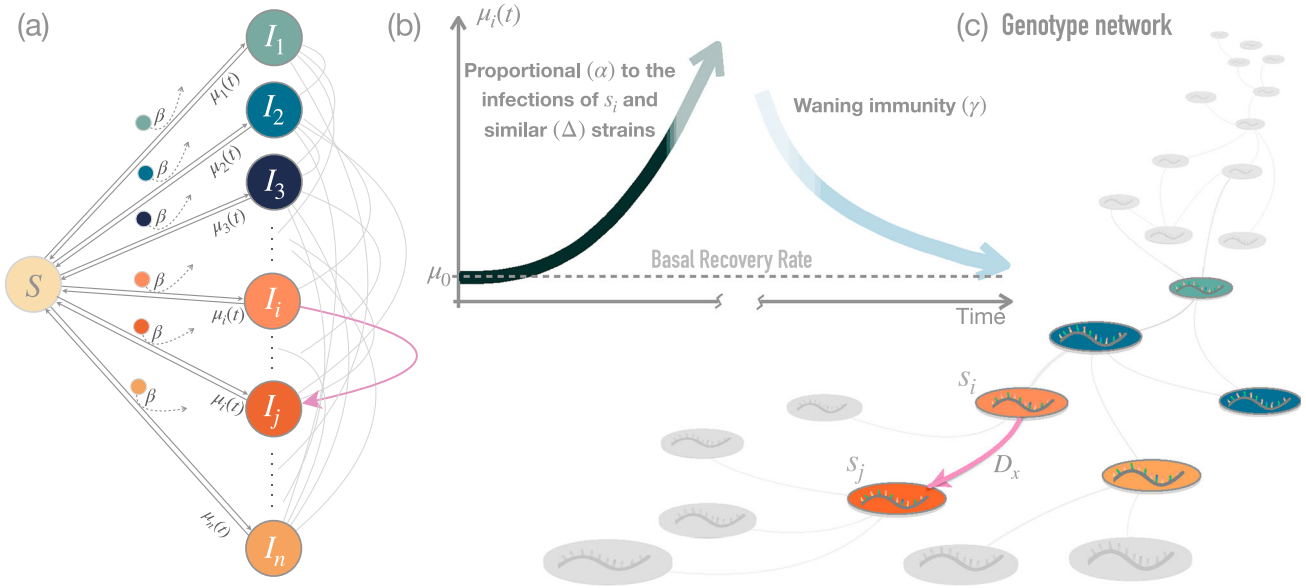

Fig. 1 | Schematic illustration of the SIMS dynamical model. Panel **a** depicts the compartmental structure, where Susceptible individuals (S) can transition to one of n Infectious compartments (I_i), each corresponding to a distinct strain s_i ($i = 1, \ldots, n$). Infection occurs at a rate β upon contact with an infectious individual carrying strain s_i. Infected individuals recover and return to the Susceptible compartment at a strain-dependent rate μ_i. Panel **b** illustrates the adaptive immune response at the population level: the recovery rate μ_i increases proportionally to the global immune pressure against strain s_i and antigenically similar strains, with proximity in the genotype network modulated by the parameter Δ. Counteracting this growth, immunity wanes over time at a rate γ, leading to a gradual reduction in μ_i toward the basal recovery rate μ_0. Panel **c** represents viral evolution through mutation. Infectious individuals carrying strain s_i can transition to an adjacent strain s_j in the genotype network at a mutation rate D_x, as indicated by the colored links in panels (**a**, **c**). This mechanism establishes a feedback loop between contagion and mutation dynamics.

The structure of the antigenic space shapes epidemic trajectories

Once presented with the SIMS model, we are interested in exploring how the complex structure of genotype networks governs epidemic dynamics. To tackle this question, we should (i) decipher what makes real genotype networks different from low-dimensional or randomized representations of the antigenic space and (ii) understand how these features alter the behavior of epidemic trajectories.

The Influenza A genotype network. The influenza A (H3N2) genotype network[41] consists of multiple connected components, with eight of them containing at least 130 nodes. Fig. 2a shows the second-largest connected component, which exhibits a rich and complex topology. For example, its degree distribution, displayed in Fig. 2b, follows a long-tailed pattern. This feature was previously noted by ref. 41, when analyzing the largest connected component, which also shows stable clustering and negative assortativity, two structural characteristics that align with generative models based on linear preferential attachment. These observations suggest the presence of non-trivial processes governing the growth and evolution of genotype networks.

To further investigate whether these structural features are purely a consequence of the connectivity distribution, we conduct a comparative analysis between the real genotype network and its randomized counterparts. Specifically, in Fig. 2c, we examine the most relevant structural metrics of the second-largest connected component and compare them to an ensemble of 100 degree-preserving randomized networks (see Methods for details). A similar analysis for the remaining components can be found in Supplementary Fig. 4 in Supplementary Note 4.

Our results reveal substantial differences between real and randomized genotype networks. First, real genotype networks exhibit longer average path lengths than their randomized counterparts, indicating a more intricate connectivity structure. Second, higher modularity values in the real network suggest the presence of mesoscale organization, where clusters of antigenically similar genotypes, commonly referred to as mutant *swarms*[33], could correspond to evolutionary lineages. Third, elevated clustering coefficients (with lower transitivity compared to randomized networks) suggest a higher likelihood of closed triangular structures, pinpointing the presence of a hierarchical organization of the networked backbone. Finally, the genotype network exhibits disassortative mixing, meaning that highly connected genotypes tend to link with those of lower connectivity. This disassortativity is consistent with the evolutionary process in which successful strains generate multiple antigenically similar offspring.

Epidemic trajectories on synthetic genotype networks. As discussed above, the H3N2 genotype network is disassortative, presents a wide range of connectivities, and a marked modular structure. To address the impact of the complex structure of the genotype network on the epidemic trajectories, we first construct a minimal synthetic network presenting these features. In particular, we model genotype networks as a concatenation of star-like clusters (mimicking the so-called mutant swarms), where an intermediate node connects the leaves of each pair of consecutive stars (see the blue and orange structures in Fig. 3a). Regarding the epidemiological parameters, throughout the manuscript we set $\mu_0^{-1} =$ 10 days and $\beta = 0.3$, yielding a basic reproduction number $\mathcal{R}_0 = 3$ (see Eq. (12) in Methods). For the acquisition of immune response against the variants, we use an immune production rate of $\alpha = 0.03$ and a cross-immunity length of $\Delta = 3$. More details on the choice of these parameters can be found in Methods.

In the presence of long-lasting immune memory, i.e., $\gamma = 0$, the constructed synthetic genotype networks produce epidemic trajectories characterized by a series of periodic waves that lead the system to a stationary pattern of seasonal cycles. When the mutant swarms are composed of the same number of leaves (blue curve in Fig. 3b), the oscillations are regular, whereas heterogeneous mutant swarms produce variability in both size and shape of the individual outbreak (orange curve in Fig. 3b). In particular, the size of a mutant swarm is proportional to the magnitude of the

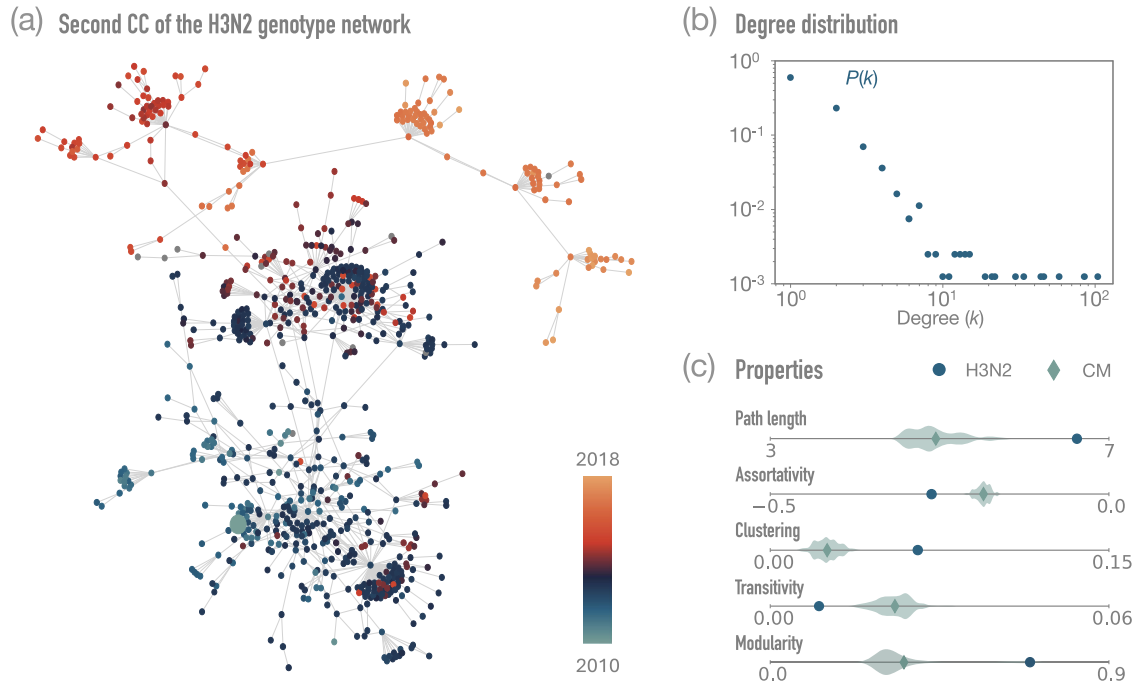

Fig. 2 | Structural properties of genotype networks. a Second connected component of the Influenza A genotype network[41], where the color-code represents the sampling year. **b** Degree distribution of the second component of the Influenza A genotype network. **c** The blue dots correspond to the values of some structural properties of the second component of the Influenza A genotype network, while the violin plot (labeled as configurational model, CM) corresponds to the values obtained for an ensemble of 100 networks sharing the same degree sequence. More details on the crafting can be found in the Methods.

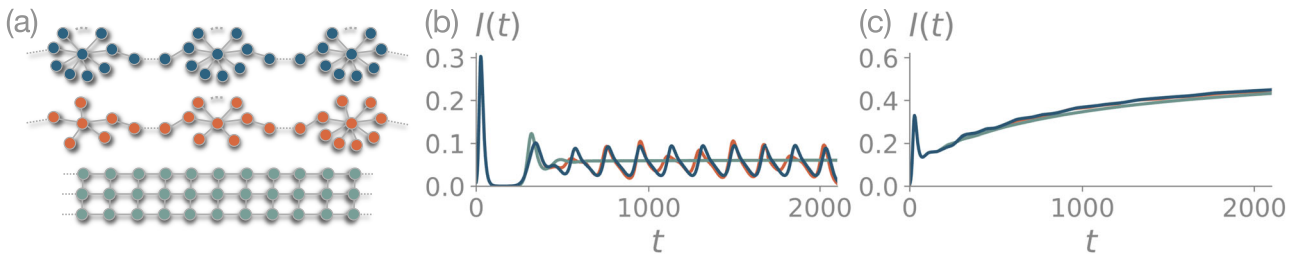

Fig. 3 | Steady versus seasonal endemicity. a Three synthetic genotype networks: a lattice, a homogeneous concatenation of star-like swarms and a heterogeneous concatenation of star-like swarms (see Supplementary Notes 6 and 7 for details on the structures). **b** Prevalence of the epidemic trajectories ($I(t)$) with the memory mechanism activated (waning immunity rate $\gamma = 0$). **c** Prevalence of the epidemic trajectories ($I(t)$) when the time scales governing immune acquisition and loss are compatible ($\gamma = \alpha$). In all cases, the infectivity rate $\beta = 0.3$, the basal recovery rate $\mu_0 = 0.1$, the immunity acquisition rate $\alpha = 0.03$, the mutation rate $D_x = 10^{-5}$, and the characteristic cross-immunity length $\Delta = 3$. Note that the colors used in panel (**a**) for each of the structures correspond to each of the curves in panels (**b**, **c**).

corresponding outbreak peak. This suggests that the emergence of mutant swarms around successful genotypes may serve to counteract cross-immunity pressure, as the simultaneous appearance of similar genotypes leads to joint outbreaks with higher overall prevalence. For further details on the trade-off between cross-immunity and star-like swarms, we refer the reader to the Supplementary Fig. 5 in Supplementary Note 5.

To round off our analysis on synthetic networks, we now introduce a lattice network in the antigenic spaces, resembling the traditional low-dimensional representations of the antigenic space. In that case, we observe how the epidemic quickly converges to a steady endemic equilibrium (turquoise curve in Fig. 3b), retrieving the well-reported results in the literature[24,25]. A complementary spectral analysis confirming the differences between trajectories is conducted in Supplementary Fig. 6 in Supplementary Note 6.

Interestingly, when time scales of immune acquisition and loss are compatible $\alpha = \gamma$, Fig. 3c shows that the epidemic trajectories for the three structures (lattice, concatenation of homogeneous swarms and concatenation of heterogeneous swarms) become macroscopically indistinguishable. Moreover, Supplementary Fig. 7 in Supplementary Note 7 shows that SIS-like phenomenology is retrieved for the three structures if immune response dynamics are neglected. Together, our findings thus confirm that the infection history of individuals, along with the complex structure of the antigenic space, are responsible for the complex epidemic trajectories observed at the population level.

Epidemic trajectories on real-world genotype networks. After demonstrating that synthetic genotype networks can shape epidemic trajectories and the role of mutant swarms, we now turn our attention to the Influenza A genotype network constructed by ref. 41. To follow the virus's evolution through antigenic space, we identified the most predominant mutant swarms in each connected component of the network using community detection (see Methods). For example, Fig. 4a displays the largest connected component of the Influenza A network, with nodes in its seven communities colored differently. We set $\gamma = 0$ hereafter, as memory B cells persist for a long time in this class of viruses. The remaining parameter choices are detailed in the Methods section and in the figure captions.

In Fig. 4b, the epidemic trajectory $I(t)$ is shown for this giant component, with the first sampled genotype serving as the wild type. The colored segments under the prevalence curve indicate the fraction of the population

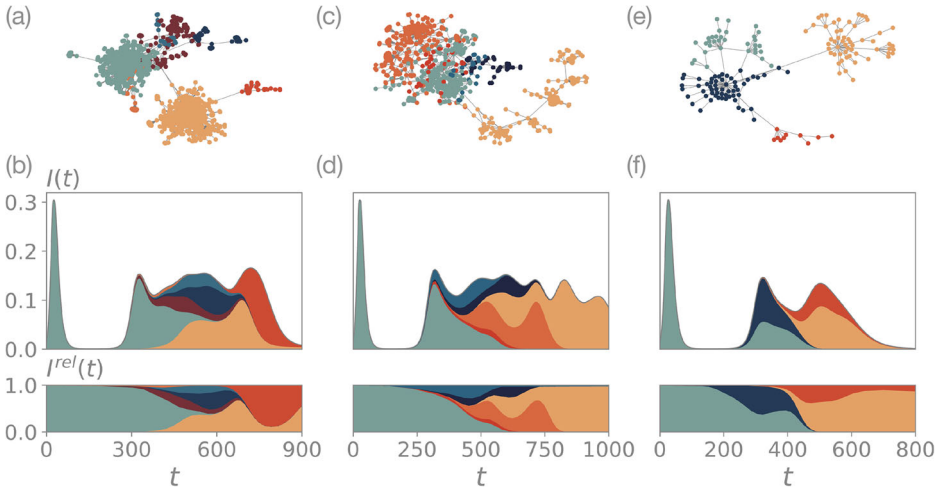

Fig. 4 | Epidemic trajectories in real-world genotype networks. Graphical representations and absolute ($I(t)$) and relative ($I^{rel}(t)$) prevalence of the epidemic trajectories for the first (**a**, **b**), second (**c**, **d**), and seventh (**e**, **f**) largest connected components of the INFV A (H3N2) network[41]. The colors correspond to the mutant swarms of nodes on each of the structures. In all panels, the infectivity rate $\beta = 0.3$, the basal recovery rate $\mu_0 = 0.1$, the immunity acquisition rate $\alpha = 0.03$, the waning immunity rate $\gamma = 0$, the mutation rate $D_x = 10^{-5}$, and the characteristic cross-immunity length $\Delta = 3$.

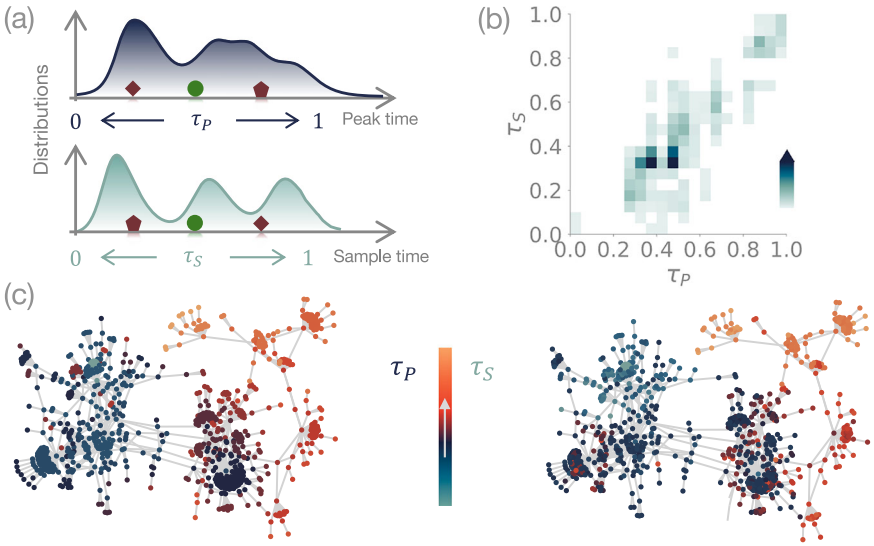

Fig. 5 | SIMS model reproduce partially the evolutionary trajectories of the H3N2 virus. a Scheme of the methodology used to compare simulations with genotype sampling times. We show example distributions of the measurements of the normalized peak and sampling times. Measurements of diamond and pentagon strains would be uncorrelated, while those of the circle strain would be correlated. **b** Density scatter plot showing the correlation between peak times (τ_P) and sampling times (τ_S) for the second-largest connected component. The darker the color, the more measurements correspond to that combination of values. **c** Visual comparison between the normalized sequence of peak times (left panel) and the normalized sequence of sampling times (right panel) for the second-largest connected component of the INFV A (H3N2) network. In the numerical simulations, the infectivity rate $\beta = 0.3$, the basal recovery rate $\mu_0 = 0.1$, the immunity acquisition rate $\alpha = 0.03$, the waning immunity rate $\gamma = 0$, the mutation rate $D_x = 10^{-5}$, and the characteristic cross-immunity length $\Delta = 3$.

infected by each community. Initially, infections are dominated by the teal swarm, but as immunity is acquired, the antigenic escape shifts the dynamics toward the adjacent swarms. Eventually, the dynamics become dominated by the orange and red swarms, which are the most antigenically different from the wild type. For clarity, the lower panel of Fig. 4b presents the relative prevalence $I^{rel}(t)$, i.e., the fraction of infections attributable to each swarm (see Eq. (11) in Methods), clearly showing the transition from the teal to the ocher communities.

In Fig. 4c, we present the second-largest component of the Influenza A network, which contains six prominent mutant swarms. Due to its more intricate topology around the wild type, the epidemic trajectory shown in Fig. 4d is correspondingly complex, with several swarms evolving almost concurrently within the most tightly connected region of the genotype network. Nonetheless, toward the end of the trajectory, the orange swarm emerges as dominant because of its greater genetic distance from the other regions. The mutant swarm landscape captured by $I^{rel}(t)$ in this component is elaborate, as several swarms appear and fade, resembling the complex lineage patterns reported in public data repositories[30].

Similarly, Fig. 4e, f depicts the seventh largest component of the Influenza A network. This component exhibits a more hierarchical structure, similar to the synthetic concatenation of star-like clusters presented in Fig. 3. As a result, identifying mutant swarms corresponding to distinct peaks in the epidemic trajectory is more straightforward.

In all the cases shown in Fig. 4, epidemic trajectories are notably more complex than those derived from lower-dimensional representations, such as the linear chain shown in Supplementary Fig. 1.1 or the lattice used in Fig. 3. Mesoscopically, our results show how antigenic drift across genotype networks yields complex dynamic variants' landscapes whose shape and dynamics, e.g., the emergence and turnover of lineages due to antigenic fitness or the clonal interference between lineages, are reminiscent of those observed in real outbreaks. This correspondence arises from the interplay between the infection history and the trajectory of the virus across the antigenic network. Supplementary Fig. 8 in Supplementary Note 8 shows how neglecting these features indeed leads to unrealistic variant landscapes, characterized by continuous genotype swarm alternations or the re-emergence of all lineages when cross-immunity or immune memory are not accounted for, respectively.

To round off our analysis, we now focus on the microscopic properties of the generated trajectories under the SIMS model. In particular, we are interested in analyzing whether the temporal dynamics of strains in our model resembles the timeline of the different H3N2 strains extracted from genomic surveillance data, i.e., the time at which the individual peak of each strain (genotype) occurred. To perform this comparison, we record the sequence of peak times of each strain i in the simulation, τ_i^P, and the day at which the first sample corresponding to the strain was recorded in the genomic surveillance data, τ_i^S. As the global time scales of the dynamics might be different in both distributions, we normalize the values as illustrated in Fig. 5a so that $\tau_i^S = 0$

($\tau_i^P = 0$) and $\tau_i^S = 1$ ($\tau_i^P = 1$) correspond to the first and last strain recorded in the simulations (real data), respectively. Figure 5b reports a high correlation between the synthetic and real time series, indicating that the SIMS model can reproduce the evolutionary trajectory of the strains composing the second-largest component of the antigenic genotype network of the H3N2 virus. This correspondence between the genomic data generated in silico by the SIMS model and the empirically sampled genomic data is further supported in Fig. 5c, d, which visualizes the synthetic and real epidemic trajectories over the genotype network. In both cases, nodes are colored according to the time of their first appearance in the epidemic trajectory.

Discussion

Antigenic drift has been long stated as the major driver of the sustained circulation of endemic viruses in real populations. From a dynamical point of view, the incidence of endemic communicable diseases spans a wide range of temporal patterns. Some diseases, such as tuberculosis[51] and malaria[52] in many regions of Africa, maintain relatively stable, quasi-constant levels of incidence while others, such as influenza-like illnesses, often exhibit quasi-periodic seasonal fluctuations[5,53]. While low-dimensional representations of antigenic spaces allow obtaining the former dynamical behavior, setting a formal connection between the structure of the antigenic space and the emergence of seasonal epidemics remains elusive. To solve this puzzle, here we have introduced the SIMS model, an eco-evolutionary framework where the antigenic space is explicitly represented as a complex genotype network. Another particularly salient feature of the model, compared to other frameworks relying on genotype networks is the incorporation of a memory core. That allows the system to keep track of the infection history of the population instead of only accounting for the most recent infection[24,50]. Unlike previous frameworks, the infection history of the population is encoded in the recovery rates for each strain rather than in the depletion of their associated pool of susceptible agents[22,24,25,49,54]. In particular, our mean-field equations assume that, when a host contracts one strain, they develop an immune response which increases the expected recovery rate of a randomly chosen agent in the population.

Focusing on the H3N2 virus, we have found that real antigenic genotype networks present complex features, such as long-tailed degree distributions, high modularity, and pronounced disassortative mixing of strains, which cannot be accommodated in low-dimensional antigenic spaces. We have proposed a minimal synthetic model presenting these features, assuming a complex network composed of a consecution of genotype swarms. The epidemic trajectories observed across this antigenic space exhibit oscillatory (seasonal) behavior, while assuming homogeneous lattices instead results in steady endemic epidemics. Remarkably, seasonality here emerges from the complex interplay between antigenic evolution and immunity[28,54] rather than from the periodic behavior of the force of infection as in other diseases[55,56]. Overall, our results show that the topology of the structure of antigenic space is a critical determinant of the dynamical behavior of epidemic trajectories.

The SIMS model not only captures macroscopic epidemic patterns but also sheds light on the complex mesoscopic and microscopic ecological dynamics created by the antigenic drift of the H3N2 virus. When analyzing mesoscale dynamics, we have found how the interplay between immune memory, cross-immunity among strains and mutant swarms yields viral trajectories mirroring the complex lineage patterns documented in public data repositories[30]. From a microscopic perspective, we have found that the emergence times of the strains in the model and their first associated record in the actual genomic data are highly correlated, showing how the SIMS model can partially reconstruct the evolutionary history of the virus.

Despite the advantages of the SIMS model, providing a full reconstruction of both epidemiological and evolutionary dynamics with our framework still remains challenging. This limitation arises because genotype networks display multiple disconnected components, which typically differ by only two or three point mutations from at least one other component, indicating the possibility of unobserved genotypes linking otherwise distinct clusters[41]. To overcome this issue, promising avenues for future research could involve leveraging Bayesian inference or machine learning techniques to reconstruct the complete network[57,58] or the formulation of generative models linking the microscopic genomic structure of viruses with the topology of their antigenic spaces[10]. Furthermore, the immunity mechanism here introduced is a mean-field approach that reduces the effective reproductive number of previously prevalent strains below 1. Although this assumption effectively reduces the strains' population fitness, it is less restrictive than considering a directed, tree-like structure that explicitly forbids loops[8] and imposes an evolutionary direction[24]. This could potentially lead to an overestimation of the role of very old infections in the population. Despite the spurious long memory effects introduced by the mean-field assumption, the SIMS model provides a minimal description of the impact of antigenic escape across complex multidimensional antigenic spaces on the epidemic trajectories.

Looking forward, our work establishes a quantitative framework for integrating genomic data into epidemic models and opens promising avenues for future research. Moving from mean-field approaches to an individual-based framework[22] would allow us to describe more accurately the infection history of hosts and introduce heterogeneous contact rates[59,60]. rates. Together with host mobility[61,62], dynamic strain-specific features (such as variable infectivity[18]), and more detailed within-host dynamics[63], this could improve the realism and predictive power of the model. Moreover, investigating the role of immunocompromised individuals[64] in facilitating extensive antigenic drift, especially for emerging pathogens like SARS-CoV-2, may yield deeper insights into viral persistence and evolution. Beyond the biological aspects, incorporating the effect of public health policies[65,66] together with the behavioral responses of individuals to them[67–69] will help to devise realistic control scenarios[70].

Overall, our findings underscore the critical influence of genotype network topology on epidemic trajectories and viral endemicity. This study provides valuable insights into the mechanisms by which emerging pathogens navigate antigenic space and lays a robust foundation for future research at the interface of viral genomics and infectious disease modeling. In doing so, it has the potential to inform public health interventions that account for both antigenic drift and viral circulation.

Methods

Genotype networks

Formally, a genotype network is defined as a graph, $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, where nodes represent the N distinct genetic sequences (genotypes) $s_i \in \mathcal{N}$ ($i = 1, \ldots, N$) associated to the same virus. Each sequence s_i consists of a concatenation of S genetic elements, i.e., $s_i = e_1^i, \ldots, e_S^i$. In addition, the links in the set $\mathcal{L}$ connect sequences that differ by a single mutation. The adjacency matrix $\mathbf{A} = \{a_{ij}\}$ encodes these connections, where $a_{ij} = 1$ if s_i and s_j differ by one genetic element, and $a_{ij} = 0$ otherwise. Another important matrix characterizing $\mathcal{G}$ is the out-degree normalized Laplacian matrix: $\mathbf{L} = \{\ell_{ij}\}$, whose elements are defined as $\ell_{ij} = \delta_{ij} - a_{ij}/k_j$ where $k_i = \sum_{j=1}^{N} a_{ij}$ is the degree of genotype i. Finally, we introduce the genetic distance matrix $\mathbf{X} = \{x_{ij}\}$, whose elements are $x_{ij} = \sum_{n=1}^{S} \delta(e_n^i, e_n^j)$, where $\delta(x, y)$ is the Kronecker delta. Notably, if $\mathcal{G}$ is connected, x_{ij} is the shortest path length between s_i and s_j.

Genotype networks can be constructed at different genetic scales. A node in the network may represent a genotype at three possible levels of resolution: (i) nucleotide sequences, the most fine-grained representation[33]; (ii) amino acid sequences, where nodes represent amino acids[41]; or (iii) gene-level representations, where mutations involve entire functional units. Coarse-grained representations introduce degeneracy, as multiple nucleotide sequences can encode the same amino acid, and different genetic sequences may correspond to the same node when only a section of the genome is considered.

Among all genomic regions, those encoding proteins involved in host–pathogen interactions are of particular interest, as they directly influence immune evasion. In particular, in this study we use the genotype network for the highly antigenic hemagglutinin protein of influenza A (H3N2), constructed in ref. 41 using data from the Influenza Research

Database[71]. This network, built at the amino acid level, connects genotypes differing by an amino acid.

Configurational model

To build the ensembles of 100 synthetic networks utilized in Fig. 2 and Supplementary Fig. 4, we have utilized the function *configuration_model* of the *networkx* library in Python[72]. Through this method, for each of the connected components of the Influenza A network, we build 100 networks that have the same degree sequence (i.e., each node has the same number of connections), but with randomized connections. Note that, in case the generated structures were disconnected, we have added the least possible number of extra links to ensure connectedness.

Structural metrics

To analyze the structural properties of the genotype networks $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, we compute the following metrics by using the *networkx* library in Python[73]:

Average shortest path length ($L(\mathcal{G})$)[74,75]:

$$L(\mathcal{G}) = \frac{1}{N(N-1)} \sum_{i \neq j} l_{ij}, \tag{1}$$

where l_{ij} is the shortest genetic distance between strains s_i and s_j, with $L(\mathcal{G}) = \infty$ if $\mathcal{G}$ is disconnected.

Degree assortativity coefficient($r(\mathcal{G})$)[76,77]:

$$r(\mathcal{G}) = \frac{\sum_{(i,j)\in\mathcal{L}} (k_i - \bar{k})(k_j - \bar{k})}{\sum_{(i,j)\in\mathcal{L}} (k_i - \bar{k})^2}. \tag{2}$$

The assortativity coefficient assesses similarity in degrees (connections) between nodes, where $\bar{k}$ is the average degree over all nodes in $\mathcal{N}$.

Average clustering coefficient ($C(\mathcal{G})$)[78]:

$$C(\mathcal{G}) = \frac{1}{N} \sum_{i\in\mathcal{N}} c_i, \tag{3}$$

where c_i is the local clustering coefficient of node i, defined as

$$c_i = \frac{2E_i}{k_i(k_i - 1)} \quad \text{for } k_i > 1, \tag{4}$$

with E_i denoting the number of edges between the k_i neighbors of node i. If $k_i \leq 1$, $c_i = 0$. $C(\mathcal{G})$ quantifies the tendency of nodes to form locally dense clusters.

Transitivity ($T(\mathcal{G})$):

$$T(\mathcal{G}) = \frac{3 \cdot \#\,\text{triangles}}{\#\text{triads}}, \tag{5}$$

where a triangle is a set of three nodes all mutually connected, and a triad is a set of three nodes where at least two are connected. $T(\mathcal{G})$ evaluates the global tendency of the network to form closed triplets.

Modularity ($Q(\mathcal{G})$)[79]:

$$Q(\mathcal{G}) = \frac{1}{2|\mathcal{L}|} \sum_{i,j\in\mathcal{N}} \left[A_{ij} - \frac{k_i k_j}{2|\mathcal{L}|} \right] \delta(g_i, g_j), \tag{6}$$

where the cardinality of the set of links $|\mathcal{L}|$ is the number of connections, and $\delta(g_i, g_j)$ equals 1 if nodes i and j are in the same community and 0 otherwise. Therefore, $Q(\mathcal{G})$ measures the density of intra-community edges relative to inter-community edges.

Note that the communities have been inferred with the Infomap method, described in the subsection below.

Community detection algorithm

Community detection was performed using the Infomap method[80,81] as implemented in the *infomap* library in Python. This algorithm encodes the network structure as a compressed map of information flow by minimizing the description length of random walks. It iteratively partitions nodes into communities based on flow patterns, recursively refining the hierarchy until further compression is not possible.

Effective dynamical equations of the SIMS model

The dynamics of the SIMS model can be captured by a set of coupled differential equations. For simplicity, we assume a well-mixed population where all individuals are equivalent from a microscopic perspective. This leads to a mean-field description in which the relevant variables are the probability of being susceptible, $\rho^S(t)$ and the probabilities of being infectious with each of the n viral strains, $\rho_i^I(t)$ ($i = 1, \ldots, n$). Equipped with these variables, we are ready to formulate the set of differential equations describing the dynamical evolution of the SIMS framework:

$$\dot{\rho}_i^I(t) = \beta \rho_i^I(t) \rho^S(t) - \mu_i(t) \rho_i^I(t) - D_x \sum_{j=1}^{n} \ell_{ij} \rho_j^I(t), \tag{7}$$

$$\dot{\mu}_i(t) = \alpha \sum_{j=1}^{n} \left(\rho_j^I(t) e^{-x_{ij}/\Delta} \right) - \gamma(\mu_i(t) - \mu_0), \tag{8}$$

where the relevant parameters and matrices introduced before are present. Since the sum of probabilities associated with the state of the population at some time t must fulfill:

$$\rho^S(t) = 1 - \sum_{i=1}^{n} \rho_i(t)\ , \tag{9}$$

the evolution of the dynamical system is fully described by the set of $2n$ Eqs. (7)–(8).

Let us describe the terms of Eqs. (7)–(8) associated with the three key mechanisms of the SIMS model. The first two terms in Eq. (7) describe the contagion dynamics following a mean-field SIS process for each viral genotype. Thus, infection occurs at a rate β, while recovery takes place at a strain-dependent rate $\mu_i(t)$. Unlike the standard SIS model, where the recovery rate is constant, here it evolves dynamically according to Eq. (8) to incorporate the effects of immune response.

In particular, Eq. (8) captures how recovery rate increases as immunity is acquired, with the gain being proportional to the prevalence of the strain at the population level $\rho_i(t)$, modulated by the immunity acquisition rate α. This term serves as a memory core, allowing the system to keep track of past infections.

In addition to immunity gained from direct infection, Eq. (8) also accounts for how individuals acquire cross-immunity from related strains. The strength of this effect depends on the genetic distance between the strains, encoded in the matrix $\mathbf{X} = \{x_{ij}\}$. Strains that are antigenically similar contribute more to immunity gain than those that are more genetically distinct. The extent of cross-immunity is controlled by the characteristic immunity length Δ. When $\Delta \to 0$, immunity is strain-specific, while for $\Delta \to \infty$, all strains confer maximal cross-immunity. Finally, the last term in Eq. (8) incorporates the effect of waning immunity over time, modeled as a decay toward the basal recovery rate μ_0 with rate γ.

To round off, we turn our attention again to Eq. (7). There, the last term accounts for mutation dynamics, that allows infected individuals to change infectious compartments during the course of their infection. This process is represented through the diffusion term (and thus is governed by the normalized Laplacian $\mathbf{L} = \{\ell_{ij}\}$) across the genotype network. The importance of this term is weighted by the mutation rate D_x.

Table 1 | Empirical measurements of mutation rates of the influenza A virus

Mutation rate per cycle	Mutation rate per day	Specifically N3H2	Reference
1.5×10^{-5}	5.14×10^{-5}	No	83
1.2×10^{-6}	4.11×10^{-6}	Yes	84
2.3×10^{-5}	7.89×10^{-5}	No	85
2.5×10^{-4}	8.57×10^{-4}	Yes	86

For Influenza A, the mutation rate per nucleotide per cycle takes values in the range $[10^{-4}, 10^{-6}]$, with a cycle lasting ~7 h[84]. The particular genotype network of influenza A used in the manuscript corresponds to the H3N2 type of influenza A, and only refs. 84,86 account for that specific strain. Therefore, we assume as a compromise that $D_x = 10^{-5}$.

Integration of the dynamical equations

Epidemic trajectories are obtained by integrating Eqs. (7)–(8) through a fourth-order Runge–Kutta with a time step $\delta t = 0.01$. There, in addition to tracking the fraction of the population infected by each viral strain, $\rho_i^I(t)$, we also use the global prevalence of the disease,

$$I(t) = \sum_{i=1}^{n} \rho_i^I(t), \tag{10}$$

and the relative prevalence of each strain,

$$I_i^{rel}(t) = \frac{\rho_i^I(t)}{I(t)}, \tag{11}$$

as the primary metrics to characterize the disease.

Control parameters

Below, we define the three control parameters utilized in the article.

Basic reproduction number of the SIMS model. Provided β is the effective infectivity rate and μ_0 the basal recovery rate, the basic reproduction number R_0 is defined as follows:

$$R_0 = \frac{\beta}{\mu_0}. \tag{12}$$

Maximum peak infectivity in the SIR model. For a pathogen characterized by R_0, the incidence in the peak of the epidemic reads as follows[82]:

$$I_{\max} = 1 - \frac{1}{R_0}\left[1 + \ln(R_0)\right]. \tag{13}$$

Effective reproduction number of the SIMS model. For a genotype s_i, provided β is the effective infectivity rate and $\mu_i(t)$ the recovery rate, the effective reproduction number $R_i^{eff}(t)$ is defined as follows:

$$R_i^{\text{eff}}(t) = \frac{\beta}{\mu_i(t)} \rho^S(t). \tag{14}$$

Values of the model parameters

For simulations in Figs. 3b, 4, and 5, we chose values within the biologically plausible range for rapidly evolving viruses. In particular, we set the mutation rate to $D_x = 10^{-5}$ based on references[83–86] (see Table 1 for details). For the basal recovery rate, we choose $\mu_0^{-1} = 10$ days. We have set the infectivity rate to $\beta = 0.3$, yielding a basic reproduction number $R_0 = 3$. For the immunity-related parameters, we set $\gamma = 0$ because memory B cells persist for a long time in this class of viruses. Furthermore, we have chosen α such that, for an isolated strain, the prevalence peak corresponds to that predicted by the traditional SIR model via Eq. (13). Finally, we set the characteristic cross-immunity length to $\Delta = 3$ to account for its trade-off with the swarm sizes. See Supplementary Fig. 5 for further analysis on the effect of this latter parameter. Finally, we would like to point out that in refs. 23,24, the authors fitted the epidemiological parameters to the data.

Data availability

The Influenza A genotype network can be obtained in ref. 41.

Code availability

The code is available at https://github.com/santiagolaot/SIMS-model.

Received: 12 June 2025; Accepted: 31 October 2025;
Published online: 19 November 2025

References

1. W.H.O. Covid-19 dashboard. https://data.who.int/dashboards/covid19/cases (2024).
2. Chen, Z. et al. Global landscape of SARS-CoV-2 genomic surveillance and data sharing. *Nat. Genet.* **54**, 499 (2022).
3. Roemer, C. et al. SARS-CoV-2 evolution in the omicron era. *Nat. Microbiol.* **8**, 1952 (2023).
4. Markov, P. V. et al. The evolution of SARS-CoV-2. *Nat. Rev. Microbiol.* **21**, 361 (2023).
5. Google Flu Trends. https://www.google.com/publicdata/explore?ds=z3bsqef7ki44ac_&hl=en&dl=en (2024).
6. Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. Evolution and ecology of influenza A viruses. *Microbiol. Rev.* **56**, 152 (1992).
7. Nelson, M. I. & Holmes, E. C. The evolution of epidemic influenza. *Nat. Rev. Genet.* **8**, 196 (2007).
8. Smith, D. J. et al. Mapping the antigenic and genetic evolution of influenza virus. *Science* **305**, 371 (2004).
9. Ferguson, N. M., Galvani, A. P. & Bush, R. M. Ecological and immunological determinants of influenza evolution. *Nature* **422**, 428 (2003).
10. Koelle, K., Cobey, S., Grenfell, B. & Pascual, M. Epochal evolution shapes the phylodynamics of interpandemic influenza A (H_3N_2) in humans. *Science* **314**, 1898 (2006).
11. 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. USA* **104**, 7711 (2007).
12. Katzelnick, L. C. et al. Antigenic evolution of dengue viruses over 20 years. *Science* **374**, 999 (2021).
13. Streicker, D. G. et al. Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. *Science* **329**, 676 (2010).
14. Moya, A., Holmes, E. C. & González-Candelas, F. The population genetics and evolutionary epidemiology of RNA viruses. *Nat. Rev. Microbiol.* **2**, 279 (2004).
15. Mittal, A., Khattri, A. & Verma, V. Structural and antigenic variations in the spike protein of emerging SARS-CoV-2 variants. *PLoS Pathog.* **18**, e1010260 (2022).
16. Guan, Y., Vijaykrishna, D., Bahl, J., Zhu, H., Wang, J. & Smith, G. J. The emergence of pandemic influenza viruses. *Protein Cell* **1**, 9 (2010).
17. Guzmán, M. G., Kouri, G. & Halstead, S. B. Do escape mutants explain rapid increases in dengue casefatality rates within epidemics? *Lancet* **355**, 1902 (2000).
18. Soriano-Paños, D. Eco-evolutionary constraints for the endemicity of rapidly evolving viruses. *PRX Life* **3**, 043001 (2025).
19. Carabelli, A. M. et al. SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nat. Rev. Microbiol.* **21**, 162 (2023).
20. Tuekprakhon, A. et al. Antibody escape of SARS-COV-2 Omicron BA.4 and BA.5 from vaccine and BA.1 serum. *Cell* **185**, 2422 (2022).
21. Van de Sandt, C. E., Kreijtz, J. H. & Rimmelzwaan, G. F. Evasion of influenza A viruses from innate and adaptive immune responses. *Viruses* **4**, 1438 (2012).

22. Bedford, T., Rambaut, A. & Pascual, M. Canalization of the evolutionary trajectory of the human influenza virus. *BMC Biol.* **10**, 38 (2012).
23. Bedford, T. et al. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature* **523**, 217 (2015).
24. Rouzine, I. M. & Rozhnova, G. Antigenic evolution of viruses in host populations. *PLoS Pathog.* **14**, e1007291 (2018).
25. Chardès, V., Mazzolini, A., Mora, T. & Walczak, A. M. Evolutionary stability of antigenically escaping viruses. *Proc. Natl Acad. Sci. USA* **120**, e2307712120 (2023).
26. Tsimring, L. S., Levine, H. & Kessler, D. A. RNA virus evolution via a fitness-space model. *Phys. Rev. Lett.* **76**, 4440 (1996).
27. Rouzine, I. M., Wakeley, J. & Coffin, J. M. The solitary wave of asexual evolution. *Proc. Natl Acad. Sci. USA* **100**, 587 (2003).
28. Yan, L., Neher, R. A. & Shraiman, B. I. Phylodynamic theory of persistence, extinction and speciation of rapidly adapting pathogens. *Elife* **8**, e44205 (2019).
29. Marchi, J., Lässig, M., Walczak, A. M. & Mora, T. Antigenic waves of virus–immune coevolution. *Proc. Natl Acad. Sci. USA* **118**, e2103398118 (2021).
30. GISAID. Global initiative on sharing all influenza data. https://www.gisaid.org (2024).
31. Nei, M. & Kumar, S. *Molecular Evolution and Phylogenetics* (Oxford Univ. Press, 2000).
32. Langedijk, A. C. et al. The genomic evolutionary dynamics and global circulation patterns of respiratory syncytial virus. *Nat. Commun.* **15**, 3083 (2024).
33. Seoane, L. F., Secaira-Morocho, H., Lázaro, E. & Manrubia, S. Hierarchical genotype networks and incipient ecological speciation in q-beta phage quasispecies. Preprint at arXiv:2411.07110 (2024).
34. Wagner, A. Genotype networks shed light on evolutionary constraints. *Trends Ecol. Evol.* **26**, 577 (2011).
35. Wagner, A. A genotype network reveals homoplastic cycles of convergent evolution in influenza A (H3N2) haemagglutinin. *Proc. R. Soc. B: Biol. Sci.* **281**, 20132763 (2014).
36. Manrubia, S. et al. From genotypes to organisms: state-of-the-art and perspectives of a cornerstone in evolutionary dynamics. *Phys. Life Rev.* **38**, 55 (2021).
37. Dábilla, N. & Dolan, P. T. Structure and dynamics of enterovirus genotype networks. *Sci. Adv.* **10**, eado1693 (2024).
38. R. Vahdati, A. & Wagner, A. Parallel or convergent evolution in human population genomic data revealed by genotype networks. *BMC Evol. Biol.* **16**, 1 (2016).
39. Newman, M. *Networks* (Oxford Univ. Press, 2018).
40. Latora, V., Nicosia, V. & Russo, G. *Complex Networks: Principles, Methods and Applications* (Cambridge Univ. Press, 2017)
41. Williams, B. J., Ogbunugafor, C. B., Althouse, B. M. & Hébert-Dufresne, L. Immunity-induced criticality of the genotype network of influenza A (H3N2) hemagglutinin. *PNAS Nexus* **1**, pgac143 (2022).
42. 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**, 233 (1989).
43. Andreasen, V., Lin, J. & Levin, S. A. The dynamics of cocirculating influenza strains conferring partial cross-immunity. *J. Math. Biol.* **35**, 825 (1997).
44. Kamo, M. & Sasaki, A. The effect of cross-immunity and seasonal forcing in a multi-strain epidemic model. *Phys. D Nonlinear Phenom.* **165**, 228 (2002).
45. Ferguson, N. & Andreasen, V. In *Mathematical Approaches for Emerging and Reemerging Infectious Diseases: Models, Methods, and Theory.* (eds Castillo-Chavez, C. et al.) pp. 157–169 (Springer, 2002).
46. Minayev, P. & Ferguson, N. Improving the realism of deterministic multi-strain models: implications for modelling influenza A. *J. R. Soc. Interface* **6**, 509 (2009).
47. Kucharski, A. J., Andreasen, V. & Gog, J. R. Capturing the dynamics of pathogens with many strains. *J. Math. Biol.* **72**, 1 (2016).
48. Williams, B. J., St-Onge, G. & Hébert-Dufresne, L. Localization, epidemic transitions, and unpredictability of multistrain epidemics with an underlying genotype network. *PLoS Comput. Biol.* **17**, e1008606 (2021).
49. Gog, J. R. & Grenfell, B. T. Dynamics and selection of many-strain pathogens. *Proc. Natl Acad. Sci. USA* **99**, 17209 (2002).
50. Lin, J., Andreasen, V., Casagrandi, R. & Levin, S. A. Traveling waves in a model of influenza A drift. *J. Theor. Biol.* **222**, 437 (2003).
51. Cobelens, F. G. et al. Risk of infection with *Mycobacterium tuberculosis* in travellers to areas of high tuberculosis endemicity. *Lancet* **356**, 461 (2000).
52. Hay, S., Smith, D. & Snow, R. Measuring malaria endemicity from intense to interrupted transmission. *Lancet Infect. Dis.* **8**, 369 (2008).
53. Hamid S, P. R. et al. Seasonality of respiratory syncytial virus — United States, 2017-2023. *MMWR Morb. Mortal. Wkly. Rep.* **72**, 355 (2023).
54. Sasaki, A., Lion, S. & Boots, M. Antigenic escape selects for the evolution of higher pathogen transmission and virulence. *Nat. Ecol. Evol.* **6**, 51 (2022).
55. London, W. P. & Yorke, J. A. Recurrent outbreaks of measles, chickenpox and mumps: I. seasonal variation in contact rates. *Am. J. Epidemiol.* **98**, 453 (1973).
56. Earn, D. J., Rohani, P. & Grenfell, B. T. Persistence, chaos and synchrony in ecology and epidemiology. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **265**, 7 (1998).
57. Bonacina, F. et al. Characterization and forecast of global influenza (sub) type dynamics. Preprint at *medRxiv* https://doi.org/10.1101/2024.08.01.24311336 (2024).
58. Parino, F. et al. Integrating dynamical modeling and phylogeographic inference to characterize global influenza circulation. *PNAS Nexus* **4**, pgae561 (2025).
59. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925 (2015).
60. Granell, C., Gómez, S., Gómez-Gardeñes, J. & Arenas, A. Probabilistic discrete-time models for spreading processes in complex networks: a review. *Ann. der Phys.* **536**, 2400078 (2024).
61. Yeakel, J. D., Gibert, J. P., Gross, T., Westley, P. A. & Moore, J. W. Eco-evolutionary dynamics, density-dependent dispersal and collective behaviour: implications for salmon metapopulation robustness. *Philos. Trans. R. Soc. B Biol. Sci.* **373**, 20170018 (2018).
62. Blot, N. et al. How host mobility patterns shape antigenic escape during viral-immune coevolution. *PRX Life* **3**, 023011 (2025).
63. Zhang, X. et al. Epidemic spreading under mutually independent intra- and inter-host pathogen evolution. *Nat. Commun.* **13**, 6218 (2022).
64. Smith, C. A. & Ashby, B. Antigenic evolution of SARS-CoV-2 in immunocompromised hosts. *Evol. Med. Public Health* **11**, 90 (2023).
65. Kojaku, S., Hébert-Dufresne, L., Mones, E., Lehmann, S. & Ahn, Y.-Y. The effectiveness of backward contact tracing in networks. *Nat. Phys.* **17**, 652 (2021).
66. Lamata-Otín, S., Reyna-Lara, A. & Gómez-Gardeñes, J. Integrating virtual and physical interactions through higher-order networks to control epidemics. *Chaos Soliton. Fract.* **189**, 115592 (2024).
67. Steinegger, B., Arola-Fernández, L., Granell, C., Gómez-Gardeñes, J. & Arenas, A. Behavioural response to heterogeneous severity of COVID-19 explains temporal variation of cases among different age groups. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **380**, 20210119 (2022).
68. Saad-Roy, C. M. & Traulsen, A. Dynamics in a behavioral-epidemiological model for individual adherence to a nonpharmaceutical intervention. *Proc. Natl Acad. Sci. USA* **120**, e2311584120 (2023).

69. Traulsen, A., Levin, S. A. & Saad-Roy, C. M. Individual costs and societal benefits of interventions during the COVID-19 pandemic. *Proc. Natl Acad. Sci. USA* **120**, e2303546120 (2023).
70. Morris, D. H., Rossine, F. W., Plotkin, J. B. & Levin, S. A. Optimal, near-optimal, and robust epidemic control. *Commun. Phys.* **4**, 78 (2021).
71. Zhang, Y. et al. Influenza research database: an integrated bioinformatics resource for influenza virus research. *Nucleic Acids Res.* **45**, D466 (2017).
72. Newman, M. E. The structure and function of complex networks. *SIAM Rev.* **45**, 167 (2003).
73. Hagberg, A., Swart, P. J. & Schult, D. A. *Exploring Network Structure, Dynamics, and Function Using NetworkX*. Report Nos. LA-UR-08-05495; LA-UR-08-5495 (Los Alamos National Laboratory (LANL), 2008).
74. Euler, L. Solutio problematis ad geometriam situs pertinentis, Commentarii academiae scientiarum Petropolitanae, 128 (1741)
75. Dijkstra, E. W. in *Edsger Wybe Dijkstra: His Life, Work, and Legacy* (ed. Apt, K. R.) Ch. 13 (Springer, 2022).
76. Newman, M. E. Mixing patterns in networks. *Phys. Rev. E* **67**, 026126 (2003).
77. Foster, J. G., Foster, D. V., Grassberger, P. & Paczuski, M. Edge direction and the structure of networks. *Proc. Natl Acad. Sci. USA* **107**, 10815 (2010).
78. Saramäki, J., Kivelä, M., Onnela, J.-P., Kaski, K. & Kertesz, J. Generalizations of the clustering coefficient to weighted complex networks. *Phys. Rev. E* **75**, 027105 (2007).
79. Newman, M. E. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113 (2004).
80. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl Acad. Sci. USA* **105**, 1118 (2008).
81. Edler, D., Holmgren, A. & Rosvall, M. The MapEquation software package. https://mapequation.org (2025).
82. Lamata-Otín, S., Reyna-Lara, A., Soriano-Paños, D., Latora, V. & Gómez-Gardeñes, J. Collapse transition in epidemic spreading subject to detection with limited resources. *Phys. Rev. E* **108**, 024305 (2023).
83. Parvin, J. D., Moscona, A., Pan, W., Leider, J. & Palese, P. Measurement of the mutation rates of animal viruses: influenza A virus and poliovirus type 1. *J. Virol.* **59**, 377 (1986).
84. Nobusawa, E. & Sato, K. Comparison of the mutation rates of human influenza A and B viruses. *J. Virol.* **80**, 3675 (2006).
85. Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M. & Belshaw, R. Viral mutation rates. *J. Virol.* **84**, 9733 (2010).
86. Pauly, M. D., Procario, M. C. & Lauring, A. S. A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses. *Elife* **6**, e26437 (2017).

Acknowledgements

S.L.-O and J.G.-G. acknowledge financial support from the Departamento de Industria e Innovaweción del Gobierno de Aragón y Fondo Social Europeo (FENOL group grant E36-23R) and from Ministerio de Ciencia, Innovación y Universidades (grant PID2023-147734NB-I00). S.L.-O. acknowledges financial support from the Gobierno de Aragón through a doctoral fellowship. A.A. acknowledges Spanish Ministerio de Ciencia, Innovación y Universidades (PID2021-128005NB-C21), Generalitat de Catalunya (2021SGR-00633), Universitat Rovira i Virgili (2023PFR-URV-00633), the European Union's Horizon Europe Program under the CREXDATA project (grant agreement no. 101092749), ICREA Academia, the James S. McDonnell Foundation (Grant N. 220020325), and the Joint Appointment Program at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated for the US Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830. D.S.-P. acknowledges financial support through grants JDC2022-048339-I and PID2021-128005NB- C21 funded by MCIN/AEI/10.13039/501100011033 and the European Union "NextGenerationEU"/PRTR".

Author contributions

J.G.-G. and D.S.-P. conceived the study. S.L.-O., A.A., D.S.-P., and J.G.-G. contributed to the design of the methodology. S.L.-O. and O.C.R.-I. performed preliminary formal analysis and numerical simulations. S.L.-O. performed the formal analysis, carried out the numerical simulations, and took care of data curation. S.L.-O., A.A., D.S.-P., and J.G.-G. contributed to the review and editing. S.L.-O., O.C.R.-I., A.A., D.S.-P., and J.G.-G. approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42005-025-02406-5.

Correspondence and requests for materials should be addressed to David Soriano-Paños or Jesús Gómez-Gardeñes.

Peer review information *Communications Physics* thanks Igor Rouzine and Ira B. Schwartz for their contribution to the peer review of this work. A peer review file is available.