Complex evolution of viral respiratory pathogens can be an unintended consequence of social distancing

Simon Plakolb[1,2*], Patrick Mellacher[2]

[1]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Oslo 0316, Norway.

[2]Graz Schumpeter Centre, University of Graz, Graz 8010, Austria.

[*]Corresponding author. Blindernveien 31, 0371 Oslo. Email: simon.plakolb@ibv.uio.no

Classification: Biological Sciences, Systems Biology

Keywords: Epidemiology, Complexity, Mathematical modeling

Abstract

In the early stage of a pandemic, social distancing effectively curbs the infection spread. However, non-pharmaceutical interventions may also affect pathogenic evolution. This is particular interesting in the context of the Covid-19 pandemic, as the evolution of the SARS-CoV-2 virus is highly complex compared to influenza viruses. Using a novel evolutionary-epidemiological model combining the classical SIR differential equations with an agent-based approach, we investigate the impact of social distancing interventions on the structural complexity of emergent phylogenetic trees. In our model, interventions are connected to evolutionary complexity through their impact on non-specific immunity in the population. For our analysis, we introduce tree complexity metrics from multiple disciplines that have previously not been used in this context. The novel model can replicate baseline viral evolution models of influenza and SARS-CoV-2. Our results show that interventions can lead to increased viral evolutionary complexity and suggest that this mechanism's salience diminishes over time, as social distancing measures are phased out. An empirical investigation of viral evolution suggests that the evolution of influenza has been affected by the pandemic, supporting our hypothesis. We find that the observed evolutionary complexity is not necessarily the result of features specific to SARS-CoV-2 but may also arise in future pandemics of highly mutable viral pathogens due to social distancing.

1 Introduction

In contrast to the seasonal influenza viruses, where evolution has been rather gradual and linear (*1*), SARS-CoV-2 has seen unexpected patterns of viral evolution that puzzled scientists (*2, 3*). This is particularly true for the viral strains – known as variants – associated with the "Omicron" lineage, which has proven to only be a distant relative to the previously dominating "Delta" variant (*4*). Accordingly, variants from the Omicron lineage tend to evade immunity acquired from other variants as well as (non-adapted) vaccinations (*5*). Hence, the non-linear evolution of SARS-CoV-2 could be a long-term public health problem.

As of this writing (March 2025), the main explanation for this phenomenon is that SARS-CoV-2 must have mutated for a prolonged period in reservoirs that were not under surveillance. Three main hypotheses have been put forward to locate such reservoirs: i) the population of countries or regions which are not sampled, ii) immunocompromised hosts (such as HIV patients) who are subject to a chronic Covid-19 infection, or iii) animal reservoirs (*2, 3, 6*). Empirical studies indicate some support for the latter two hypotheses, as some studies suggest that a predecessor of Omicron has developed in mice (*7, 8*), while other studies show that the virus is able to survive and evolve in some HIV patients, which causes idiosyncratic patterns of viral evolution (*9*). There is also evidence that immunosupressed who are treated with antiviral therapies may add to viral diversity (*10*).

The aim of our paper is to develop and test a novel additional hypothesis for this phenomenon, namely that social distancing may also have contributed to the complex viral evolution as observed in SARS-CoV-2 in contrast to influenza viruses. We develop a novel evolutionary-epidemiological model that illustrates how such a scenario can occur even if there are no hidden 'reservoirs' which interact infrequently with the general population. This insight is important, because social distancing measures are eased over time, which could mean that the strong non-linearity and evolutionary complexity of the SARS-CoV-2 virus could decrease over time.

2 Related literature

Interestingly, the *absence* of non-linear viral evolution in influenza viruses has already been studied in detail, using an evolutionary-epidemiological agent-based model that represents each host, as

well as their immune history, explicitly (*1*). This model was used to systematically investigate the necessary conditions to reproduce the evolutionary patterns observed in various influenza viruses. The authors show that they could only reproduce this pattern by introducing the assumption that an infection with one influenza variant induces a short-lived perfect cross-immunity across all strains (*1*), changing the resulting epidemiological dynamics drastically. Given an adequate mutational adaption rate, the combination of high transmissibility and strain-transcendent immunity causes an epidemiological pattern that is characterized by successive waves of infections (*11*). During each wave, many new variants are born, because a large number of hosts are infected and prove to be breeding grounds for new mutations. However, due to the infection wave, *all* viral strains 'run out of hosts', as almost all people have acquired strain-independent cross-immunity. Hence, this mechanism introduces a 'bottleneck' as a natural barrier to continued viral evolution. This is in stark contrast to a situation without perfect cross-immunity. In such a case, antigenic diversity proves to be an evolutionary advantage. Widespread immunity against certain variants cannot establish a bottleneck for viral spread as a whole, which leads to rapid speciation (*11*). In a two-strain model without perfect cross-immunity, the specific characteristics of immunity against the infecting variant determine whether two strains co-circulate or whether they replace each other as dominant variants (*12*). The variant-specific immunity has been the main driver for SARS-CoV-2's evolution since the arise of the Omicron lineages in 2022 (*13*). In a multi-strain model, the epidemiological pattern resulting from an absence of short-lived perfect cross-immunity stabilizes with a continuously high number of infected in the long run [see (*14*)]. These dynamics resemble a SIS model [see (*15*)], in which infected simply become susceptible again after the end of their infection.

These results are the starting point of our analysis. What if the viral transmission is curbed due to non-pharmaceutical interventions, and the 'natural' longer-lasting bottleneck of strain-transcendent immunity is replaced by artificial short-term bottlenecks introduced by lockdowns? We would expect that such a change could give the virus more time and space to evolve, possibly driving a more complex and non-linear path of viral evolution.

3 Methods

3.1 Overview

There is a growing literature on models coupling evolutionary and epidemiological dynamics. For reasons of tractability, these studies often focus on the case of two different variants [see (*12, 16–22*)] or a linear strain space [see, e.g., (*23*)]. Our research agenda, on the other hand, naturally calls for a model that allows for complex viral evolution with the potential for a much higher number of variants than two. There are also several recent models which feature multi-strain evolution (*1, 11, 24, 25*).

To develop a suitable framework for our analysis, we combine an evolutionary-epidemiological agent-based model developed for the Covid-19 pandemic (*14*) (referred to as the SARS-CoV-2 reference model) and a multi-regional evolutionary-epidemiological agent-based model developed to study influenza (*1*) (referred to as the influenza reference model). We combine the power of equation-based and agent-based modeling by i) simplifying the modeling of the viral transmission using ordinary differential equations which are derived from the well-known equation-based and deterministic epidemiological model (*26*) and ii) preserving the stochastic nature of mutations by modeling each virus variant as an agent containing variant-specific characteristics such as viral infectiousness, as well as the compartments necessary to model the spread of each virus variant.

This approach reduces the computational power needed to analyze the complex evolutionary-epidemiological dynamics, and hence allows us to engage in a detailed sensitivity analysis. Our model simplifies the approach of the influenza reference model (*1*) by assuming that the population within each region of this multi-regional model mixes homogeneously. This is a necessary change due to our equation-based approach. Nevertheless, our model reproduces the results of the influenza reference model (*1*) regarding linear evolutionary patterns (see Supplementary Text).

Like the SARS-CoV-2 reference model [see (*14*)], we simulate detailed compartments (susceptible, exposed, pre-symptomatic infectious, symptomatic infectious, asymptomatic infectious, recovered and dead), and incorporate mutations that cover both changes in the antigenic profile of a virus as well as characteristics of its associated disease (such as lethality and infectiousness). Both are key features of SARS-CoV-2's evolution (*27–29*). In fact, we find that adding mutations in disease characteristics is vital to reproduce the evolutionary patterns observed in SARS-CoV-2 (see figure D.6) in our model. We simulate a multi-regional model with 20 different regions that are split

to a Northern and a Southern hemisphere with corresponding seasonality affecting viral dynamics, and base our mutation and cross-immunity model on stylized biological facts [see also (*1*)].

Apart from combining the models and making their computation more efficient, we further implement intervention policies. In our model, non-pharmaceutical interventions are implemented after surpassing a specific threshold τ_b of deaths in the last 10 days. These policies reduce the viral infectiousness through the factor β_b. By varying the threshold, we can study how such interventions could affect the direction of viral evolution. A hypothetical no-intervention baseline scenario can be found by setting the parameters such that there are no interventions at all. In contrast to the influenza reference model's simulations, our scenarios deal with a novel pathogen. Thus, our simulations do not start at an equilibrium.

We calibrate our model using the parameters given by the influenza and SARS-CoV-2 reference models (see Table S1) and explore multiple parameter configurations regarding the non-pharmaceutical interventions.

Mutability is one of the core parameters in the influenza reference model's analysis. We consider mutation rates of $1 \cdot 10^{-6}$, $2.5 \cdot 10^{-6}$ and $1 \cdot 10^{-5}$ per-base per-infected host per-day. The reported mutation rates of SARS-CoV-2 (*30*) and influenza A and B viruses (*31*) lie within this range. We present our findings for the highest mutation rate of $1 \cdot 10^{-5}$. To evaluate our model results, we introduce a metric for tree linearity and utilize existing measures of structural tree complexity. Our novel metric quantifies the linearity of a tree structure. A perfectly linear tree has a linearity of 1. A tree root with an infinite number of attached leafs yields a linearity of ~ 0. In addition, to quantify the structural tree complexity (*32*) we use a multi-disciplinary approach with the tree degree entropy $H^*(T)$ of (*33*) stemming from computer science and the phylogenetic entropy index H_p (*34*) rooted in biology. Here, a higher value indicates a higher structural complexity.

We then investigate how differences in the share of population exhibiting perfect cross-immunity due to having recently recovered from an infection are correlated to these metrics. Importantly, social distancing interventions create significant variation in the share of perfectly cross-immune people, hence linking social distancing to viral evolution.

For brevity, we use Newton's dot notation for differentiation and implicit notation of endogenous time dependence. All compartments are endogenously time-dependent as they evolve according to the ODE system. Exogenous time dependence of a variable is explicitly notated with a subscript t.

For our compartmental model, we use the term *compartment* for those variables that are part of the population's mass balance. The term *state variable* is reserved for those variables that fall out of the mass balance.

3.2 Epidemiological model

The ODE system, derived from the SARS-CoV-2 reference's agent-based model [see (*14*)], is a SEPIARD model. Such a model extends the classic SIR model (*26*) with compartments for exposed (E), pre-symptomatic (P), asymptomatic (A) and deceased (D) individuals that are necessary to formulate more detailed intervention scenarios. We add an additional compartment for the short-lived non-specific immunity of the influenza reference model [see (*1*)] and a state variable for cross-immunity among virus variants.

Each compartment and state variable is a vector of 20 state variables, one for each geographical patch. In total N individuals inhabit these patches, which are uniformly populated with N_p hosts per patch p. All compartments are subject to a death rate determined by the life expectancy L. An exception to this rule is D_i which tracks the deaths caused by a variant i. Two compartments are variant-independent: The susceptible population S and the population that currently exhibits a short-lived non-specific immunity F_p. Individuals who are immune and get into contact with a variant will observe a boost of their non-specific immunity and return to F_p [see (*1*)]. If they were immune due to their non-specific immunity, they restart their residence in F_p.

Due to cross-immunity, the number of effectively susceptible individuals in each patch varies per variant. Some cross-immune individuals may be infected with another variant j or experience boosted non-specific immunity. In other words, $C_{p,i}$ may contain a sub-population which is currently not part of S_p. This sub-population cannot be subtracted from the susceptible population. Let $C_{p,i,I+F}$ denote the cross-immune sub-population currently infected with another variant or experiencing non-specific immunity due to a recent recovery from another variant. In addition, let $C_{p,i,B}$ hold the cross-immune sub-population currently experiencing a boosted non-specific immunity. The population that is susceptible to a variant i can then be defined as:

$$S_{p,i} = S_p - C_{p,i} + C_{p,i,I+F} + C_{p,i,B} \tag{1}$$

where p is the geographical patch, i is the variant in question, $C_{p,i}$ gives the number of cross-immune hosts of i in p.

Due to the varying cross-immunity between strains, the sub-population sizes are specific to each pair of variants. Thus, to calculate the sub-populations we add variant-specific helper state variables. For a variant j the variable $F_{p,j}$ holds people who currently wield a non-specific immunity and have recovered from j. The variable $B_{p,j}$ contains the population experiencing a boost in non-specific immunity following a contact with j.

Using the variant-specific non-specific immune population $F_{p,i}$, we can calculate $C_{p,i,I+F}$ as:

$$C_{p,i,I+F} = \sum_j (E_{p,i} + I_{p,i} + P_{p,i} + A_{p,i} + F_{p,j})(1 - f(d_{i,j}))\frac{C_{p,i}}{N_p} \tag{2}$$

where $f(d_{i,j})$ is the cross-immunity with $d_{i,j}$ as the antigenic distance between variants i and j (see Section 3.3). Note that we sum over a population possibly including individuals without immunity. We are, however, only interested in individuals cross-immune to variant i. Thus, we multiply our result with the fraction of cross-immune individuals within the total population.

The number of cross-immune individuals currently experiencing a non-specific immunity boost is given by:

$$C_{p,i,B} = \sum_j B_{p,j} f(d_{i,j}) \tag{3}$$

These individuals' chance to have been boosted increases with cross-immunity between i and j. Since only immune individuals can be boosted, a scaling by the fraction of cross-immune individuals as in Equation 2 would be incorrect.

We use a seasonality factor in each patch p of $s_p = 1 + \epsilon_p \sin \frac{2\pi t}{365}$. This seasonality affects the infectiousness of the virus. A share of the population is mobile $N'_p = (1-h)\frac{N_p}{N_{tot}}$, meaning that it can infect the inhabitants of other patches. This heterogeneous mixing between geographical patches is governed by the homophily parameter $h \in [0, 1]$ (*35*). Setting $h = 1$ leads to no mixing between patches, while $h = 0$ distributes the population of every patch to all others in every integration step. Finally, the infectiousness of each virus variant depends on an innate variant-specific parameter

$\beta_{W,i}$. Both $\beta_{W,i}$ and the homophily h are set to resemble (*1*)'s basic reproduction numbers and given in Table S1. Thus, the time-dependent adjusted contact rate $b_{p,i}$ for variant i can be defined as:

$$b_{p,i} = \beta_{b,t}\beta_{W,i}\left((h + N'_p)s_p I_{tot,p,i} + \sum_{q \neq p} N'_q s_q I_{tot,q,i}\right) \tag{4}$$

with the fraction of infectious population $I_{tot,p,j}$ given by:

$$I_{tot,p,i} = \frac{I_{p,i} + P_{p,i} + A_{p,i}}{N_p} \tag{5}$$

The adjusted contact rate $b_{p,i}$ also acts as the point of leverage for the intervention model through $\beta_{b,t}$ (see Section 3.4). It is directly used by the variant-independent susceptible compartment derivative:

$$\dot{S}_p = \frac{N_p}{L} + \tau\left(F_p - \sum_i b_{p,i}F_p\right) - \sum_i b_{p,i}S_p - \frac{S_p}{L} \tag{6}$$

Following the influenza reference model (*1*), we assume that the birth rate is determined by the life expectancy L and that the population stays constant, leading to a constant influx of susceptible individuals. The susceptible compartment also grows through the decay of non-specific immunities τ. However, some non-specific immunities do not decay due to an immunity boost upon exposition to a variant i. The last two terms express a loss of susceptible population. First, through new expositions to variants i determined by their specific susceptible population $S_{p,i}$ and the number of contagious contacts $b_{p,i}$ of individuals infected with i. Second, through natural death determined by the life expectancy L.

As described above, each variant i holds the number of its recovered, non-specific immune hosts in patch p in a state variable $F_{p,i}$. Its influx is determined by the variant-specific gross recovery rate $\dot{R}'_{p,i}$ defined in Equation 14. It expresses all novel recovered cases and differs from the net recovery rate by excluding those recovered hosts that die from natural causes. With its decay governed by τ, the derivative of the variant-specific non-specific immune population is given by:

$$\dot{F}_{p,i} = \dot{R}'_{p,i} - F_{p,i}(\tau + L^{-1}) \tag{7}$$

Note that individuals are only included in the cross-immunities $C_{p,i}$ once their non-specific immunity decays. Boosted non-specific immunities of cross-immune individuals caused by an exposure to variant i in patch p are collected in $B_{p,i}$. These are, by definition, already part of $C_{p,i}$ and calculated from the non-susceptible hosts in S_p and F_p:

$$\dot{B}_{p,i} = b_{p,i}(S_p - S_{p,i}) - \tau(B_{p,i} - b_{p,i}B_{p,i}) - \frac{B_{p,i}}{L} \tag{8}$$

The first term in the definition of $\dot{B}_{p,i}$ collects all susceptible individuals who came into contact with variant i but were cross-immune to it. This state variable only collects individuals whose immunity is boosted due to cross-immunity. Hence, we do not include boosted immunities of the non-specific immune population. Next, the decay of boosted immunities is considered. The non-specific immunity of boosted individuals that were re-boosted does not decay. In the last term, the death rate from natural causes is determined by the life expectancy L.

Both $F_{p,i}$ and $B_{p,i}$ are necessary installments to facilitate the calculation of the variant-specific number of susceptible hosts $S_{p,i}$. The population in these state variables is not part of the total stock's mass balance, but helps to keep track of a separate amount of (potentially) cross-immune hosts. In contrast, the variant-independent compartment of non-specific immunities is part of the mass balance and contains all boosted and recently recovered non-specific immune hosts:

$$\dot{F}_p = \sum_i \dot{R}'_{p,i} + b_{p,i}(S_p - S_{p,i}) - \tau\left(F_p - \sum_i b_{p,i}F_p\right) - \frac{F_p}{L} \tag{9}$$

In the first term, recovered individuals become non-specifically immune. Additionally, part of the cross-immune population $\sum_i S_p - Sp, i$ is exposed and receives a non-specific immunity. The second term denotes the decay of non-specific immunity determined by τ. Non-specifically immune individuals experiencing an immunity boost $\sum_i b_{p,i}F_p$ cannot lose it. In the final term, the life expectancy L determines a decline due to natural deaths.

The other SEPIARD model compartments are specific to each variant i in the phylogenetic tree. With the exception of $D_{p,i}$, they are all part of the mass balance, just as S_p and F_p.

The exposure rate $\dot{E}_{p,i}$ increases with the number of exposures that will lead to an infection $b_{p,i}S_{p,i}$. Some exposures cannot lead to an infection due to cross-immunity. These exposures lead

to a boost of non-specific immunity, as outlined above. Such exposures are not considered for this compartment. Instead, this compartment captures people who are infected, but not yet infectious. This period is called the latent period and its average duration is given by by α_i.

$$\dot{E}_{p,i} = b_{p,i} S_{p,i} - E_{p,i}(\alpha_i + L^{-1}) \tag{10}$$

After the latent period, a pre-symptomatic period P follows. Its duration is determined by μ_i. In this stage, individuals are infectious, but do not yet show any symptoms.

$$\dot{P}_{p,i} = \alpha_i E_{p,i} - P_{p,i}(\mu_i + L^{-1}) \tag{11}$$

After the pre-symptomatic period, some people exhibit symptoms and others stay asymptomatic. The variant-specific parameter $\nu_i \in [0, 1]$ describes the share of symptomatic infections. Regardless of whether agents are symptomatic or not, they will cease to be infected after – on average – γ_i^{-1} periods.

$$\dot{I}_{p,i} = \mu_i P_{p,i} \nu_i - I_{p,i}(\gamma_i + L^{-1}) \tag{12}$$

$$\dot{A}_{p,i} = \mu_i P_{p,i}(1 - \nu_i) - A_{p,i}(\gamma_i + L^{-1}) \tag{13}$$

For performance reasons, if the number of active hosts $N_a = E + P + I + A$ falls below 10^{-65} we assume a variant will not re-emerge and set these four compartments to 0. Since we still track all immunities caused by these variants, this does not affect our model results.

Recoveries of symptomatic and asymptomatic infections are tracked in R. We define a gross recovery rate $\dot{R}'_{p,i}$ as an intermediate result that also determines the influx of non-specific immunities (see Equation 9).

$$\dot{R}'_{p,i} = \gamma_i \left(A_{p,i} + I_{p,i} \left(\lambda_i + (1 - \lambda_i) \psi \frac{R_{p,tot}}{N_p} \right) \right) \tag{14}$$

Asymptomatic cases in A will not die from the disease and obtain a short-lived non-specific immunity in addition to their strain-specific immunity. Symptomatic cases in I have a chance of a lethal outcome of infection given by the viral lethality λ_i. Previous recoveries from any variant $R_{p,tot} = \sum_i R_{p,i}$ yield a cross-protection against a lethal outcome, reducing the death rate by ψ (*14*).

The net recovery rate can easily be defined by using the gross recovery rate $\dot{R'}_{p,i}$ and the natural death rate following the life expectancy L as:

$$\dot{R}_{p,i} = \dot{R'}_{p,i} - \frac{R_{p,i}}{L} \tag{15}$$

The compartment for dead individuals $D_{p,i}$ only tracks variant-specific fatalities and is not influenced by the death rate determined through the life span L.

$$\dot{D}_{p,i} = \gamma_i I_{p,i} \left(1 - \lambda_i + (1 - \lambda_i)\psi \frac{R_{p,tot}}{N_p}\right) \tag{16}$$

As mentioned above, we assume that only symptomatic cases can die from the disease and that individuals gain cross-protection by recovering from a previous infection, which drastically lowers the lethality of the virus.

This leaves us with only the cross-immunity derivative to be defined in the next subsection.

3.3 Mutation and cross-immunity model

A small number of genes seem to be particularly important to understand the direction of viral evolution. For instance, for influenza type H3, more than one third of the mutations in the HA1 domain are limited to merely 18 out of 329 codons (*1*). Similarly, only a small number of non-synonymous mutations in SARS-CoV-2 are driving an immune escape (*36*).

Thus, we base the cross-immunity between strains on the influenza reference model (*1*). Each variant is accompanied by a simulated RNA strand divided into $AC = 12$ codons. A codon consists of 3 nucleotide bases. Each codon can be translated to an amino acid using an RNA codon table. An cross-immunity $f(d_{i,j}) \in [\theta_0, 1]$ between two strains i and j is then derived by comparing their amino acid sequences. Any site with differing amino acids increments the difference $d_{i,j} \in [0, 12]$.

With $f(d) = 1$ for $d_{i,j} = 0$ perfect cross-immunity is achieved among equal sequences. Then, until a threshold n_t the cross-immunity is determined by θ_1 giving $f(d_{i,j}) = \theta_1$ for $d_{i,j} >= n_t$. Below n_t the cross-immunity is given by $f(d_{i,j}) = \theta_1 + (\theta_0 - \theta_1)(d_{i,j} - n_t)/(AC - n_t)$ for $0 < d_{i,j} < n_t$ (*1*).

Thus, the genetic distance $d_{i,j}$ serves as a proxy for the probability of immune escape and gives rise to the cross-immunity differential $\dot{C}$:

$$\dot{C}_{p,i} = \sum_j f(d_{i,j})\tau F_{p,j} - \frac{C_{p,i}}{L} \tag{17}$$

Its influx is determined by the decay of non-specific immunity. The life expectancy L governs its decrease due to natural death.

In contrast to the influenza reference model, our model collects cross-immunities per virus variant and not per agent. An individual which was previously infected by two distinct strains i, j with $d_{i,j} = 12$ would, therefore, retain a cross-immunity of $2\theta_0$ versus θ_0 in the influenza reference model (*1*).

To counter this difference, we reduce θ_0 from 0.25 to 0.15 and verify its fit through reproduction of the influenza reference model's results (see Supplementary Text). For further details on the cross-immunity model refer to (*1*) from which we also derive our parameters listed in Table S1.

Like in the SARS-CoV-2 reference model (*14*) the strain-specific parameters of the epidemiological model are also subject to mutation. For each parameter r_i of variant i a random variable $x \sim N(0, \sigma), x \in (-1, 1]$ determines its mutated value in an offspring j: $r_j = r_i(1 + x)$. For simplicity and in accordance with the influenza (*1*) and the SARS-CoV-2 (*14*) reference models, we disregard any potential functional links between the disease characteristics and the antigenic profile, and simulate their mutations independently.

3.4 Interventions

A key component to test our hypothesis is the implementation of non-pharmaceutical interventions against the viral spread. In public discourse, such interventions are commonly known as social distancing. During the Covid-19 pandemic, varying social distancing policies have been implemented. Furthermore, individuals also engaged in voluntary social distancing beyond government mandates. The common intent of social distancing was to reduce the contact rate β. Our intervention model

aims at capturing this effect regardless of concrete policies. It is rooted in the assumption that interventions are a reaction to the perceived pathogenic severity in the recent history.

More precisely, interventions start as soon as the number of deaths per capita in the previous 10 days exceeds a certain threshold $\tau_b \geq 0$. In this case, the contact rate $\beta_{W,i}$ of all strains i is multiplied by a reduction factor $\beta_b(t) \in [0, 1]$ in Equation 4, hence uniformly reducing the contact between all population groups.

We further assume that interventions can be phased out altogether after some cut off day d_c, after which $\beta_b(t) = 1$ for $t > d_c$.

3.5 Vaccination

A prominent feature of the Covid-19 pandemic was also the rapid development of vaccines. In many countries, interventions were only phased out after vaccines became widely available. We incorporate this mechanism by adding a simple vaccination policy that assumes a certain fraction δ_v of the population can be vaccinated per day. Only the susceptible and non-specifically immune population is assumed to be able to benefit from vaccination. Respecting the already vaccinated population in V_p, the vaccinated population increase $\dot{V}_p$ in a patch p is then determined by:

$$\dot{V}_p = \begin{cases} \delta_v N_p & \text{for } \delta_v N_p \leq S_p + F_p - V_p \\ max(0, S_p + F_p - V_p) & \text{for } \delta_v N_p > S_p + F_p - V_p \end{cases} \tag{18}$$

The vaccination effect is two-fold: First, it delivers a short-lived non-specific immunity boost and second a long term cross-immunity equal to an immunity against the initial strain. We assume vaccinations start 100 days before the end of social distancing. Thus, a targeted vaccination rate of 1 % per day could protect the entire population. Similar vaccination rates were reached by the State of Israel during the Covid-19 pandemic (*37*) and serve as an optimistic upper bound. We conduct a sensitivity analysis to investigate the effect of failing to achieve such a rate.

3.6 Metrics of structural tree complexity

We use several metrics from multiple disciplines to evaluate the evolution of the virus in our simulations quantitatively. First, like in the influenza reference model (*1*), we use pairwise diversity

PD as a measure of genetic diversification. We use a weight function f_w that yields the current infected hosts of a variant to find the diversity at each time step in the simulation.

$$PD = \frac{\sum_i \sum_j f_w(i) f_w(j) d_{i,j}}{\sum_i \sum_j f_w(i) f_w(j)} \tag{19}$$

The pairwise diversity utilizes the genetic distance $d_{i,j}$ and sums it over all pairs of strains i, j to calculate a weighted mean.

To quantify the effects of interventions on the phylogenetic structure, we utilize tree-specific entropy metrics. As a metric from the field of computer science, we use tree degree entropy H^* (*33*). This metric can be rooted in the literature and serves as a good proxy for structural complexity (*32, 33*).

$$H^* = \sum_c \frac{n_c}{n} \log \frac{n}{n_c} \tag{20}$$

H^* corresponds to a Shannon entropy using the ratios of the number of tree vertices with $c \in \mathbb{N}$ children n_c and the total number of vertices n as probability values. A weighted tree degree entropy variant H^*_w can be defined using a weight function f_w and substituting n and n_c in 20 for sums of vertex weights n_w and $n_{w,c}$:

$$n_w = \sum_{i \in n} f_w(i) \tag{21}$$

$$n_{w,c} = \sum_{i \in n_c} f_w(i) \tag{22}$$

In addition, we utilize an entropy metric rooted in biology. The phylogenetic entropy index H_p (*34*) is derived from the Shannon index over tree branches b using the tree branch length $l(b)$ and the proportion of individuals, in our case infected hosts, in the community $p(b)$ stemming from b as:

$$H_p = -\sum_b l(b) p(b) \ln p(b) \tag{23}$$

It combines the quantification of variant distances with the structural component inherent in the phylogenetic tree (*38*).

Lastly, we introduce a metric targeted at the key feature of phylogenetic linearity. We base this metric on the impact $\iota(i)$ a variant i has within the phylogeny. Our quantification can use a weight function $f_w \geq 0$, with $f_w(i) > 0$ for at least one variant. This impact $\iota(i)$ is defined as the weight of a variant and the sum of the impacts of all its descendants:

$$\iota(i) = f_w(i) + \sum_{j \in c(i)} \iota(j) \tag{24}$$

where c covers all children of variant i. The mean linearity Λ can then be based on the ratio of maximum child impact and total strain impact:

$$\Lambda = 2 \left(\frac{\sum_i f_w(i)(f_w(i) + \max_{j \in c(i)} \iota(j))}{\sum_i f_w(i)(f_w(i) + \sum_{j \in c(i)} \iota(j))} - \frac{1}{2} \right) \tag{25}$$

In effect, Λ is defined in the interval $(0, 1]$. A linearity of 1 represents a tree where each node, barring a single final leaf, bears exactly 1 child. In this case Λ is independent of the weight function f_w. In contrast, a low linearity can be achieved by distributing the weight uniformly among a large number of children.

Structural changes in the phylogenetic tree are inevitably bound to its growth. Thus, there is a correlation between tree growth and changes in our metric values. In tables S3 and S4 we quantify the correlation between the tree size and our metrics both in absolute terms as well as in terms of their dynamics. We find that H_p exhibits the strongest correlation to tree size and its change. The other indicators show a low correlation to the absolute size. The tree degree entropy H^* is the only metric that is negligibly correlated to the tree size growth. H^* also exhibits the highest correlation to the peak non-specific immunity (see Fig. 2) indicating its usefulness in quantifying structural tree complexity. The strong correlation of H_p to tree size growth explains its response to the pandemic phase in Fig. S12 which is the inverse of what would be expected of a measure of structural complexity.

3.7 Data analysis and figure details

All simulations for the main results were carried out 100 times with varying random number generator seeding. Data analysis was performed using Python version 3.12.9 and the packages "seaborn" version 0.13.2 (*39*) and "statannotations" version 0.7.1 (*40*) for the statistical plotting and analysis. Box-plots show the quartiles as box limits and the median as center line within the box. The whiskers denote the entire range of the data, covering all outliers. Line-plots depict the standard deviation as a shaded area around the graph line that represents the mean. Violin-plots depict the median and quartiles as lines through the distribution representation. To compare the different phases of the pandemic and its effects on the tree measures statistically Welch's t-test was utilized. Linear regressions through the collected data points were performed using the "scipy" package version 1.15.2 which alongside the regressions also calculated the values for Pearson's r stated in the relevant figures (*41*).

4 Results

Fig. 2 pane **D** shows the relationship between the peak non-specific immunity and the mean of different metrics of phylogenetic complexity using a fixed intervention threshold of $\tau_b = 10^{-6}$. Each circle depicts the mean result for the first two simulated years in one scenario combining an intervention effect parameter β_b and individual pseudo random number generator seeding. The colors indicate β_b values. All entropy metrics show higher complexity as they increase when the peak non-specific immunity decreases. This decrease in non-specific immunity is a direct cause of the introduction of non-pharmaceutical interventions as can be gathered from the last sub-pane in Fig. 2 pane **D**. Interestingly, the pairwise diversity [used by (*1*)] is not as sensitive to this increase in complexity witnessed by the entropy metrics. The linearity metric introduced below shows a similar response to the entropy scales. These results highlight the importance of taking the structural properties of phylogenetic trees into account.

As visible in pane **A** of Fig. 2, we can categorize the intervention scenarios of different parameter combinations of τ_b and β_b into three broad categories. These categories lead to characteristic peaks in the non-specific immunity compartment during the interventions and thereafter (see pane **B** of Fig. 2). To gather a more detailed look into the complex dynamics we present an illustrative

scenario for each category. Fig. 2 pane **C** shows the first four years of these examples including one without interventions and one with vaccination. An example scenario for category C uses a low threshold $\tau_b = 10^{-6}$ combined with a strong intervention effect $\beta_b = 0.1$. In this category of intervention scenarios, the population is largely unaffected from infections. When the interventions are artificially lifted after two years, however, the number of infections spikes due to a lack of built-up immunity. This spike can be mitigated by introducing vaccinations. However, this prevents a variant-induced bottleneck effect and increases the phylogenetic complexity. The example scenario for category B raises the threshold to $\tau_b = 2.5 \cdot 10^{-5}$. It represents a category where the interventions lead to a rather slow building up of immunity within the population and the interventions can be subsequently phased out. In contrast, the example scenario for category A retains the threshold $\tau_b = 10^{-6}$ but lowers the intervention effect to 60 % ($\beta_b = 0.4$). The population gains immunity quickly and the interventions are phased out. It bares the most resemblance to the scenario without any interventions.

We refrain from reporting the mortality that would be associated with the particular scenarios, as it is not the focus of our analysis and our model is not well-equipped to do so. In particular, we do not consider the non-linear impact of the number of concurrent infections on the number of deaths that may happen when the health care systems are overwhelmed [see (*42*)].

5 Empirical validation

As a final step in our investigation, we want to explore our hypothesis empirically by analyzing the empirical development of viral pathogenic evolution before, during and after the Covid-19 pandemic. Our hypothesis is that non-pharmaceutical interventions increase the structural complexity of phylogenetic trees of viral pathogens such as influenza and SARS-CoV-2 due to the fact that the population builds less widespread non-specific immunity. Due to the global nature of the response, the pandemic presents a unique opportunity to study the impact of non-pharmaceutical interventions on pathogenic development.

We study empirical phylogenetic trees from the nextstrain platform (*43*). To stabilize the temporal development, we use the ready made trees based on 12 years of influenza's evolutionary history for HA and NA genes respectively. For SARS-CoV-2 we use all available data, since it is only

available for a much shorter period at the time of this writing. We prune the trees to their state up to each month from January 2018 to December 2024. Compared to our synthetic phylogenetic trees, the structure of these empirically observed trees cannot be determined with full certainty. In particular, the early evolution of SARS-CoV-2 can only be inferred. This explains the initial low linearity and higher tree degree entropy we find for its phylogenetic tree in Fig. S12 **A**.

Furthermore, the evolution of SARS-CoV-2 had been affected by non-pharmaceutical interventions almost from the beginning. This is why SARS-CoV-2 represents a poor case to study our hypothesis empirically and why we instead use a simulation approach to investigate this 'what if?' type question.

Since we have no information about the infection numbers for specific variants in these trees, we can only use the unweighted versions of our metrics.

During the pandemic, tree linearity Λ and the tree degree entropy H^* suggest a higher evolutionary complexity in SARS-CoV-2 compared to the influenza samples. This matches reported observations in the literature (*2, 3*). The metrics seem to respond to a seasonal development. This is particularly evident in our linearity metric during the winters of 2020 and 2021. Qualitatively, H^* bears the most resemblance to our simulated vaccination example scenario. This metric also shows the highest correlation to the peak non-specific immunity in our simulation study. The phylogenetic entropy index H_p shows a lower value for SARS-CoV-2 compared to influenza viruses. This is likely due to its comparatively low tree size. In both entropy metrics, the development of SARS-CoV-2 appears to align with the phylogenetic trees of influenza viruses after the pandemic.

The analyzed trees represent a filtered subset of the full evolutionary picture. Thus, a quantitative comparison of synthetic and observed phylogenetic trees is difficult. Nonetheless, our analysis unveils characteristic changes in the pathogenic evolution (see Fig. S12 **B**). We test whether the evolution is significantly different using monthly data on the change in each metric.

Starting in March of 2020, for influenza trees, the development of the phylogenetic entropy index H_p and our linearity metric stagnates. Furthermore, it returns, for almost all lineages that we consider, to its previous trend (or closer to the previous trend). This is an important finding which suggests that this is indeed due to the pandemic. These changes in the trend are significant until the end of 2022. For H3N2, this effect is less prominent. A significant difference for the trend in H_p could only be established until October of 2021. Furthermore, the linearity metric shows

no significant differences in post-pandemic development for the NA gene. Contrary to our other findings, the NA gene of the influenza B victoria lineage exhibits an even stronger stagnation in its development after the pandemic. Further research is necessary to explain these differences and to investigate whether they will persist.

We observe a strong effect of the Covid-19 pandemic on the tree degree entropy metric H^*. From June 2020 until the end of 2022 we observe a significant mean increase of entropy in contrast to a decrease before and thereafter. There are no significant differences between the pre- and post-pandemic development, which again supports our hypothesis that the pandemic is the driving force behind these results.

To conclude, our empirical analysis uses ready-made phylogenetic trees that are pruned to specific dates. It shows that the phylogenetic tree structure metrics used in our analysis have a potential to shed light on viral pathogenic evolution. Importantly, we find that the pandemic seems to have affected influenza's evolution, albeit this is likely of a temporary nature. While our empirical analysis is not able to establish a causal link, our simulations point to social distancing as a driver of these changes. Our findings highlight the essential link between social and natural systems and may act as an important stepping stone for further research.

6 Discussion

In this paper, we developed a novel model capturing epidemiological and evolutionary dynamics to study the impact of non-pharmaceutical interventions on the (non-)linearity of viral evolution and viral heterogeneity. Our model draws on two agent-based models developed for influenza (*1*) and the Covid-19 pandemic (*14*), respectively and features a large number of virus variants that are endogenously created by a mutation process. However, in contrast to the two reference models, our approach captures the viral spread using equations which are based on the well-known SIR approach. Hence, we do not model each individual host explicitly, but only each virus variant. This approach drastically improves the efficiency of our evolutionary-epidemiological framework, allowing us to compute a large number of simulations.

We showed that in a baseline calibration derived from the two models, non-pharmaceutical interventions, such as lockdowns or voluntary social distancing, may trigger a more complex

and non-linear evolution as an unintended consequence. While the complexity of viral evolution becomes lower after lifting social distancing measures without vaccination policies, vaccinations are able to preserve these dynamics over the longer term.

The driving assumption behind our results is that we assume that – similar to influenza [see (*1, 11*)] – an infection with one variant creates short-lived broad cross-immunity against all other variants (in addition to long-term cross-immunity against the same and highly similar variants). This mechanism causes large infection waves to become evolutionary bottlenecks that inhibit other variants to spread, therefore limiting the complexity of viral evolution. We thus illustrate a potential explanation for the observed complexity in the SARS-CoV-2 phylogeny.

However, our results do not preclude explanations to the origins of Omicron based on hidden human or animal reservoirs or immunocompromised human carriers, but merely illustrate that they are not necessary to explain a highly non-linear viral evolution as observed in SARS-CoV-2. Instead, we show that such a non-linearity can emerge endogenously through interventions common during the Covid-19 pandemic.

Finally, a short empirical exploration using publicly-available data from the nextstrain platform (*43*) using our metrics seem to broadly support our hypothesis that social distancing can have an impact on phylogenetic complexity and linearity: First, the evolution of influenza has become significantly less strongly linear during the pandemic for all six influenza trees analyzed, and more strongly linear after the pandemic again for four out of six trees (we only observe a further significant decrease after the pandemic for one tree).

Subsequent work is necessary to expand our understanding of the concurrent evolution of competing viral pathogens and their interaction with socio-economic responses.

Importantly, our analysis should not be viewed as a basis for normative judgment about how to *best* design social distancing policies, as this would require a much more complex model that takes – among others – the health and economic impacts of each scenario better into account.

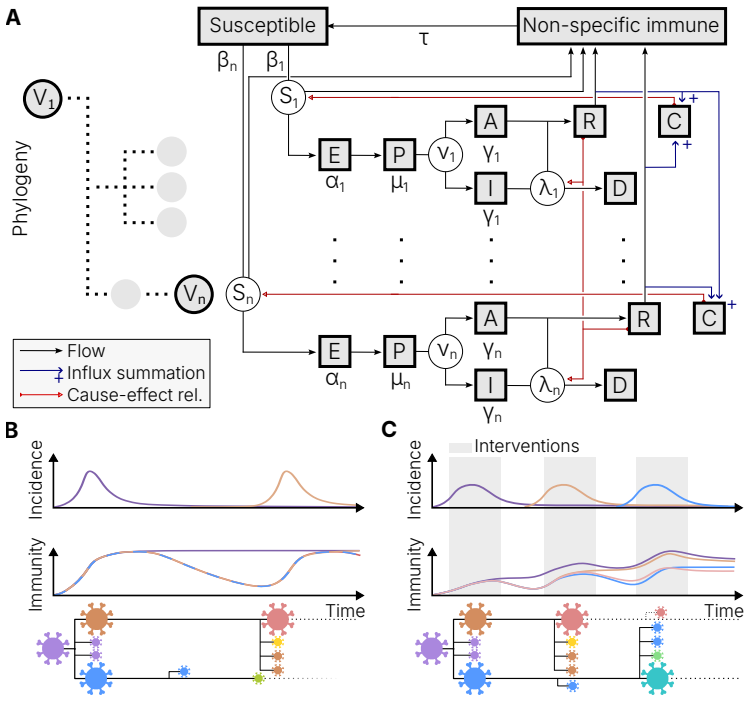

Figure 1: **Flow diagram of our model (A) and schematic representation of the non-specific immunity component's effect (B & C).** (**A**) Population flow diagram for our model. The compartments for the susceptible and non-specific short-lived immune population are not specific to variants. All other compartments are variant-specific. Black arrows show the population flow between compartments. Red arrows point out cause-effect relationships. The state variable for the cross-immune population *C* is excluded from the mass balance as indicated by the plus signs next to the blue influx arrows. (**B**) Schematic of the suppression of diversification as an effect of the non-specific immunity effect [see (*1*)]. Infection waves caused by a virus variant boost the immunity against all variants regardless of antigenic distance, thus, creating a bottleneck for diversification. (**C**) Schematic showing how non-pharmaceutical interventions may drive the mutational dynamics towards a more complex pattern. In contrast to **B** the interventions curb both the amplitude of infection waves and their effect on widespread perfect cross-immunity.

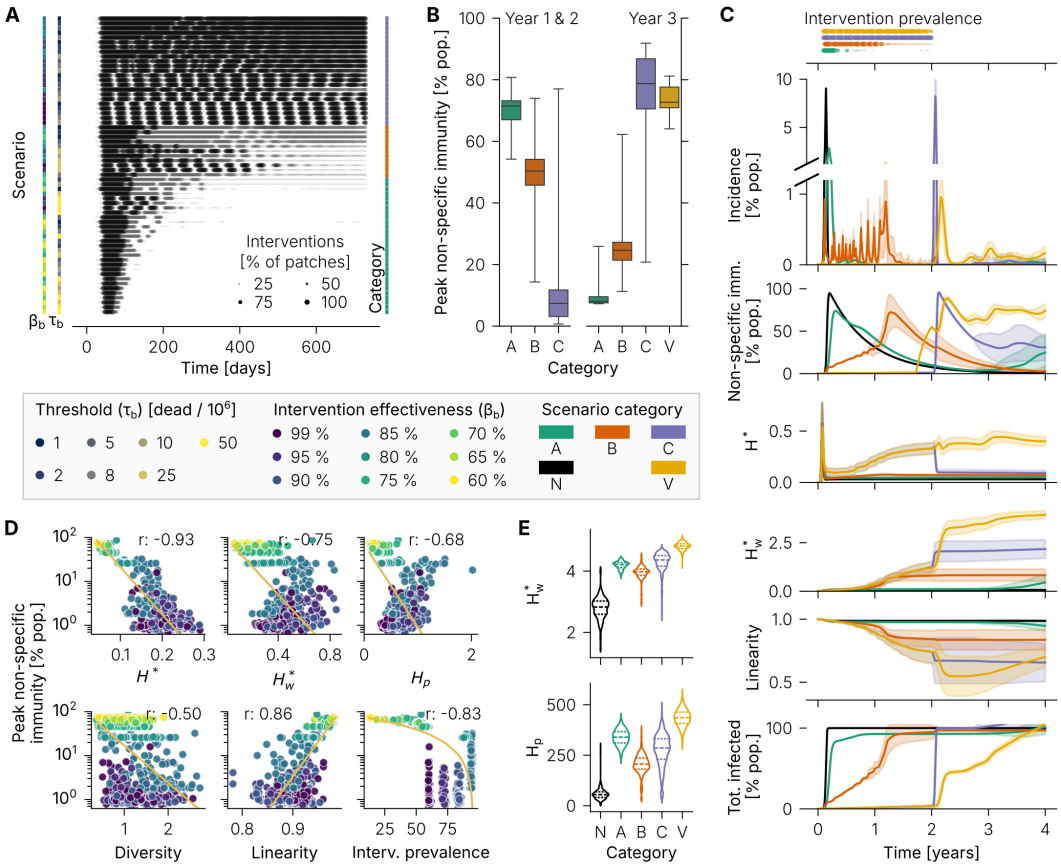

Figure 2: **Non-pharmaceutical interventions drive the evolutionary complexity of our modeled viral pathogen.** (**A**) The intervention scenarios can be categorized into three types differentiated by the emerging intervention prevalence. In category A short initial interventions are followed by a non-restrictive phase. In category B the interventions slowly fade out while in C they last. (**B**) The scenario categories exhibit characteristic non-specific immunity peaks during the intervention phase and the year thereafter. A fourth category V shows the effect of vaccinations on the non-specific immunity peak after interventions are lifted in category C. (**C**) The development of five example scenarios in the initial four years reveals more qualitative differences between the categories. Counter-intuitively, vaccinations in our model support a lasting high complexity by preventing a variant-induced bottleneck effect and introducing evolutionary pressure to escape cross-immunity. (**D**) The correlation between the peak non-specific immunity and various metrics of phylogenetic complexity in 100 simulation runs with 9 different values for β_b. Interventions and peak non-specific immunity have a linear relationship. High mean entropy metrics during interventions and our linearity metric correlate with a lower non-specific immunity. (**E**) Two entropy metrics in the 15th simulated year show a long-lasting vaccination effect on evolutionary complexity.

Acknowledgments

We want to thank the University of Graz for providing a local compute cluster and access to the Vienna Scientific Cluster (VSC). The computational results presented have been achieved [in part] using the Vienna Scientific Cluster (VSC).

Funding sources: This research was funded in whole, or in part, by the Austrian Science Fund (FWF) [P 35228]. For the purpose of open access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Author contributions: S.P. and P.M. both contributed to the equally to the idea and the writing of this manuscript. S.P. designed and implemented the computer model, analyzed the results and prepared all figures.

Competing interests: There are no competing interests to declare.

Data and materials availability: The model result data is provided alongside the model source code which can be used to generate it (*44*). Data form the empirical analysis of phylogenetic trees was retrieved from the *NextStrain* platform (*43*).

Code availability: The entire code of the model will be available under a permissive open-source license upon publication (*44*). We use the *nix* build tool for reproducibility.

References

1. Ferguson, N. M., Galvani, A. P. & Bush, R. M. Ecological and immunological determinants of influenza evolution. *Nature* **422**, 428–433 (2003). Number: 6930 Publisher: Nature Publishing Group.

2. Tegally, H. *et al.* Emergence of SARS-CoV-2 Omicron lineages BA.4 and BA.5 in South Africa. *Nature Medicine* **28**, 1785–1790 (2022). Number: 9 Publisher: Nature Publishing Group.

3. Markov, P. V. *et al.* The evolution of SARS-CoV-2. *Nature Reviews Microbiology* **21**, 361–379 (2023). Number: 6 Publisher: Nature Publishing Group.

4. Bansal, K. & Kumar, S. Mutational cascade of SARS-CoV-2 leading to evolution and emergence of omicron variant. *Virus Research* **315**, 198765 (2022).

5. Pulliam, J. R. C. *et al.* Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa. *Science* **376**, eabn4947 (2022). Publisher: American Association for the Advancement of Science.

6. Roemer, C. *et al.* Sars-cov-2 evolution in the omicron era. *Nature microbiology* **8**, 1952–1959 (2023).

7. Wei, C. *et al.* Evidence for a mouse origin of the SARS-CoV-2 Omicron variant. *Journal of Genetics and Genomics* **48**, 1111–1121 (2021).

8. Sun, Y., Lin, W., Dong, W. & Xu, J. Origin and evolutionary analysis of the SARS-CoV-2 Omicron variant. *Journal of Biosafety and Biosecurity* **4**, 33–37 (2022).

9. Cele, S. *et al.* SARS-CoV-2 prolonged infection during advanced HIV disease evolves extensive immune escape. *Cell Host & Microbe* **30**, 154–162.e5 (2022).

10. Feng, S. *et al.* Evidence of SARS-CoV-2 convergent evolution in immunosuppressed patients treated with antiviral therapies. *Virology Journal* **21**, 1–13.

11. Yan, L., Neher, R. A. & Shraiman, B. I. Phylodynamic theory of persistence, extinction and speciation of rapidly adapting pathogens. *eLife* **8**, e44205.

12. Park, S. W., Cobey, S., Metcalf, C. J. E., Levine, J. M. & Grenfell, B. T. Predicting pathogen mutual invasibility and co-circulation. *Science* **386**, 175–179 (2024).

13. Raharinirina, N. A. *et al.* SARS-CoV-2 evolution on a dynamic immune landscape 1–9.

14. Mellacher, P. Endogenous viral mutations, evolutionary selection, and containment policy design. *Journal of Economic Interaction and Coordination* **17**, 801–825 (2022).

15. Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Physical review letters* **86**, 3200 (2001).

16. Eletreby, R., Zhuang, Y., Carley, K. M., Yağan, O. & Poor, H. V. The effects of evolutionary adaptations on spreading processes in complex networks. *Proceedings of the National Academy of Sciences* **117**, 5664–5670 (2020).

17. Espinoza, B. *et al.* Coupled models of genomic surveillance and evolving pandemics with applications for timely public health interventions. *Proceedings of the National Academy of Sciences* **120**, e2305227120 (2023).

18. Sood, M. *et al.* Spreading processes with mutations over multilayer networks. *Proceedings of the National Academy of Sciences* **120**, e2302245120 (2023).

19. Gozzi, N. *et al.* Estimating the impact of covid-19 vaccine inequities: a modeling study. *Nature Communications* **14**, 3272 (2023).

20. Delli Gatti, D., Reissl, S. & Turco, E. V for vaccines and variants. *Journal of Evolutionary Economics* 1–56 (2023).

21. Basurto, A., Dawid, H., Harting, P., Hepp, J. & Kohlweyer, D. How to design virus containment policies? a joint analysis of economic and epidemic dynamics under the covid-19 pandemic. *Journal of Economic Interaction and Coordination* **18**, 311–370 (2023).

22. Azizi, A., Kazanci, C., Komarova, N. L. & Wodarz, D. Effect of human behavior on the evolution of viral strains during an epidemic. *Bulletin of Mathematical Biology* **84**, 144 (2022).

23. Ye, Y. *et al.* Equitable access to covid-19 vaccines makes a life-saving difference to all countries. *Nature human behaviour* **6**, 207–216 (2022).

24. Gog, J. R. & Grenfell, B. T. Dynamics and selection of many-strain pathogens. *Proceedings of the National Academy of Sciences* **99**, 17209–17214.

25. Marchi, J., Lässig, M., Walczak, A. M. & Mora, T. Antigenic waves of virus–immune coevolution. *Proceedings of the National Academy of Sciences* **118**, e2103398118.

26. Kermack, W. O. & McKendrick, A. G. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **115**, 700–721 (1927). URL `https://www.jstor.org/stable/94815`. Publisher: The Royal Society.

27. Chen, J., Wang, R., Wang, M. & Wei, G.-W. Mutations strengthened sars-cov-2 infectivity. *Journal of molecular biology* **432**, 5212–5226 (2020).

28. Plante, J. A. *et al.* The variant gambit: COVID-19's next move. *Cell Host & Microbe* **29** (2021).

29. Yuan, M. *et al.* Structural and functional ramifications of antigenic drift in recent sars-cov-2 variants. *Science* **373**, 818–823 (2021).

30. Amicone, M. *et al.* Mutation rate of SARS-CoV-2 and emergence of mutators during experimental evolution. *Evolution, Medicine, and Public Health* **10**, 142–155 (2022).

31. Nobusawa, E. & Sato, K. Comparison of the Mutation Rates of Human Influenza A and B Viruses. *Journal of Virology* **80**, 3675–3678 (2006).

32. Ganczorz, M. Using Statistical Encoding to Achieve Tree Succinctness Never Seen Before. In *37th International Symposium on Theoretical Aspects of Computer Science (STACS 2020)* (Leibniz, 2020).

33. Jansson, J., Sadakane, K. & Sung, W.-K. Ultra-succinct representation of ordered trees with applications. *Journal of Computer and System Sciences* **78**, 619–631 (2012).

34. Allen, B., Kon, M. & Bar-Yam, Y. A New Phylogenetic Diversity Measure Generalizing the Shannon Index and Its Application to Phyllostomid Bats. *The American Naturalist* **174**, 236–243 (2009).

35. Brauer, F. Epidemic Models with Heterogeneous Mixing and Treatment. *Bulletin of Mathematical Biology* **70**, 1869–1885 (2008).

36. Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology* **19**, 409–424 (2021). Number: 7 Publisher: Nature Publishing Group.

37. Mathieu, E. *et al.* A global database of COVID-19 vaccinations. *Nature Human Behaviour* **5**, 947–953 (2021).

38. Mouchet, M. A. & Mouillot, D. Decomposing phylogenetic entropy into α , β and γ components. *Biology Letters* **7**, 205–209 (2011).

39. Waskom, M. L. seaborn: statistical data visualization. *Journal of Open Source Software* **6**, 3021 (2021). URL https://doi.org/10.21105/joss.03021.

40. Charlier, F. *et al.* Statannotations (2022). URL https://doi.org/10.5281/zenodo.7213391.

41. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**, 261–272 (2020).

42. Walker, P. G. *et al.* The impact of covid-19 and strategies for mitigation and suppression in low-and middle-income countries. *Science* **369**, 413–422 (2020).

43. Hadfield, J. *et al.* Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics (Oxford, England)* **34**, 4121–4123 (2018).

44. Plakolb, S. & Mellacher, P. Model source code and result data for "complex evolution of viral respiratory pathogens can be an unintended consequence of social distancing" (2025). URL https://dx.doi.org/10.5281/zenodo.14864344.

45. Liu, Y. & Rocklöv, J. The effective reproductive number of the Omicron variant of SARS-CoV-2 is several times relative to Delta. *Journal of Travel Medicine* **29**, taac037 (2022).

46. Ewald, P. W. Evolution of virulence. *Infectious Disease Clinics of North America* **18**, 1–15 (2004).

Supplementary information Supplementary Information is available for this paper:

Supplementary Text

Figs. S1 to S12

Tables S1 to S5

References *(35-46)*

Materials and Correspondence Correspondence and requests for materials should be addressed to Simon Plakolb (simon.plakolb@ibv.uio.no).

Extended data and supplementary information for Complex evolution of viral respiratory pathogens can be an unintended consequence of social distancing

Simon Plakolb*, Patrick Mellacher

*Corresponding author. Email: simon.plakolb@ibv.uio.no

This PDF file includes:

Supplementary Text

Figures S1 to S12

Tables S1 to S5

Other supporting materials for this manuscript include the following:

Model source code S1

7 Supplementary Text

7.1 Calculation of contact rate values

Our model is based on an influenza reference model (*1*) to describe the presence of low phylogenetic diversification in influenza strains. This reference model utilizes an agent-based approach where all individuals of the population are simulated individually. These agents are distributed uniformly within $M = 20$ geographically distinct patches. This method is computationally expensive, if the simulated number of people is large. We thus simplify this model to carry out the simulations in a more efficient way: We assume homogeneous mixing within each geographical patch. This allows us to model the spread of the virus using the well-known SIR equations. However, since we are interested in tracking viral evolution, we also want to explicitly model a potentially large number of virus variants. To do so, we use an agent-based approach where we model each variant as an agent (instead of each person).

To account for the differences in the modeling approach, we have to carefully adjust the contact rates in our model. The influenza reference model differentiates between local transmission with a contact rate determined by β_L, mixing within a patch with β_W and contact between patches with β_B. The reported basic reproduction number for transmission within local groups is $R_0 = 5$. Within a geographical patch, transmission is less likely with $R_0 = 0.4$ and transmission in between patches is given with $R_0 = 0.02$.

We find our values for $\beta_{W,0}$ and the homophily parameter h using equations S1 and S2 (*35*):

$$\beta_{W,0}(h + \frac{1-h}{M}) = \frac{5.4}{\gamma_0^{-1} + \mu_0^{-1}} \tag{S1}$$

$$\beta_{W,0}\frac{1-h}{M} = \frac{0.02}{\gamma_0^{-1} + \mu_0^{-1}} \tag{S2}$$

The resulting parameter values are listed in Table S1. To gain an intuition for these equations, one can rewrite their left-hand side (LHS) with $M = 20$. Equation S1's LHS then reads as $\frac{1}{20}\beta_{W,0}(19h + 1)$. It can be understood as the local transmission component. In a completely homophilic $h = 1$ scenario $\beta_{W,0}$ equals the right-hand side (RHS) of S1. For equation S2 the LHS can be rewritten as

$\frac{1}{20}\beta_{W,0}(1-h)$. In a scenario with homogenous mixing between all patches ($h = 0$), $\beta_{W,0}$ corresponds to the RHS of S2. Intuitively, the homophily parameter h regulates the balance between both extremes.

7.2 Reproduction scenarios

Despite our methodological changes, our model can reproduce the results of the influenza reference model [see (*1*)]. As listed above [see Methods in Section 3], all but one parameter are set to the values of the reference model. Our method necessitates some changes to the immunity model. Therefore, as described in our methods section, the cross-immunity's minimal effect θ_0 was reduced.

Fig. S1 shows our reproduction of the influenza reference model's results in two key scenarios that follow the methodology of its source (*1*). A setting with low mutability and *only* cross-immunity is compared to a scenario paring high mutability with an additional non-specific immunity. Both scenarios were evaluated with each using 50 simulation runs over 30 virtual years.

Pane **A** of Fig. S1 shows the weekly incidence in the first geographical patch. Much like in the influenza reference model's associated report, the scenario excluding the non-specific immunity peaks between $10,000$ and $15,000$ inhabitants in the first 15 years. The incidence for the scenario including the non-specific immunity is lower. In the second half of the simulation, it peaks around $3,000$ inhabitants, again resembling the reference result (*1*). The reference's units are reported to be per $100,000$ inhabitants (*1*). We find that this may be an error in these units, as the percentage of infected inhabitants appears unrealistic. In any case, our model produces a similar difference between both scenarios. More importantly, our observations for pairwise diversity match those reported for the influenza reference model (*1*).

In pane **B** of Fig. S1 a lack of non-specific immunity proves to lead to rapid diversification. Meanwhile, the added non-specific immunity component limits diversification drastically, despite a tenfold increase in mutability. In both cases, the range of the pairwise diversity closely resembles that reported for the influenza reference model (*1*).

Equations of the reproduction model

Our model includes a pre-symptomatic phase and asymptomatic cases. In contrast, the influenza reference model (*1*) uses a single mode of symptomatic infection. Furthermore, it does not consider the possibility of a lethal infection outcome. Therefore, some equations listed in our manuscript's methods section can be simplified to reproduce this model.

The population that is cross-immune to variant i and currently within the compartments specific to any variant can be calculated without the $P_{p,i}$ and $A_{p,i}$ compartments:

$$C_{p,i,I+F} = \sum_j (E_{p,i} + I_{p,i} + F_{p,i})(1 - f(d_{i,j}))\frac{C_{p,i}}{N_p} \tag{S3}$$

We want to reiterate the intuition for this expression. For a variant i, the sum of the cross-immune population needs to be scaled to reflect a good estimate. We assume a reasonably good mixing within all compartments. Then, the fraction of the cross-immune individuals within the total population $\frac{C_{p,i}}{N_p}$ can be used for the variant-specific cross-immunity compartments. However, the likelihood of a cross-infection is inversely correlated to the cross-immunity. Thus, an additional scaling by $1 - f(d_{i,j})$ is necessary.

Due to the reduction of infection compartments the normalized contagious population $I_{tot,p,i}$ simplifies to a mere fraction of symptomatic individuals within the population:

$$I_{tot,p,i} = \frac{I_{p,i}}{N_p} \tag{S4}$$

Since there are no more pre-symptomatic and asymptomatic cases, the exposed population in $E_{p,i}$ flows directly into $I_{p,i}$:

$$\dot{I}_{p,i} = \alpha_i E_{p,i} - \gamma_i I_{p,i} - \frac{I_{p,i}}{L} \tag{S5}$$

Likewise, the absence of lethal consequences simplifies the compartment outflow of $\dot{I}_{p,i}$. As in all compartments, the life expectancy L governs the rate of natural death. Besides the natural deaths, only γ_i regulates the outflow by determining the mean infection duration. Consequently, the net recovery rate reduces to only the net flux out of the infected compartment:

$$\dot{R'}_{p,i} = \gamma_i I_{p,i} \tag{S6}$$

For the scenario excluding the non-specific immunity the model can be further simplified. In this scenario the non-specific immunity compartment F_p and the associated, variant-specific state variable $F_{p,i}$ can be excluded from the model. The component for the variant-specific susceptible population $S_{p,i}$ can then be rewritten as:

$$S_{p,i} = S_p - C_{p,i} + C_{p,i,I} \tag{S7}$$

It uses the cross-immune population currently infected by other variants $C_{p,i,I}$. It is an alternative to $C_{p,i,I+F}$ that excludes the non-specific immunity compartment F_p:

$$C_{p,i,I} = \sum_j (E_{p,i} + I_{p,i})(1 - f(d_{i,j}))\frac{C_{p,i}}{N_p} \tag{S8}$$

Excluding F_p also changes the flux of recovered individuals. The recovered population flows directly into the susceptible compartment, for which the derivative changes to:

$$\dot{S}_p = \frac{N_p}{L} + \left(sum_i \gamma_i I_{p,i} - b_{p,i} S_{p,i}\right) - \frac{S_p}{L} \tag{S9}$$

Without the non-specific immunity, recovered individuals directly contribute to the cross-immunity. Therefore, the cross-immunity derivative is now given by:

$$\dot{C}_{p,i} = \sum_j f(d_{i,j})\gamma_j I_{p,j} - \frac{C_{p,i}}{L} \tag{S10}$$

The influenza reference model (*1*) deals with the evolution of a pathogen with widespread antigenic adaption in the population. Thus, simulations start "near the single-strain equilibrium" (*1*). To reproduce this, we disable all mutations in the first 100 simulated years to arrive at the (dynamic) single-strain equilibrium. Afterwards, we enable mutations. As in the influenza reference model, in our reproduction model the epidemiological parameters are not affected by the mutation but constant. The parameters used in the reproduction model are listed in Table S2.

7.2.1 Our model under equilibration

The substitution of the host-agents with ODEs is not the only difference the between the influenza reference model (*1*) and ours. Due to the focus on a different, and novel, pathogen our method introduces new compartments and the mutation of epidemiological parameters. In addition, the scenarios in our study start before the single-strain equilibrium establishes itself. Arguably, there is a large gap between the reproduction scenarios and our main study. To bridge this gap, we additionally investigate a scenario of our novel model starting close to the single-strain equilibrium. This scenario uses the main study's base scenario parameters. There are no pharmaceutical or non-pharmaceutical interventions. Note that these parameters do not differ drastically from the reproduction model.

The additional compartments of the epidemiological model exhibit only a minor impact on the population dynamics. The most notable difference to the reproduction scenarios appears to be the quicker establishment of a dynamic equilibrium. This is likely the effect of the extended latent period.

Our scenario includes a non-specific immunity component. Consequently, the pairwise diversity is reduced. For the influenza reference model, one can show that functional constraints on the viral evolution do not significantly impact the results (see Supplementary Material of (*1*)). Using our model, which decouples the stylized RNA mutation from the evolution of epidemiological parameters, we find that only after the first $\sim$ 15 years the parameter mutation starts to impact the pairwise diversity. We, therefore, conclude that the long term effects of the interventions in our model should be investigated at the 15 year mark.

7.3 Sensitivity Analysis

To improve the insight into our model's dynamics, we perform a sensitivity analysis beyond the scenarios shown in Fig. 2. Each parameter combination is evaluated at least 50 times with varying pseudo-random number generator seeding. The sensitivity analysis focuses on the state after 2 and 4 simulated years. Some parameter combinations, however, cover a time span of 15 or 30 years to investigate the long term dynamics.

7.3.1 Intervention model analysis and long term results

Our scenarios cover a large range of the two main intervention model parameters. In our main results figure, we show the correlation of various metrics with the peak size of the non-specific immunity population for a varying intervention effect β_b. The same analysis can be done for the intervention threshold τ_b. Fig. S3 shows these correlation plots using a fixed intervention effect β_b of 90 %. Most R^2 values are lower compared to the main results pane **D** in our manuscript. Only the pairwise diversity metric appears to react stronger when varying τ_b with a fixed β_b.

Due to a lack of space in our main results figure, only two entropy metrics show the results after 15 years. In Fig. S3 we show the state of other key metrics in our study throughout the 15 th year. We find that the pairwise diversity reaches a similar equilibrium in all scenarios. The size of the population with non-specific immunity is, however, much larger in the scenario without interventions. This is reflected in the entropy metrics which are lower in these scenarios. Contrary to our intuition, without the interventions the linearity is lower than in any of the intervention scenarios on average, even though the variance is higher. This could be a latent effect of the delayed second infection wave.

In Fig. S4, we take a look at some exemplary phylogenetic trees. All example trees show the 500 variants with the largest recovered or infected population and their ancestry. The blue trees show the phylogeny after 15 years without the interventions (**A**) and for a scenario of category C (**B**). Significantly more diversification events are visible in the blue tree shown in **B**. This explains the difference in the Λ values. Table S5 lists the associated metrics for the aforementioned trees. It is easy to see, that the entropy measures and the tree linearity focus on the entire structure of a phylogenetic tree. In contrast, pairwise diversity (*PD*) quantifies the temporary state. This is especially noticeable in the values for the black trees which show the state after 30 years. An additional example of a scenario with vaccinations is shown in pane **C**. At its right end this tree shows more concurrently circulating variants the other examples in **A** and **B**. It exhibits a pairwise diversity of 7.14. In comparison, this is significantly higher. This difference is not reflected in the rest of our used metrics. This points to the complementary nature of these metrics and to the benefit of their use in unison.

Our simulations end after a period of 30 years. The effects of the parameter mutation on the

evolutionary patterns already explored in Fig. S2 can be revisited here. Fig. S4 **D** suggests that the scenario lacking interventions may be more resilient to the effect of the parameter mutation. In contrast to the other scenarios, its pairwise diversity is less elevated. This can also be seen in the example trees (**A** - **C**) that show the phylogeny after 30 years.

7.3.2 Analysis of epidemiological and mutational characteristics

The main driver for the evolutionary dynamics in our focus is the intervention model. However, there is a complex interplay between it and the other model parameters. We extend our analysis to cover the parameters that had to be changed from the source values or were newly introduced. First, we focus on the state after 2 years, *i.e.*, the state immediately preceding the end of the interventions. The model component parametrized based on the influenza reference model (*1*) differs in two parameters from the source. The contact rate β_W uses a homogeneous mixing based estimate. The minimal effect of the cross-immunity θ_0 was adapted to the differences in the cross-immunity model.

The scenario category C covers the scenarios with the strongest effect of the non-pharmaceutical interventions on our results. Thus, we model the interventions for this category with a threshold of $\tau_b = 10^{-6}$ and an intervention effect β_b of 90 %. All other parameters are chosen in line with Table *S*1 of our manuscript with the exception of the singular parameters we vary for each analysis. Our analysis in Fig. S5, shows that our main results are resilient to a change in either of the two parameters. We attribute the small differences to a feedback from the intervention model.

Our model simulates the viral mutation in a two-fold process. The stylized RNA mutation gives rise to new virus variants. This RNA mutability is governed by δ. A new variant mutates its individual epidemiological parameters that it inherited from its parent variant. This parameter mutation distance follows a (limited) normal distribution that depends on σ. The expected effect of a lower RNA mutability is a reduced evolutionary complexity. Indeed, Fig. S5 shows this effect. It also shows that reducing the parameter mutation range has a strong impact on the evolutionary dynamics. By setting $\sigma = 0$ we effectively disable the parameter mutation. The result is a strong reduction of the pairwise diversity and all measures of entropy.

The effect of the parameter mutation cannot be attributed to a singular parameter. As shown in Fig. S6 **A** the absence of the mutation of individual parameters exhibits only a minor effect. The

biggest difference can be observed in a combination of the parameters directly affecting the basic reproduction number R_0: β, γ and μ. This, as shown in **C**, is not due to a rapid increase of the basic reproduction number. Rather, a small differentiation in virulence is sufficient, albeit vital, to produce the observed effect on the viral diversification. The simulated mutations cause R_0 to reach a value, shown in **D**, that is within the observed range for the Omicron variant of SARS-CoV-2 (*45*). Fig. S6 **V** depicts examples of the synthetic phylogeny after four years with and **E** without the parameter mutation. As visible, a lack of the parameter mutation leads to an unrealistic diversification where the ancestral variant remains the most potent pathogen across all recurring infection waves.

The lethality of our simulated disease interacts with the intervention model. In Fig. S7 we explore this relationship. We consider the example scenarios for our three intervention scenarios and compare them to versions with no initial lethality and no cross-protection. As expected, we find that the initial survival chance λ_0 has a strong impact on the results. An initially low lethality tends to reduce the phylogenetic complexity. This can be attributed to a reduction of the intervention prevalence. In **A** showing the example of the scenario category A, however, we find the opposite effect. Here, after four years the complexity is increased. The likely explanation is a larger susceptible population left for the variants that are children of the ancestral strain. This exemplifies the complex dynamics that unfold through the interplay of the epidemiological, social and evolutionary systems.

Due to the parameter evolution, a lack of the initial lethality does not entirely remove the interventions. Contrary to our intuition, in our model λ appears to evolve towards a higher lethality (see Fig. S6) if the standard deviation of mutations is very high. This is likely because the initial value of the ability of hosts to survive the disease is close to its upper bound of 1. Our model is not suited to fully unravel the complexities of pathogens within human hosts. Usually, within hosts a trade-off between contagiousness and host health emerges (*46*). What is more, since there are no small host communities in our model, highly lethal variants may not go extinct as quickly as expected. Nevertheless, our viral variants are subject to selective pressure due to the interventions and the susceptible population size. Our results could point towards a trade-off on a global epidemiological level where the lethality lowers the cross-immunity through differentiation but inhibits the transmission due to the interventions. The induced delay between recovery and susceptibility due to the non-specific immunity may limit the evolutionary impact of the lethality on the susceptible host population size.

In other words, the non-specific immunity in combination with the lethality dependent interventions may change the landscape of the evolutionary stable strategies such that pathogens benefit from (a small) lethality.

The cross-protection against lethal infection outcomes ϕ shows less impact on our results. Its damping effect on the lethality can be explored in Fig. S7 **B**. By removing the cross-protection entirely, we find an increased complexity after four simulated years. With cross-protection, the population's adaption to a novel pathogen leads to less lethal cases. In turn, the intervention prevalence drops. Removing this dynamic in the scenario category B leads to ongoing interventions and an infection spike after their discontinuation. This drives the example towards the scenario category C, which exhibits increased late stage complexity metrics.

7.3.3 Effects after four years

The intervention discontinuation after two years causes a significant change of the resulting phylogenetic complexity. In scenario category C, the infection peak that follows the end of interventions is especially large. Therefore, we also analyze the sensitivity two years after the end of the non-pharmaceutical interventions. Fig. S8 depicts the same analysis as above but two years after the intervention discontinuation.

The infection wave following an intervention cessation appears to have little effect on the qualitative analysis of our model's sensitivity. The changed epidemiological model parameters remain rather inconsequential to our results. Both mutational parameters continue to drive the resulting complexity. While the entropy and linearity metrics may exhibit latent structural effects of early differences, the pairwise diversity reflects a lasting impact of these parameters after four years. The unweighted tree degree entropy H^* exhibits an elevated value for the scenario without the parameter mutation. This points to diversification events in child variants where the resulting sub-variants remain unfit to compete against the ancestral strain. These events are likely the result of the infection spike after the intervention discontinuation.

7.3.4 Effects in other scenario categories

The previous analysis focused on the scenario category C. In the other categories, the effect of the non-pharmaceutical interventions on the phylogenetic complexity differs significantly. Hence, we

extend our analysis to two example scenarios from the categories A and B. We use the same example scenarios as in our main study. For the scenario category B we set the intervention threshold to $\tau_b = 2.5 \cdot 10^{-5}$ and the intervention effect β_b to 90 %. The example of the category A uses an intervention threshold of $\tau_b = 1 \cdot 10^{-6}$ and an intervention effect β_b of 60 %.

Figures S9 and S10 show the effect of various parameters on the non-pharmaceutical intervention phase and thereafter for examples of the scenario categories A and B. Overall, the results remain similar to the previous example of the category C. Again, the effect of the mutation parameters is more pronounced. Especially the category A is characterized by a reduction of interventions. Similarly, the effect of the parameter mutation on the phylogenetic complexity metrics is also reduced, albeit still visible.

In both categories, the effect of the adapted epidemiological parameters becomes more pronounced. This is especially evident in the results after four years. An initial R_0 decrease appears to leave more room for a later differentiation. As the interventions become less prevalent in category B and, especially, A, the minimal cross-immunity parameter θ_0 becomes more important for the epidemiological dynamics. Lowering θ_0 increases the resulting evolutionary complexity. This is expected, since this parameter induces a long-lasting non-specific immunity. The larger recovered population in these categories, thus, reduces the mutational complexity. In contrast, a lower recovered population reduces the impact of the cross-immunity on the dynamics in category C (see figures S5 and S8). This emphasizes the fundamental impact of the non-specific immunity on the pathogenic evolutionary complexity.

7.3.5 Vaccination model analysis

We model vaccinations to gauge the possible effect they could have on the progress of the phylogenetic complexity after the repeal of non-pharmaceutical interventions. Our vaccination model is rather simple. A vaccination adds (cross-)immunity against the ancestral strain and a short-lived non-specific immunity. The scenario shown in our main results is optimistic. In it, all citizens that have not yet been infected can be vaccinated at a rate of 1% of the population per day.

In Fig. S11 **B** we also cover less successful scenarios with lower vaccination rates. One scenario is aimed at vaccinating 81 % of the population and reaching the theoretical herd immunity for the initial R_0 value of 5.4. By the time the vaccine can be distributed, the R_0 value has shifted due to

the parameter mutation. In a pessimistic scenario we analyze the effect of a low vaccination rate set to half of that in the optimistic scenario.

Unsurprisingly, the peak infections are lowest in the most effective vaccination scenario. Effects of the vaccination rate on the phylogenetic complexity are visible. In our model, the pharmaceutical interventions reduce the peak of the non-specific immunity that may follow a repeal of the non-pharmaceutical interventions. As a result, the bottleneck effect is reduced.

Such a bottleneck effect may instead be inducible by distributing the vaccine at a very high rate. However, our optimistic vaccination scenario already uses the highest achieved vaccination rate during Covid-19 (*37*).

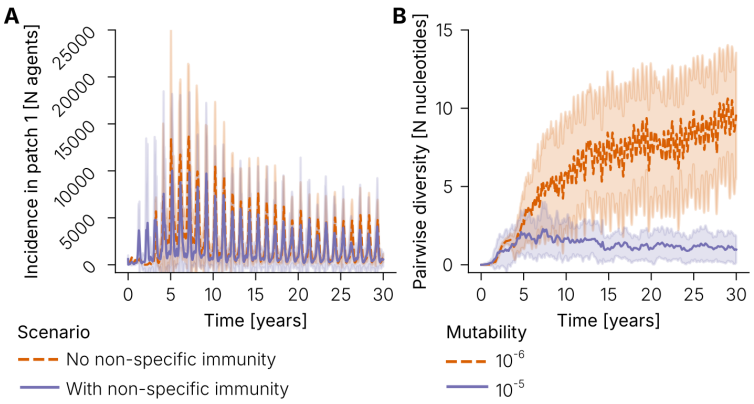

Figure S1: **Reproduction of the influenza reference model's results [see (*1*)] using our model with the assumption of perfect mixing withing geographical patches.** Both panes show the mean and standard deviation of 20 simulation runs comparing a scenario with low mutability but no non-specific immunity component with a scenario using high mutability and a non-specific immunity. (**A**) Incidences observed in the first geographical patch matching the reported numbers closely. (**B**) Pairwise nucleotide diversity weighted by the case abundance. The effect of the non-specific immunity component is evident in the low diversity exhibited despite a higher mutation rate in the scenario with non-specific immunity.

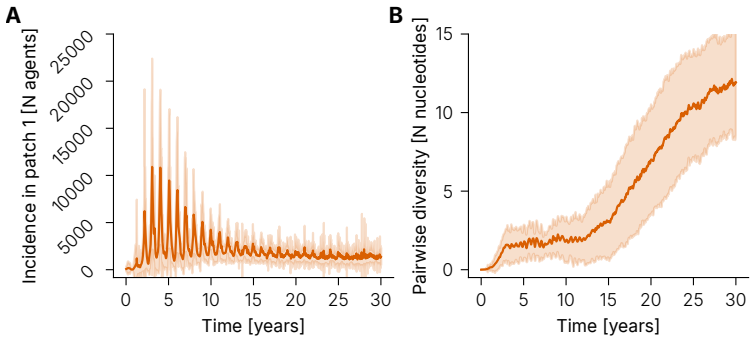

Figure S2: **Starting simulation runs close to the single-strain equilibrium using our adapted model without interventions shows a similar evolution to the influenza reference model [see (*1*)] until approximately year** 15**.** (**A**) The seasonality effects in the first geographical patch's incidence fade over time. (**B**) After some 15 years the effect of the parameter mutation drives our results towards a higher pairwise diversity. As shown in our manuscript, it is vital to assess the impact of the parameter mutation for a novel pathogen. However, our model does not set any limits to the evolution of the epidemiological parameters beyond their defined range. We, therefore, conclude that our long term analysis should focus on the model state after 15 years.

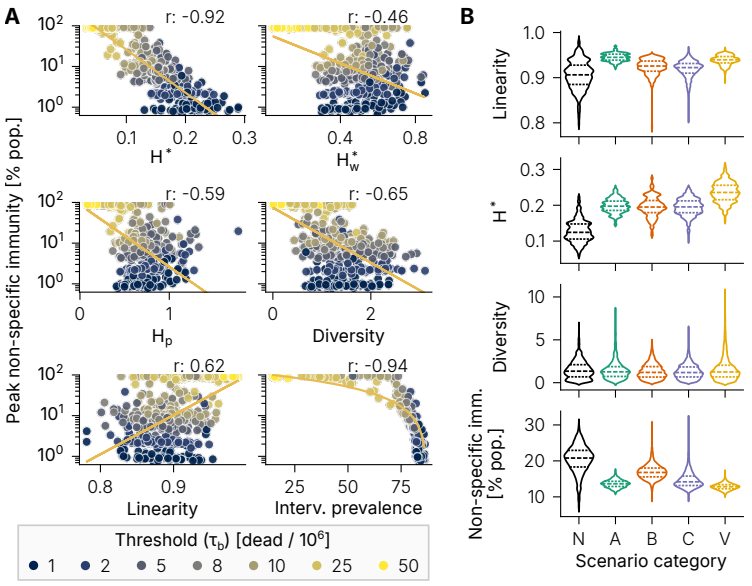

Figure S3: **The same correlation analysis exercised for the intervention effect β_b in our manuscript's main results can be extended to the intervention threshold τ_b.** (**A**) The correlation of most metrics is not as strong for this intervention parameter. Only the pairwise diversity metric gives a stronger response than when varying the intervention effect which was set to 90 % for this graph. (**B**) The state in year 15 of variables that were omitted in our main results for spatial reasons. The long term effects of the interventions seem to increase our linearity metric. The entropy metrics remain elevated. A possible explanation for this phenomenon is the increasing number of diversification events and the comparatively late emergence of differentiation in the scenario without interventions. The pairwise diversity appears to find a similar equilibrium in all scenarios. The population with non-specific immunity is larger in the scenario without interventions. In Fig. S4, we find that this difference reduced after 30 simulated years.

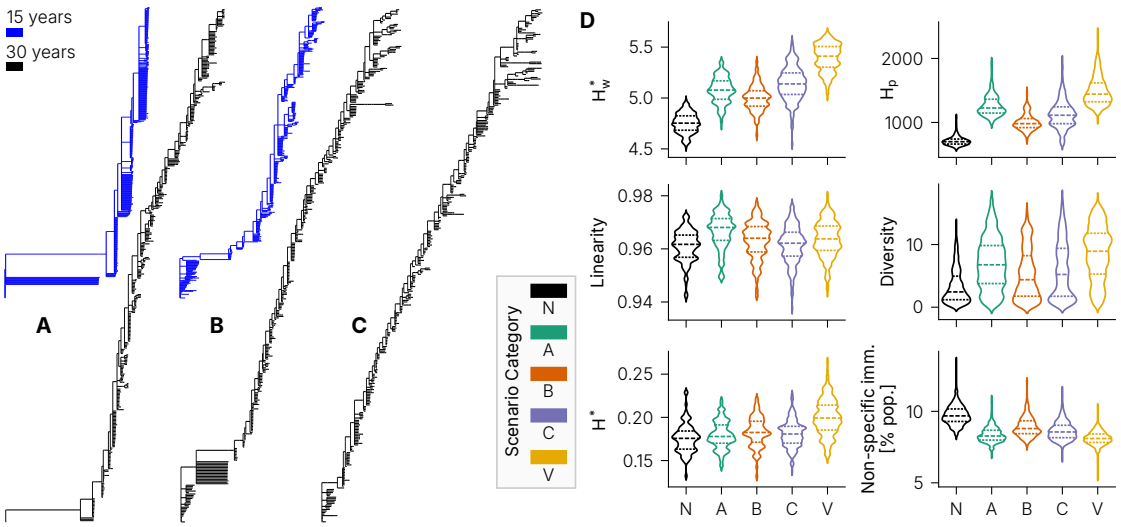

Figure S4: **In the 30th simulated year the complexity remains higher for scenarios with interventions.** There seems to be an increasing effect of the parameter mutation that drives diversity in these scenarios. The scenario without interventions appears to remain more resilient to this effect. (**A**) Example trees after 15 (blue) and 30 (black) years for a scenario without interventions. (**B**) Synthetic phylogenetic trees for the scenario category C after 15 (blue) and 30 (black) years. (**C**) An example tree for a vaccination scenario with 1 % vaccination rate after 30 years. For all phylogenetic trees we only show the 500 variants with the largest recovered population and their ancestors. For the sake of comparison, all trees are generated using the same random number generator seed. The associated metrics for these trees can be found in Table S5. (**D**) The state of our focused metrics after 30 simulated years. The entropy metrics are mostly differentiated through latent effects of early diversification events. Pairwise diversity is driven higher in scenarios with interventions due to the epidemiological parameter mutation.

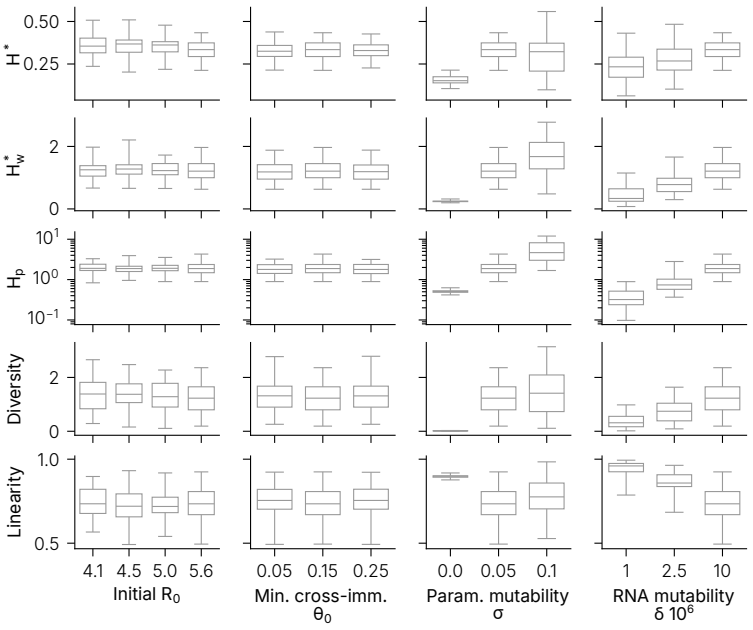

Figure S5: **A sensitivity analysis of the intervention scenario with $\tau_b = 10^{-6}$ and an intervention effect β_b of 90 % shows a large impact of changes in mutational dynamics on our main results.** The effect of the adapted epidemiological parameters is comparatively small. The investigated scenario was chosen as an example of the scenario category C. The R_0 values refer to the transmission within geographical patches. The fourth column shows the three mutability values for the stylized RNA also used by the influenza reference model [see (*1*)]. A lower mutability leads to less variant offspring and, thus, a reduction of the diversity and the opportunity for the epidemiological parameters to diverge. The parameter mutation, shown in the third column, only takes effect in new variants. Effectively disabling the parameter mutations by setting $\sigma = 0.0$ strongly reduces the effect of the interventions on all used metrics.

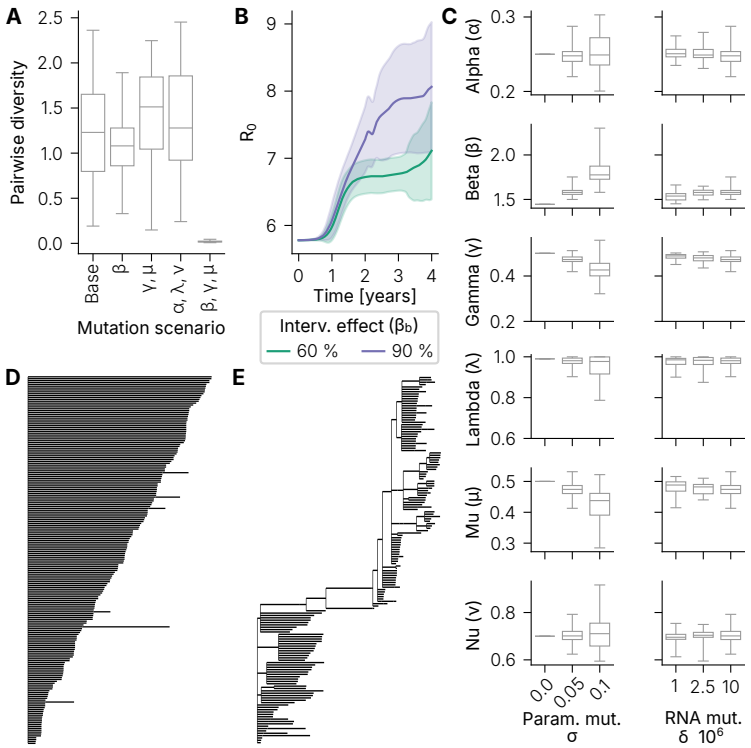

Figure S6: **The consequences of removing the parameter mutation cannot be explained by a drastic impact on one singular parameter.** Removing the mutation of singular parameters does not show the same effect on the pairwise diversity. (**A**) Results of removing the mutation in various combinations of epidemiological parameters. A *Base* scenario allows all the epidemiological parameters to mutate. Only the absence of the mutation in all parameters affecting the basic reproduction number (β, γ and μ) leads to a large drop of the pairwise diversity. (**B**) In the example scenario for the category C, the R_0 values mutate towards the range that could be observed for the Omicron variant of SARS-CoV-2 (*45*). (**C**) Despite the large effect, the amplitude of these mutations is not excessive. A visual comparison of the resulting phylogeny without (**D**) and with (**E**) the parameter mutation underlines its effect and importance. In each tree we show the 200 variants with the most infections and recovered hosts and their ancestors in each tree.

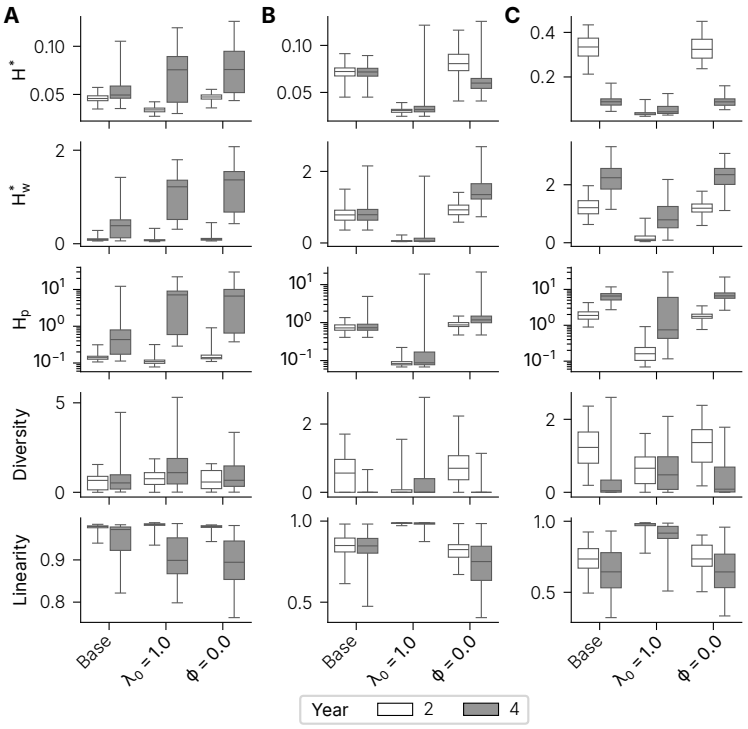

Figure S7: **Raising the initial survival chance λ_0 to 100 % decreases the evolutionary complexity by reducing the intervention prevalence.** (**A**) Result of a scenario without lethality and one without cross-protection for an example of the scenario category A. (**B**) The example scenario for the scenario category B, where the impact of lethality is especially visible. (**C**) An example of the category C, where this effect is also pronounced. In contrast, (**A**) shows that a lowered lethality can increase the pairwise diversity. We attribute this to a smaller impact of the large first infection wave on the susceptible population available to the variants emerging later. Overall the cross-protection ϕ against lethal infection consequences appears to have less impact. In (**B**) its removal slightly increases the observed mutational complexity after four years. This is likely due to the damping effect effect of the cross-protection on the lethality. The lack of this effect leads to prolonged interventions and a larger impact of their discontinuation after two years. The base scenarios refer to the category examples with interventions but unchanged lethality parameters.

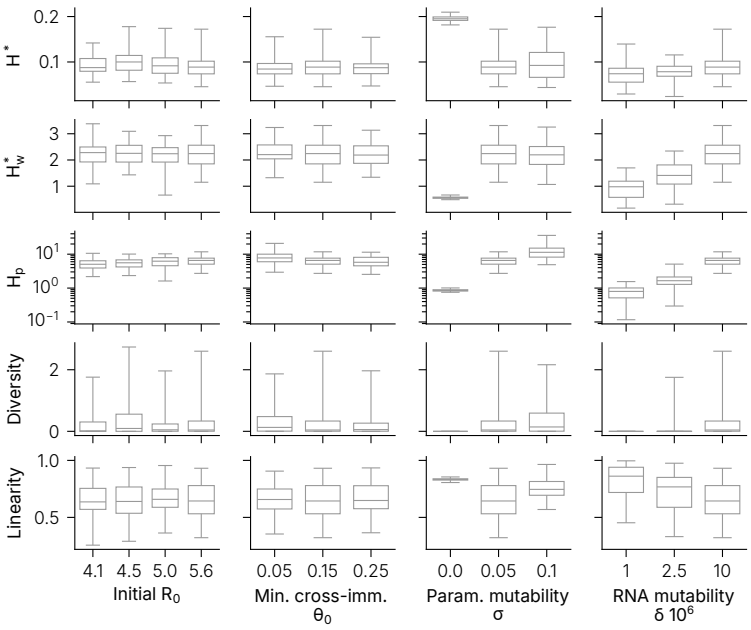

Figure S8: **Two years after the discontinuation of the non-pharmaceutical interventions there is no drastic change in the results of our sensitivity analysis.** The adapted epidemiological parameters continue to show little effect on the main results. In contrast, the mutational parameters still show a strong effect on the results. Lowering the mutation rate, both for the parameter mutation as well as for the stylized RNA mutation, reduces the evolutionary complexity. This effect remains somewhat visible in the pairwise diversity which is less prone to the latent effects of early dynamics. However, the unweighted tree degree entropy H^* shows an increased value for the lowest parameter mutability. This hints at speciation events where child variants spawn offspring still unfit in their competition against the ancestral strain.

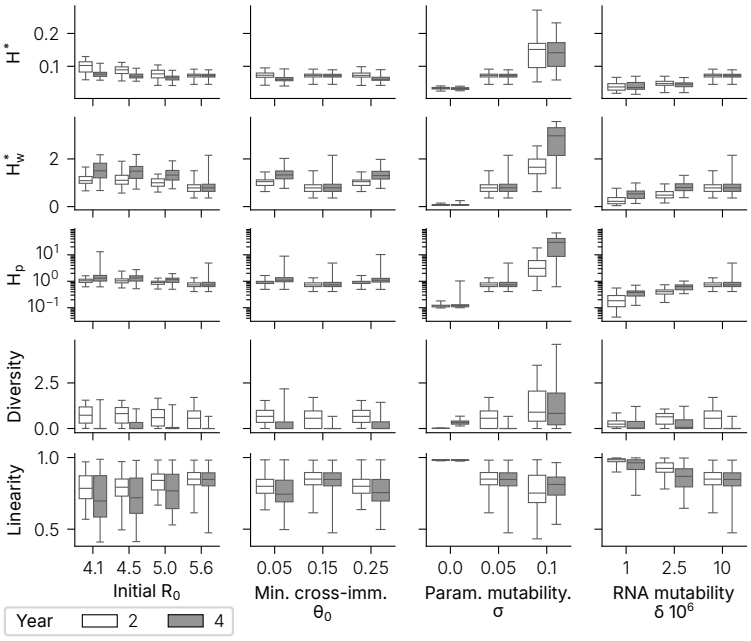

Figure S9: **In the scenario category B the observed effects after** 2 **years of non-pharmaceutical interventions remain mostly unchanged.** A slight reduction of the parameter mutation's effect can be attributed to the lower intervention prevalence. This is also (quantitatively) evident in the results for a higher mutability. Lowering the initial R_0 value seems to leave slightly more room for a later differentiation. A reduced minimal cross-immunity also impacts the long term result. This points to the reduced bottleneck effect of the non-specific immunity component in this scenario.

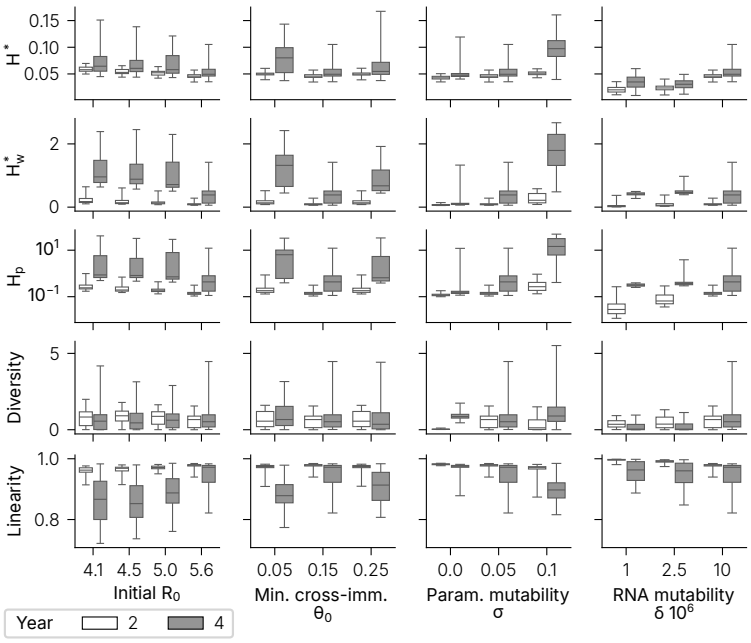

Figure S10: **For the scenario category A we observe a reduced impact of the mutational parameters on the phylogenetic complexity.** This can be attributed to a reduction of the non-pharmaceutical intervention prevalence. The effect still remains visible, pointing to the lasting effect of the initial interventions. The adapted epidemiological parameters show a noticeable effect after the intervention phase. Like in Fig. S9, a reduction of the initial R_0 value may leave more space for a later differentiation. In the category A, the cross-immunity component's effect is even more pronounced. Reducing the minimal cross-immunity parameter θ_0 increases the phylogenetic complexity after four years.

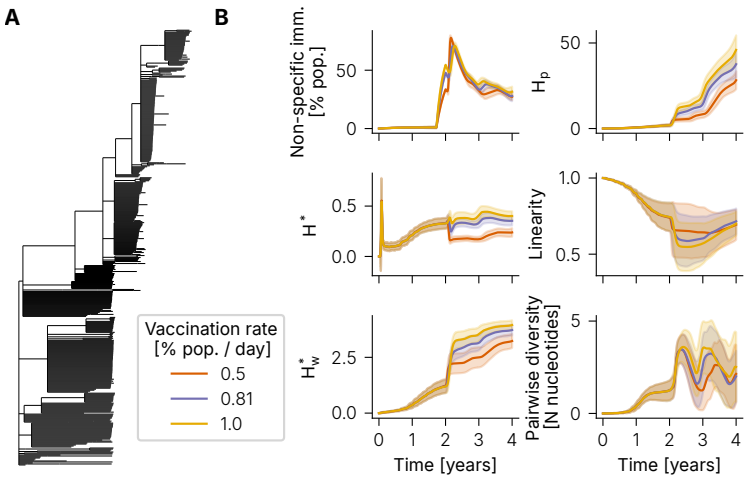

Figure S11: **Different vaccination rates affect the phylogenetic complexity by reducing the effective peak size of the non-specific immunity.** (A) An example of a synthetic phylogenetic tree produced by a scenario with the highest vaccination rate. The vaccination mitigates a variant-induced bottleneck effect and, thus, creates an equal playing field for all variants. This increases the evolutionary complexity as it is quantified by our metrics. (B) The temporal development in the first for years in three vaccination scenarios. The reduced bottleneck effect can be traced in all indicators. A lower vaccination rate leads to a higher entropy, a slight increase of the pairwise diversity and a decreased linearity. The vaccination rates were chosen based on three scenarios. In the optimistic scenario, the population is vaccinated at a rate of 1 % per day. Not all citizens can be vaccinated due to some having been infected recently. An additional scenario was chosen where the theoretic herd immunity should be reached at 81 % of vaccinated citizens. Finally, the lowest rate is set to half of the optimistic scenario in an attempt to model an insufficient vaccination progress.

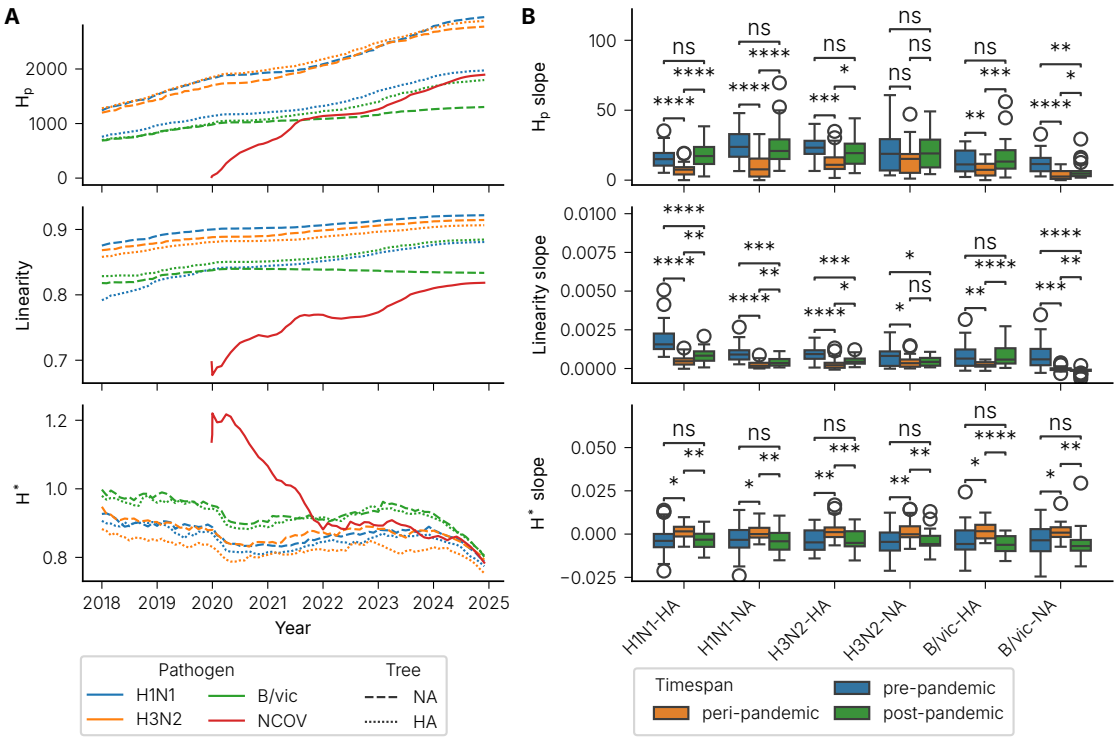

Figure S12: **Observed phylogenetic trees of influenza exhibit characteristic changes in the development of entropy and linearity during the pandemic.** (**A**) Temporal development of our complexity metrics applied to phylogenetic trees of influenza viruses in comparison to SARS-CoV-2 covering pre-pandemic to post-pandemic evolution. Our tree linearity metric Λ can observe the higher complexity of SARS-CoV-2 described in the scientific literature. It also appears to show an effect of the pandemic on the evolution of influenza viruses. Their tree degree entropy H^* exhibits a rapid initial decline with the onset of interventions followed by a subsequent growth that may be the result of reduced non-specific immunity. The tree degree entropy of SARS-CoV-2 peaks during 2020 and starts to align with influenza results after two years. The phylogenetic entropy index H_p of SARS-CoV-2 is lower due to its size. After the pandemic, the development approaches that of influenza. All metrics exhibit seasonal effects with differing prominence. (**B**) The pandemic phase significantly differs from its preceding and following development. However, different temporal selections have to be made. For the phylogenetic entropy index H_p, we find a slowed growth that is likely due to the early, intended effect of interventions following March 2020, most probably due to its correlation with tree size growth (see tables S3 and S4). Our linearity metric also exhibits a slowed growth that lasts from March 2020 to the end of 2021. The tree degree entropy H^* shows a significant growth from June 2020 until the end of 2022. This trend follows an initial decrease that is likely caused by the onset of interventions.

Table S1: The parameters used for the simulation study. Some parameters, such as the intervention effect and threshold vary by simulated scenario and are not listed here.

Name	Value	Description	Source
L	30	Agent lifespan.	(*1*)
M	20	Number of geographical patches.	(*1*)
N_p	$\frac{12{,}000{,}000}{M}$	Number of inhabitants per patch p.	(*1*)
α	0.25	Average latent period of 4 days.	(*14*)
$\beta_{W,0}$	$\frac{289}{200}$	Infectiousness within a patch with $R_0 = 5.4$.	(*1, 35*)
h	$\frac{269}{289}$	Infectiousness across patches with $R_0 = 0.02$.	(*1, 35*)
γ_0	0.5	Average 2 days of (a-)symptomatic period.	(*14*)
λ_0	0.99	Chance of survival.	(*14*)
μ_0	0.5	Average 2 days of pre-symptomatic period.	(*14*)
ν_0	0.7	Fraction of symptomatic infections.	(*14*)
ϵ_p	-0.25 / 0.25	Seasonality, negative if patch $p > M/2$.	(*1*)
ψ	0.99	Cross-protection against death.	(*14*)
τ	$\frac{1}{270}$	Decay rate of non-specific immunity.	(*1*)
σ	0.05	Standard deviation of parameter mutation.	(*14*)
δ	10^{-6} - 10^{-5}	Mutation rate of nucleotide bases per infectious host per day.	(*1*)
θ_0	0.15	Minimum cross-immunity effect.	(See Section 3.2)
θ_1	0.99	Maximum imperfect cross-immunity effect.	(*1*)
n_t	2	Antigenic distance threshold.	(*1*)

Table S2: The parameters used for the reproduction model. All parameters are derived from the influenza reference model [see (*1*)]. The minimal cross-immunity effect θ_0 is adapted to the differences in the the cross-immunity model. The contact rate $\beta_{W,0}$ and the homophily h are derived from the influenza reference model's values as described in Section 7.1.

Name	Value	Description
L	30	Agent lifespan.
M	20	Number of geographical patches.
N_p	$\frac{12{,}000{,}000}{M}$	Number of inhabitants per patch p.
α	0.5	Average latent period of 2 days.
$\beta_{W,0}$	$\frac{289}{200}$	Infectiousness within a patch with $R_0 = 5.4$.
h	$\frac{269}{289}$	Infectiousness across patches with $R_0 = 0.02$.
γ_0	0.25	Average 4 days of infected period.
ϵ_p	-0.25 / 0.25	Seasonality, negative if patch $p > M/2$.
τ	$\frac{1}{270}$	Decay rate of non-specific immunity.
δ	10^{-6} - 10^{-5}	Mutation rate of nucleotide bases per infectious host per day.
θ_0	0.15	Minimum cross-immunity effect.
θ_1	0.99	Maximum imperfect cross-immunity effect.
n_t	2	Antigenic distance threshold.

Table S3: Correlation between metrics and the phylogenetic tree size.

	Tree size	PD	Λ	H^*	H^*_w	H_p
Tree size	1.000000	-0.045520	-0.216350	-0.361966	0.326739	0.536647
PD	-0.045520	1.000000	-0.306345	0.343543	0.316502	0.250051
Λ	-0.216350	-0.306345	1.000000	-0.427979	-0.915686	-0.628884
H^*	-0.361966	0.343543	-0.427979	1.000000	0.424715	0.212379
H^*_w	0.326739	0.316502	-0.915686	0.424715	1.000000	0.837813
H_p	0.536647	0.250051	-0.628884	0.212379	0.837813	1.000000

Table S4: Correlation between temporal metrics differences and the phylogenetic tree size differences.

	Δ Tree size	ΔPD	$\Delta\Lambda$	ΔH^*	ΔH^*_w	ΔH_p
Δ Tree size	1.000000	0.394419	-0.725089	0.102415	0.802855	0.894748
ΔPD	0.394419	1.000000	-0.309248	0.051063	0.391713	0.439709
$\Delta\Lambda$	-0.725089	-0.309248	1.000000	-0.098356	-0.934172	-0.780769
ΔH^*	0.102415	0.051063	-0.098356	1.000000	0.115777	0.107378
ΔH^*_w	0.802855	0.391713	-0.934172	0.115777	1.000000	0.916460
ΔH_p	0.894748	0.439709	-0.780769	0.107378	0.916460	1.000000

Table S5: Associated values for complexity metrics of the example trees shown in Fig. S4.

Year	β_b [%]	τ_b	$\delta_v \left[\frac{\%}{\text{day}}\right]$	Scen. Category	Λ	PD	H_p	H^*	H^*_w
15	0	-	-	N	0.89	0.63	25.12	0.10	2.31
15	99	10^{-6}	-	C	0.96	2.07	266.54	0.18	4.30
30	0	-	-	N	0.97	0.32	646.20	0.16	4.60
30	99	10^{-6}	-	C	0.97	2.50	1004.85	0.16	5.13
30	99	10^{-6}	1	V	0.99	7.14	1367.60	0.19	5.35