An allele-based evolution model of the population spread of SARS-CoV-2

James S. Koopman[1,2,6]

Marisa C. Eisenberg[1,2]

Austin Polanco[5]

Carl P. Simon[1,3,4,*]

ABSTRACT

We present and analyze a dynamic evolution-oriented model of the spread of SARS-CoV-2 in a (homogeneous) population. Unlike previous such models, we use alleles at the various virus sites as our basic building blocks – an approach that we find more natural and more effective in studying the evolution of viral strains. Our model parameters include: transmissibility, immunity levels, immunity waning rates, residual waned immunity levels, infection duration, mutation rates, epistasis, and possible antibody dependent enhancement. To compensate for the model complexity required for all of these features, we restrict to two virus sites with two alleles at each site, allowing us to explore how each of these features interacts with the basic evolutionary processes involved in viral strain changes. Our results show several qualitative patterns of interest in understanding viral strain dynamics, including founder effects of the seeded strain and an interplay between ease of evolution to neighboring strains and immune advantages of complementary strains. Overall, our allele-based approach uncovers insights about virus evolution, including the effects of parameter choices on the timing and patterns of the evolution of new strains, the staying power of the initial (seeded) strain, and the key roles that immunity plays.

1 Center for the Study of Complex Systems, The University of Michigan, Ann Arbor, MI
2 Department of Epidemiology, School of Public Health, The University of Michigan, Ann Arbor, MI
3 Gerald R Ford School of Public Policy, The University of Michigan, Ann Arbor, MI
4 Department of Mathematics, The University of Michigan, Ann Arbor, MI
5 Department of Physics, The University of Michigan, Ann Arbor, MI
6 Deceased, December 2024
*Corresponding Author: cpsimon@umich.edu

Author Summary

We developed an allele-based model of pathogen evolution, coupled with a population-based transmission model that captures a variety of susceptibility, infectiousness, and immune effects. Instead of describing viruses only by their "strain," as most other population-based models do, we represent them through their genetic components—*alleles*—that shape transmissibility and immune escape. This approach provided a natural way to model how new variants arise through mutation and how population immunity influences their success. Our default model had the seeded strain-and its complement (with opposing mutations at both sites in the simplified 2-site 4-strain model we examine here) dominating the pandemic. We examined how the appearance of strains with alleles with different properties would affect this pattern. In some cases, there was minimal effect, e.g., changes in the mutation rate or in the residual immunity after waning; in other cases, e.g., longer infection duration or faster waning immunity, new strains came to dominate. By coupling allele-level dynamics with epidemiological processes, we hope this work begins to bridge laboratory and population scales.

1 Introduction and motivation for an allele-based approach

Laboratory studies have generated an impressive amount of genomic data about the evolution of SARS-CoV-2. These studies of nucleotide changes often focus on mutations that escape immunity or mutations that increase transmissibility. In the process they have also generated hypotheses around the drivers of the evolution of SARS-CoV-2. But as Saad-Roy, Metcalf, and Grenfell [SA22] point out: "To probe likely evolutionary trajectories, it is necessary to understand how immunity intersects with transmission and thus population outcomes, especially in partially immune individuals." This paper takes a systems level point of view to examine the feedback loops between viral evolution and patterns of immunity and transmissibility in the overall population. The model we present and examine is an extremely simplified one, but it still provides insights on the evolutionary dynamics of the pandemic.

It is not just differences in infection forces experienced by individuals that generate new infection patterns. How those forces generate different population patterns of immunity is also important. As also noted by Saad-Roy et al [Sa20]: "relying on the status of infection of an

individual as the main observable during an ongoing epidemic is insufficient to characterize the complex immune landscape generated by the pandemic."

We are able examine these patterns of immunity and transmissibility because we made a major change in how we modeled immunity and transmissibility effects. We assigned such effects to genetic segments within the virus, which we call "alleles," rather than to virus characteristics observed at the *strain* level. We use the term "alleles" here to refer to some portion of the virus gene sequence that may vary (e.g. single nucleotides, codons or genes). Our goal here is to explore how different parts of the genome may experience evolutionary pressure differently. By looking below the whole-genome level, we believe that we can examine more systematically how acquiring immunity against one strain affects immunity against other strains.
Figure 1 encapsulates our simplified virus/site/strain/allele/mutation model.

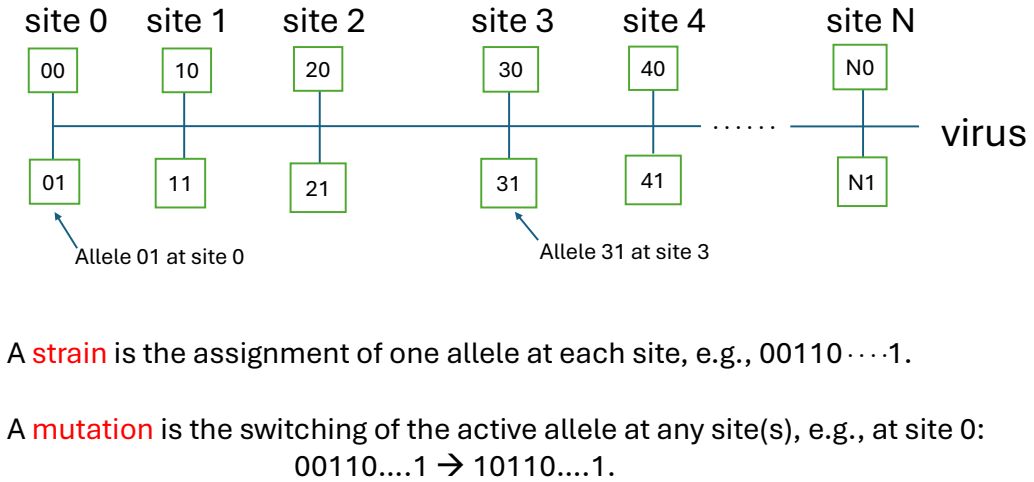

Figure 1. Our virus/site/strain/allele/mutation model, with N sites and two alleles per site.

As an example of the types of evolutionary changes we are modeling, an outstanding feature of coronaviruses is the diversity of 1) receptors on the host cell to which they can attach and 2) the resulting success of that attachment. A mutation at some site could change the type of receptor on

the host cell for viral attachment and therefore the success rate of that attachment. Suppose, for example, that:

- the current strain is 000000 in the notation of Figure 1,
- the two alleles 00 and 01 at site 0 determine different key amino acids for the structure of the receptor-binding subunit S1,
- the activity of allele 01 leads to more successful virus-cell binding than does that of allele 00,
- this more successful binding leads to higher transmissibility of the virus,
- this advantage carries over to all strains for which allele 01 is the active allele at site 0.

This is the kind of phenomenon we have in mind when we examine the effects of increasing the transmissibility (or the immune-avoidance) of allele 01.

We list here some of many reasons why an allele-based model is worth pursuing. First, mutations, the drivers of evolution, take place at the allele level (in this case, truly the single nucleotide level). In the simplest such scenario, a strain mutates to another when one of its component alleles switches to another at a given site on the virus, as illustrated in Figure 1. In the usual strain approach, one strain can mutate to another if the two are adjacent on some pre-assigned network of strains. In any case, allele mutations are the building blocks of strain mutations.

In fact, most dynamical systems models of the population spread of SARS-CoV-2 with which we are familiar have not included mutation. For example, Saad-Roy and colleagues have published a series of path-breaking papers [Sa24b, Sa21a, Sa24a, Sa20, Sa23, Wa21, Es24] on the transmissibility and immune-response of SARS-CoV-2, all focused around the infections and reinfections of a single strain. [Sa23] and [Dys21] added a second strain as an "immune-escape variant." In contrast, Sood et al. [So23] used a "mutation matrix" in their branching-process model to quantify possible mutations between any two strains, but with no underlying evolution and the requirement of full cross-immunity between any two strains. In our models with *multiple* strains, [Getz21] used a mutation matrix and [K-S21] posited a network of strains; in the latter case, one strain could drift to another if the two strains shared an edge in the network.

Believing that the transmissibility and immune-response of new variants play key roles in the timeline of the pandemic, we hope that our multi-scale allele-based model — coupled with population level transmission and immune dynamics -- will open the door to a more realistic approach to the *evolution* of new strains.

We want our modeling program to learn from and hopefully give guidance to the laboratory work that has thus far laid bare the very nature of coronaviruses. An allele-based approach is far closer to the nucleotides and other cellular and subcellular structures on which this lab work has focused. In this vein, this paper addresses the challenge of Saad-Roy, Metcalf, and Grenfell [Sa22] for "epidemiological studies that simultaneously quantify immunity and transmission across populations and at different biological scales … along with the development of cross-scale modeling approaches."

SARS-CoV-2 researchers have shown just how complex the evolution of the virus has been [Ko22, Zh25, Ing21]. Our strain-based models [K-S21] led to a small number of possible paths, with the strongest example being one in which the initial strain mutated to the furthest strain on the strain network, which in turn mutated back to the original strain, and so on. We will see this behavior in our simplest models below. Our allele-based model yields far more complexity because our formulation captures the separate contributions of alleles more realistically.

Finally, an allele-based approach provides explanations and mechanisms that shed light on the evolutionary paths our model uncovers. Such explanations play a central role in the exposition that follows.

Transmission system models need to be classified by both virus characteristics and host characteristics. Our model focuses on the virus characteristics. So, as a first step, all hosts are treated as identical. We model processes that are biologically rather than sociologically determined and ignore all the effects of population heterogeneity and *contact* patterns that could be affecting the factors on which we focus.

However, this allele-based approach still involves huge increases in the details that need to be modeled. In particular, we need to keep track of everyone's allele-based immunity status over time. A compartmental population model, that includes three sites, two alleles per site, and, say, four levels of immunity, would require 4,096 compartments of susceptibles with their immune history. To keep our model tractable, we constrain it to the unrealistic situation of two mutation sites, two alleles per site and four levels of immunity. Even with this simplification, such a model entails a system of nearly 300 differential equations.

Nevertheless, the simplification of two sites each with two alleles allows us to examine details as to what happens within the two sites and two alleles and to perceive more clearly what is driving each phenomenon. This constraint also allows us to include many more relevant parameters, such as:

- **Transmissibility** contribution of each virus allele (our parameter b_i),
- Underlying maximum **strength of immunity** at each immunity site (m_i),
- **Waning rates** of immunity at different stages of waning at each immunity site (w_{ij}),
- Possible **residual immunity** after "completed" waning (nu_i),
- Genetic **mutation rates** in each allele (f_{ij}),
- The **duration of contagiousness** in an infected person (v_{ij}).
- Joint allele effects on immune escape (m_{ij}),
- Joint allele effects on transmissibility (b_{ij}),
- Antibody dependent enhancement (ADE) (d_i).

The joint allele effects formulate **epistasis**, whereby the activity of one allele may be affected by the absence or presence of another allele. In fact, transmissibility is determined by many things that involve the interaction of many alleles, e.g., to make spike proteins, define replication rates.

ADE involves events in which the immune response to previous coronavirus infections generates antibodies that have special capacities to help the virus enter and grow in immune system cells [Ik23; Song23]. When these antibodies attach to a coronavirus, they help coronaviruses enter cells. There must be some immune response to a previous SARS-CoV-2 infection in the past for an ADE event to occur. We note that while typically considered in the more classic ADE example of dengue [Fe99], there is some evidence of ADE in other coronaviruses (SARS-CoV and

MERS-CoV) [Tip14]. All this has raised ADE as a potential concern for SARS-CoV-2 [Sa20, Te20, Ik23, Song23], and so we consider this potential process here.

Changing the impact of any of the nine viral components listed above influences the evolution of the underlying pandemic, as Saad-Roy and colleagues [Sa20] argue:

> "variations in the immune response to primary SARS-CoV-2 infections… lead to markedly different immune landscapes and burdens of critical cases…, and illustrate likely complexities in future CoV-19 dynamics.."

Regarding immune-response and waning parameters m_i and w_{ij} explicitly, they conclude in [Sa24a]:

> "the strength of immunity is a central parameter that shapes medium-term immuno-epidemiological dynamics."

A central role of this paper is to build on their single-strain modeling by including mutations to new strains at the very core of our presentation and by including virus characteristics in our population models.

In line with the central role of immunity in transmission dynamics, our model also keeps track of just how susceptible the population is to each infection strain over time.

In what follows,

- we present the states and the transitions in our model (Part Two);
- we establish a default collection of homogeneous parameter values and initial conditions for our model and analyze the resulting dynamical system (Part Three);
- we examine how the appearance of strains with alleles of different parameter values would affect this default dynamic (Part Four);
- we look for patterns among these variations, contrast our allele-based approach to strain-based approaches, and point out the important role that immunity plays in the whole process (Part Five).

Future steps in this program include: 1) incorporating realistic numbers and characteristics of alleles and mutations into the model, 2) elaborating population patterns of contact that influence

the flow of infection through populations, 3) including varying virulence and long SARS-CoV-2 infection, 4) elaborating the processes in the model to capture knowledge generated by laboratory studies, and 5) relating more directly to the role of RNA dependent RNA polymerase in virus evolution. We elaborate on these future tasks in the discussion at the end of this paper.

2 Methods: Our Compartmental Model of SARS-CoV-2 Evolution

As noted above, we present a model of the coupled dynamics between evolutionary and disease transmission/immunity at the population scale, where we track individual infection and immunity at the allele level rather than the strain level. Below we outline the model structure, beginning with the overall compartmental structure, and then presenting the different characteristics that will distinguish the compartments:

A. States and Compartments

The Compartments:

There are three classes of compartments (states) for the population under study.

- *Susceptibles* S: those who have never been infected with the virus (and have no acquired immunity).
- *Infectives* I: Those currently infected by a particular strain.
- *Recovereds* R: Those previously infected by some strain or strains (with some immunity history), but who are now not infected.

The Virus:

- We assume that the virus has 2 sites: sites 0 and 1. See Figure 2.
- With two alleles at each site:
    - alleles 0-0 and 0-1 at site 0 and
    - alleles 1-0 and 1-1 at site 1.
- We will refer to these four alleles not in the above binary notation, but by their decimal equivalents:
    - alleles 0-0 and 0-1 at site 0 become alleles 0 and 1
    - alleles 1-0 and 1-1 at site 1 become alleles 2 and 3.

- A *strain* is defined by a choice of an allele at each site.
    - Our model has four strains: 02, 03, 12, and 13.

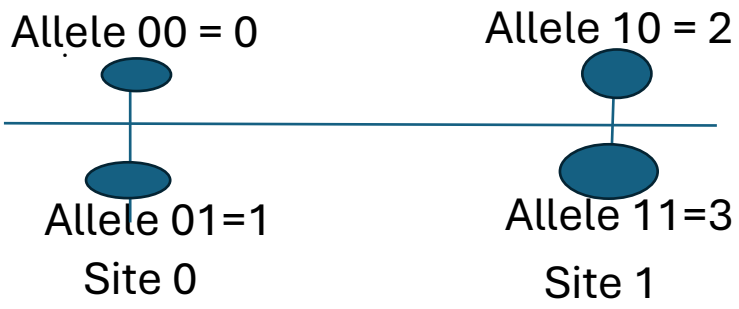

Figure 2. Our simplified 2-site, 2-allele model.

- A strain mutates when one allele is replaced by the other at a given site.
    - For example, a mutation of strain 02 at site 0 mutates strain 02 to strain 12
- Contagion, transmission, and immunity are all modeled at the allele level for every strain in the assumed homogeneously mixing population.

Immunity

Immunity, and its complement susceptibility, plays a central role in our model.

In its current form, our model allows for *four levels of immunity* to each allele.

- Level 3: no immunity, because no previous infection by a strain which includes that allele.
- Level 2: first stage of full immunity," which arises just after recovery from an infection by a strain with that allele active (or vaccination)
- Level 1: second stage of full immunity to infection by that allele, but immunity whose strength is beginning to wane.
- Level 0: fully waned immunity, which may retain some level of immunity.

As we keep track of infection history, these immunity indicators are used to specify characteristics of both the I and R populations.

Waning

Immunity strength moves:

- from level 3 to level 2 upon infection by a strain with that allele active.
- from level 2 to level 1 over time.
- from level 1 to level 0 over time *(waning).*
- from any level to level 2 upon infection by that allele, and its subsequent recovery

Immunity Status Vector

1. In the R-compartments of recovered individuals, we keep track of the immune history of its agents. So, $R(K_0, K_1, K_2, K_3)$ is the compartment of all formerly infected, but now recovered and potentially partially susceptible, individuals who are at level K_0 for allele 0, level K_1 for allele 1, level K_2 for allele 2, and level K_3 for allele 3, where each K_i is 0, 1, 2, or 3, as above.
    - For example, compartment R(2,3,2,3) is composed of all those who have just recovered from infection by strain 02 and have never been infected by strains that contain alleles 1 or 3.
2. To label the I-compartments (I designates "currently infected"), we include the identity of the infecting strain and the immune history of those in that compartment. We need not and do not include their immune statuses for the infecting alleles since they will automatically move to level 2 upon recovery.
    - Compartment $I(a_0, a_1; H_0, H_1)$ is composed of all those who are currently infected by the strain where the allele present at site 0 is a_0 (0 or 1), where the allele present at site 1 is a_1 (2 or 3), whose immune status to the allele not present at site 0 is H_0 (0, 1 ,2, or 3), and whose immune status to the allele not present at site 1 is H_1 (0, 1, 2, or 3).
    - For example, compartment I(0,2;3,3) is composed of all those who are currently infected by strain 02 and have never been infected by strains with alleles 1 or 3. Upon recovery, they will transit into compartment R(2,3,2,3).

Transmission

We write b_{hi} for the transmission rate for an encounter between an infected with strain (h,i) and a never-infected susceptible. In our allele-based approach we view b_{hi} as a combination of the transmissibility b_h of allele h at site 0 and the transmissibility b_i of allele i at site 1, possibly modified by an epistasis factor e_{hi}, which we ignore for the moment.

What is a natural way to formulate b_{hi} in terms of b_h and b_i? In a sense, the functional form is not so important. We will start in our *basic model* with equal b_h's and equal b_{hi}'s and be concerned with the impact of both large and small *changes* in individual b's. In an allele-based approach, a small change in, say, b_0 would be reflected in small changes of the basic values of *both* b_{02} and b_{03}. In a strain-based approach, we would simply change the value of the b_{hi} for the strain (h,i) under consideration.

For simplicity's sake, we compare $b_0 + b_2$ and $b_0 \cdot b_2$. Multiplication would be natural if we were parameterizing frequencies in the sense of Hardy-Weinberg Equilibrium [H08, W08]; after all, Prob(AB) ≈ Prob(A)·Prob(B). On the other hand, following Fisher [F18], the additive version is more natural if we are interested in aggregating effects. Since our concerns are more in line with effects than with frequencies, we use the simpler additive formulation in our model. Once again, for our concerns with incremental changes, it really does not make a difference.

Susceptibility/Immunity

We include parameters to quantify the immune response to an infecting strain.

Current immunity to infection by allele j has two components:

- the maximum *relative immune strength* m_j to infections by that allele j

  The parameter m_j ($0 \leq m_j \leq 1$) parameterizes the average immunity strength to virus allele j of those with level 1 or level 2 immunity to allele j. It is a measure of the population's maximal immune strength when encountering virus allele j.
- the current *parameterized immune status* κ_j of immune status K_j to allele j of the susceptibles in compartment $R(K_0, K_1, K_2, K_3)$.

  The parameter κ_j gives a numerical value to each immunity level K_j:
  - $\kappa_j = 0$ for immunity level K_3 (never infected, no immunity),

- $\kappa_j = 1$ for level K_2 or K_1 (full immunity), and
- κ_j = some value nu_j less than one for level K_0 (fully waned).

$\kappa_j = nu_j$ measures individuals' total "waning" of immunity from its max value.

We write $\kappa_j \cdot m_j$ for the current immunity for those in compartment $R(K_0, K_1, K_2, K_3)$ to infection by allele j: (maximal (relative) immunity m_j) modified by (immune status parameter κ_j). Since susceptibility is the complement of immunity, we write $(1-(\kappa_j \cdot m_j))$ for the *susceptibility* to allele j for those in compartment $R(K_0, K_1, K_2, K_3)$. See, for example, [Dys21]. Once again, the precise formulation used to denote the transmissibility-reduction factor is not important, for the same reasons presented in our discussion of the aggregation of transmissibilities; though, our formulation seems a natural one.

Examples:

- Someone never been infected by a strain with allele 2 will have $\kappa_2 = 0$ and full susceptibility $1 = (1-(\kappa_2 \cdot m_2)) = (1-(0 \cdot m_2))$ to allele 2.
- Someone just recovered from infection with an allele-2 strain ($\kappa_2 = 1$) in a world where full immunity to allele 2 is possible ($m_2 = 1$) will have zero susceptibility $(1-(1*1))$ to strains with allele 2.

Total susceptibility in compartment $R(K_0, K_1, K_2, K_3)$ to infection by the strain with alleles i and j is:

$$Susc(i,j) = (1-(\kappa_i \cdot m_i))\,(1-(\kappa_j \cdot m_j))\,(1-(\kappa_i \cdot \kappa_j \cdot m_{ij})),$$

where m_{ij} measures the *relative joint immune strength* for alleles i and j.

As noted in the Introduction, this model suffers from substantial combinatorial explosion—however, an illustration of the first few compartments following the fully susceptible compartment (i.e. R(3,3,3,3)) through the initial infection are given in Figure 3.

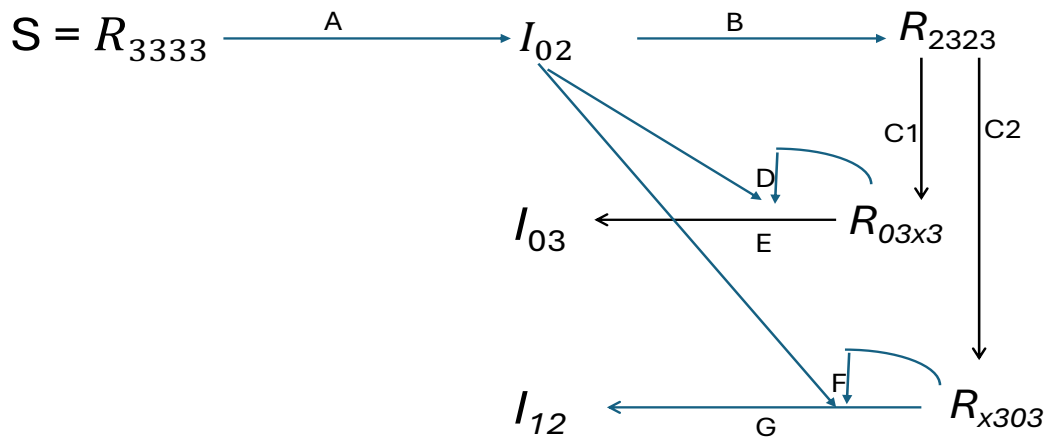

Figure 3. Illustration of the first few compartments of the model, following susceptible individuals through their initial potential infection and recovery. In a population of susceptibles, a few 02-infectives spread the infection (A). Upon recovery (B), the 02-infecteds move to recovered compartment R_{2323}, with temporary strong immunity to infection by alleles 0 and 2, but no immunity to infection by alleles 1 and 3. This immunity wanes (C) over time. Once immunity to allele 0 diminishes (C1), new infections (D) yield mutations (E) to strain 03. Once immunity to allele 2 diminishes (C2), new infections (F) yield mutations (G) to strain 12. The 03-infecteds recover to R_{2xy2}; the 12-infecteds recover to R_{x22y}. But all the other active R-compartments have low or no immunity to strain 13, which mutates from strains 12 and 03. Strain 13 soon dominates.

ADE (Antibody Dependent Enhancement)

Recall that ADE involves events in which the immune response to previous coronavirus infections generates antibodies that help the virus enter immune system cells, thus contributing to rather than preventing infection. To model ADE, we assign a number d_i to each allele i, a number that reflects how the immune response to that allele can increase transmissibility via ADE. We include $d_0 + d_1 + d_2 + d_3$ in the transmission term, but we set d_i to zero for susceptibles in R-compartments that have never been infected with a strain that included allele i.

B. Model Dynamics

Transmission Factor

Suppose an infected in compartment $I(h,j; H_{h'},H_{j'})$ encounters a recovered susceptible in compartment $R(K_0, K_1, K_2, K_3)$. The infecting strain consists of allele h in site 0 and allele j at site 1; h' is the other allele at site 0 and j' is the other allele at site 1. The transmission term for such an encounter is:

$$\mathbf{FI}(i,j; K_0, K_1, K_2, K_3) \equiv \text{Full Transmission Factor} \times \text{Susceptibility Reduction Factor}$$

$$= [(b_i + b_j + e_{ij}) + (d_0 + d_1 + d_2 + d_3)] \times [(1-(\kappa_i \cdot m_i))\,(1-(\kappa_j \cdot m_j))\,(1-(\kappa_i \cdot \kappa_j \cdot m_{ij}))].$$

The reduction in transmissibility via multiplication by a susceptibility factor is a key ingredient in the literature we have cited; see, for example, [Sa20, Sa24a, Sa24b, Bu25].

Note that **FI** is independent of the immune status of the infector. Write J(i,j) for the total of those currently infected with strain (i,j), summing over all immune histories of the infected:

$$J(i,j) \equiv \sum \{\, I(i,j; H_{i'},H_{j'}) : \text{all 4 immune levels } H_{i'} \text{ and all 4 immune levels } H_{j'} \},$$

where i and i' are the two alleles at site 0 and j and j' the two alleles at site 1.

Putting all this together, we ask how many new additions are there to compartment $I(i,j; H_{i'},H_{j'})$ in any time period Δt.

The answer:

[total number of those infected by strain (i,j)] ×
[total number of susceptibles with immune status $H_{i'}$ for allele i' and $H_{j'}$ for allele j'] ×
[Transmission Factor **FI** for such an encounter] × rate of such encounters × Δt.

Example:

The number of additions (new infecteds) to compartment $I(1,3; H_0, H_2)$ in **a time period** Δt is

$\sum J(1,3) \cdot R(H_0, K_1, H_2, K_3) \cdot \mathbf{FI}\,(1,3;\ H_0, K_1, H_2, K_3) \cdot c \cdot \Delta t$, summed over all K_1 and K_3,

where c is the **effective contact rate**, the rate of possibly infection-transmitting contacts in this homogeneously mixing population.

Mutation/Drift

Finally, we include the dynamic of mutation/drift, an essential component of our model. Note that we assume that mutations happen observationally as an infector of one strain generating an infectee of a different, related strain (i.e. the mutation occurs in the original infected person but is not noticable/substantial). Once more, for any choice of strain (i,j) based on allele i at site 0 and allele j at site 1, write i' for the other allele at site 0 and j' for the other allele at site 1. In a single mutation, strain (i,j) can mutate to strain (i,j') or to strain (i', j). To this end, we formulate drift rates $f_{jj'}$ and $f_{ii'}$. Importantly, mutation (i,j) → (i',j) can only occur if the infectee has minimal immunity to allele i'. Mutation (i,j) → (i,j') can only occur if the infectee has minimal immunity to allele j'; specifically, only if immunity level $K_{j'}$ for allele j' = 0 or 3.

For example, the expression for $\Delta I(1,2; H_0, H_3)$, the number of new strain (1,2) infections in time interval Δt, would include mutations from strain (1,3) to strain (1,2) when immunity to allele 2 is weak, i.e., K_2 = 0 or 3:

$$f_{32} \cdot \sum_{K_1=0,1,2,3} J(1,3) \cdot \mathrm{R}(\mathrm{H}_0, K_1, 0, H_3) \cdot \mathrm{FI}(1,3;\ H_0, K_1, 0, H_3) \cdot \ \mathrm{c} \ \cdot \ \Delta\mathrm{t}$$

$$+$$

$$f_{32} \cdot \sum_{K_1=0,1,2,3} J(1,3) \cdot \mathrm{R}(\mathrm{H}_0, K_1, 3, H_3) \cdot \mathrm{FI}(1,3;\ H_0, K_1, 3, H_3) \cdot \ \mathrm{c} \ \cdot \ \Delta\mathrm{t}$$

and mutations from strain (0,2) to (1,2) when immunity to allele 1 is weak, i.e., K_1 = 0 or 3

$$f_{01} \cdot \sum_{K_2=0,1,2,3} J(0,2) \cdot \mathrm{R}(\mathrm{H}_0, 0, K_2\ , H_3) \cdot \mathrm{FI}(0,2;\ H_0, 0, K_2, H_3) \cdot \ \mathrm{c} \ \cdot \ \Delta\mathrm{t}$$

$$+$$

$$f_{01} \cdot \sum_{K_2=0,1,2,3} J(0,2) \cdot \mathrm{R}(\mathrm{H}_0, 3, K_2\ , H_3) \cdot \mathrm{FI}(0,2;\ H_0, 3, K_2, H_3) \cdot \ \mathrm{c} \ \cdot \ \Delta\mathrm{t}$$

To achieve the complete dynamics for $\frac{d}{dt} I\left(i, j; , H_{i'}, H_{j'}\right)$, add in expressions for

the strain recovery rate $\quad -\mathrm{v}_{\mathrm{ij}} \cdot I\left(i, j; , H_{i'}, H_{j'}\right)$

and background death $\quad -\mathrm{mu} \cdot I\left(i, j; , H_{i'}, H_{j'}\right)$.

Complete expression for $\frac{d}{dt}$ R(K_0, K_1, K_2, K_3) includes:

- + : Recoveries of infected
- −: new infections via contacts with infected in the J(i, j)s
- + and − : Waning between R(K_0, K_1, K_2, K_3). levels
- − : Background deaths

We present the complete set of equations in the Supplementary Material.

C. Population Susceptibility

In our analysis of the evolution of the pandemic -- as one dominant strain mutates to another variant, the population level of immunity/susceptibility to the variant strain plays a major role. As Meijers et al. [Me23] note in opening their paper: "SARS-CoV-2 variants compete on a fitness landscape shaped by population immunity." Accordingly, we include a measure of population immunity – or its complement, population susceptibility. Our model has a natural choice for such a measure.

In compartment R(K_0, K_1, K_2, K_3) of recovered individuals, we defined the susceptibility to any strain (h,j), where h=0 or 1, and j=2 or 3, as:

$$(1-m_h \cdot \kappa_h)\ (1-m_j \cdot \kappa_j)\ (1-m_{hj} \cdot \kappa_h \cdot \kappa_j).$$

The total susceptibility in the population to strain (h,j) would then be:

$$\mathrm{Susc}(h,j) \equiv \sum R(K_0, K_1, K_2, K_3) \cdot (1-m_h\, \kappa_h)\ (1-m_j\, \kappa_j)\ (1-m_{hj}\, \kappa_h\, \kappa_j),$$

summed over all admissible (K_0, K_1, K_2, K_3).

The expression J(h,j) ·Susc(h,j)

[number infected by strain (h,j)] times [number susceptible to strain (h,j)]

plays a role like the effective population reproduction number R (as opposed to R_0) in that strain (h,j) takes off to an epidemic wave when J(h,j) ·Susc(h,j) increases past some threshold.

Note 1. In the special situation where recovery from an infection leads to full immunity (for a while) – all $m_i = 1$ – and all immunity eventually wanes to zero – all $nu_i = 0$, then the above expression becomes simply:

$$\text{Susc}(h,j) \equiv \sum \{R(K_0, K_1, K_2, K_3) \mid \text{all admissible } (K_0, K_1, K_2, K_3) \text{ with } K_h = 0 \text{ or } 3,\ K_j = 0 \text{ or } 3\}.$$

In other words, in this case Susc(h,j) is the fraction of the population with *no immunity* to both alleles h and j. The graph over time of this simplified measure of population susceptibility usually runs parallel to and just below the time graph of the full Susc(h.j) expression above.

Note 2. See [Top24] for an overview and Xu et al. [Xu15] for an early attempt to find a measure of immunity/susceptibility *for individuals.*

3 Methods: Parameter Choices and Base Case Values

A. Baseline Parameter Values

We begin our analysis of this system by summarizing our notation for the parameters and presenting the baseline values for these parameters that we use to start this analysis.

Parameters in model	**Base value**		**Parameter Name in Madonna Program**
b_j = transmissibility of allele j	$b_1 = b_2 = b_3 = b_4$	= 0.3	bj
b_{jk} = joint transmissibility of alleles j and k	$e_{02} = e_{12} = e_{03} = e_{13}$	= 0.0	bjk
v_{jk} = rate of recovery from infection by strain jk	$v_{02} = v_{12} = v_{03} = v_{13}$	= 0.1	vjk
m_j = relative full immunity strength to allele j	$m_0 = m_1 = m_2 = m_3$	=0.7	mj
m_{jk} = relative full joint immunity strength to alleles j and k	$m_{02} = m_{12} = m_{03} = m_{13}$	=0.0	mjk
nu_j = relative fully waned immunity strength to allele j	$nu_0 = nu_1 = nu_2 = nu_3$	=0.05	nuj
$f_{jk} = f_{j\to k}$ = probability of drift from allele j to	$f_{01} = f_{10}$	=0.00001	f01, f10

allele k at same site	$f_{23} = f_{32}$	=0.00001	f23, f32
w_{j2}= transition rate for allele j from immune state 2 to immune state 1	$w_{02} = w_{12} = w_{22} = w_{32}$	=0.003	wj2
w_{j1}= waning rate for allele j from immune state 1 to immune state 0	$w_{01} = w_{11} = w_{21} = w_{31}$	=0.003	wj1
d_j = A.D.E. transmission factor for allele j	$d_0 = d_1 = d_2 = d_3$	= 0.0	dj
c = effective contact rate	c	=0.5	chi
mu = population birth/death rate	mu	=0.0001	mu

B. Notation Summary

Below we include a quick-reference summary of the notation presented in the previous section.

Levels of immunity. Four levels of immunity:

- Level 3, K_3 = zero immunity state, for those never infected
- Level 2, K_2 = "first full immunity state," for recently recovered or vaccinated
- Level 1, K_1 = "second full immunity state," for those whose full immunity is just starting to wane
- Level 0, K_0 = state of fully waned immunity

κ_j = numerical value assigned to immunity level K_J

$\kappa_3 = 0$, $\kappa_2 = 1$, $\kappa_1 = 1$, $\kappa_0 = nu_i$ (allele dependent, as listed above)

Population Compartments. $R(H_0, H_1, H_2, H_3)$, where for each allele j, H_j is one of four immunity states, is the size of the compartment of those whose immune state for allele 0 is H_0, immune state for allele 1 is H_1, immune state for allele 2 is H_2, and immune state for allele 3 is H_3.

$I(j, h, H_{j'}, H_{h'})$ is the compartment of those infected by strain (j,h) whose immune status for allele j' is $H_{j'}$ and for allele h' is $H_{h'}$, where j and j' are the two alleles at site 0 and h, and h' are the two alleles at site 1.

S is the compartment of never-infected susceptibles, which we will sometimes write, in convenient abuse of notation, as simply R(3,3,3,3).

J(h,k) = the number of those currently infected by strain (h,k), where h ϵ {0,1} and k ϵ {2,3}. We have written J(h,k) in Madonna code as Totlhk.

In the following exposition, we will usually write the names of the R-compartments, the I-compartments, and the total strain sums without commas or parentheses. For example, we write R(2,3,2,3) as R2323, we write I(1,2; 3,1) as I1231, and we write J(0,2) simply as 02.

C. Baseline Initial Conditions:

Since population size is conserved, we can and do work with total population size = 1.0.

In nearly all the runs in this paper, we seed an otherwise completely susceptible population with a few infections of strain 02 ("seeded strain"). In this case, our dynamical system has initial condition:

INIT I0233 = 0.00001, INIT R3333 = 0.99999, all other initial conditions = 0.

Note that with recovery rate at v_{02}=0.1, at the outset of the infection, before immunities kick in, each infected will yield roughly

$$(b_0 + b_2) \cdot c \cdot (1/v_{02}) = .6 \cdot .5 \cdot 10 = 3$$

new infections, an R_0 consistent with the early data for SARS-CoV-2, as discussed in [Ko22].

Nevertheless, with our 2-site-2-allele model and its many other simplifying assumptions, we are not trying to capture the real life dynamics of the current SARS-CoV-2 pandemic. Our focus is on the relative effects of parameter changes on model behavior and on illustrating the benefits of an allele-based multi-layer approach.

D. Using no-drift model to understand the Susceptibility curves.

We study our model with the above default parameters, the "basic parameter set." We begin with the simplest case; we set the mutation-rate parameters f_{ij} all to zero. If we seed the model only

with strain 02 infectives, then no other strain can appear, and we have a simple single strain model. The only *active* compartments are I(0,2;3,3) and the R(x,3,y,3) for x,y=0,1,2,3.

The graph in Figure 4 presents the prevalence of strain 02 using the left vertical axis, and the graphs of the Susceptibility Curves using the right vertical axis. We usually include the Susceptibility Curves because there is so much information in them. To keep the graphs from interfering with each other, we move the 0-level for the Susceptibility Curves higher on the right vertical axis. To clarify the information in any given graph, we will often vary the time scale on the horizontal axis and prevalence scale on the left vertical axis

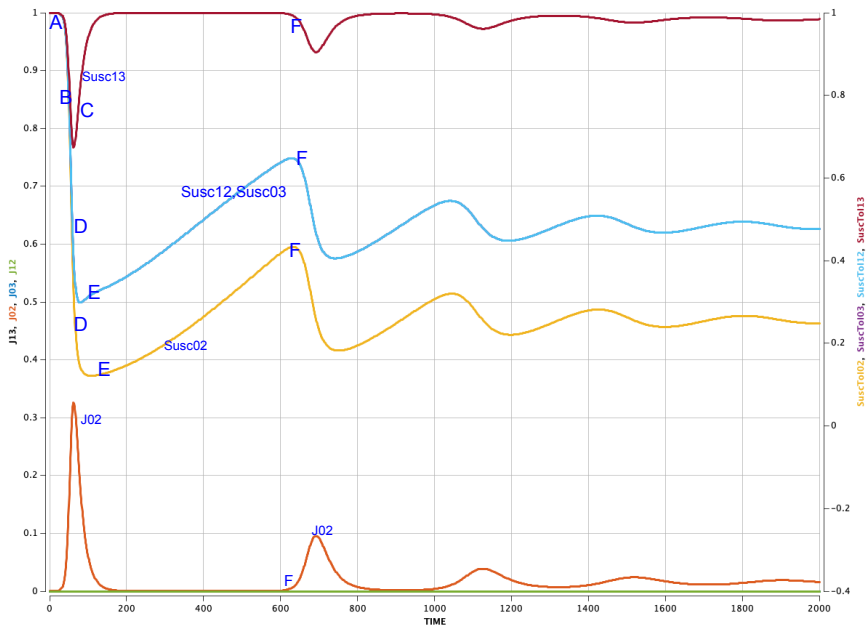

Figure 4. The graphs of strain 02 and of the four susceptibility curves for the no-drift situation.

Despite the simplicity of this one-strain situation, the patterns in Figure 4 occur regularly, and so we describe them here in some detail. The letters on the left refer to the letters in Figure 4.

(A) For t near 0, nearly everyone is in pure susceptible compartment R3333, i.e., compartment S. All the susceptibility curves start near 1 since R3333 is a component of all them.

(B) With this high level of susceptibility, the existence of even a few strain-02 infectives leads to the rapid growth of strain 02, leading to a peak prevalence of ~0.326 at t =65.

As the new infections spread, the size of compartment S = R3333 decreases rapidly and so do all the Susceptibility Curves.

(C) The recovered 02-infectives move to R2323 and eventually -- through waning immunity – into R2313, R1323, and R1313. These Rx3y3's and the not-yet-infected R3333's are all full components of Susc13 (the Susceptibility curve for strain 13) and only of Susc13. Around t=65, the growth of these Rx3y3 compartments outpaces the number of new 02 infections; Susc13 begins to rise dramatically, nearly all the way back to level 1.

(D) These Rx3y3's play only a minor role in Susc12 and Susc03 and almost no role at all in Susc02. So these latter susceptibility curves keep falling while Susc13 rises, with Susc12 and Susc03 larger than Susc02.

(E) At some point (t=108), the population immunity reaches a point where the number of 02-infectives are few and far between (~0.04) and new infections are outpaced by new "births" into R3333. This turn-around allows the Susc12, Susc03 and Susc02 curves to begin to rise and to keep rising roughly at the birth rate. This rise is eventually abetted by the full waning of the recovered R2323's to compartment R0303, also a component of all 4 susceptibility curves.

(F) While Susc02 is rising, there are very few strain 02 infectives. However, at some point (t=633) the number of infected with strain 02 times the population susceptibility Susc02 reaches a threshold so that a new strain 02 epidemic wave begins. The new infections in this wave decrease the sizes of R3333 and R0303 so that all four susceptibility curves begin to decline, and the cycle continues.

Given the high susceptibility of strain 13 after the 02 epidemic wave in Figure 4, it would take just a few 13 infections to set off a 13 epidemic wave, as we will soon see.

The model behind Figure 4 is basically the one-strain model that Saad Roy et al. [Sa20, Sa21a, Sa21b, Sa22, Sa23, Sa24a, Sa24b, Es24, Wa21] use to begin their analyses, before adding vaccination and non-pharmaceutical interventions. One of their focal points is "accumulating immunity," a phenomenon visible in the decline in the heights of the epidemic peaks and in the Susc02 curve in Figure 4.

E. Our basic system (with drift)

Two important takeaways from Figure 4 are:

1) the pattern of widely-spaced epidemic waves with little infection in the gaps between them, and

2) the strong build-up in susceptibility to strain 13 from recoveries of strain 02 infections.

Strains 02 and 13 have no alleles in common. We call such pairs *complementary strains* or simply *complements*. Recoveries from any infection strain build up susceptibilities to its complementary strain.

The high level of susceptibility of strain 13 after the episode of strain 02 suggests that any introduction of strain13 infection will lead to a strain 13 epidemic. To see this more clearly, we move to the next step and re-introduce a small rate of drift/mutation, all $f_{ij} = 0.00001$, so that the parameters of our model are exactly the default values listed above.

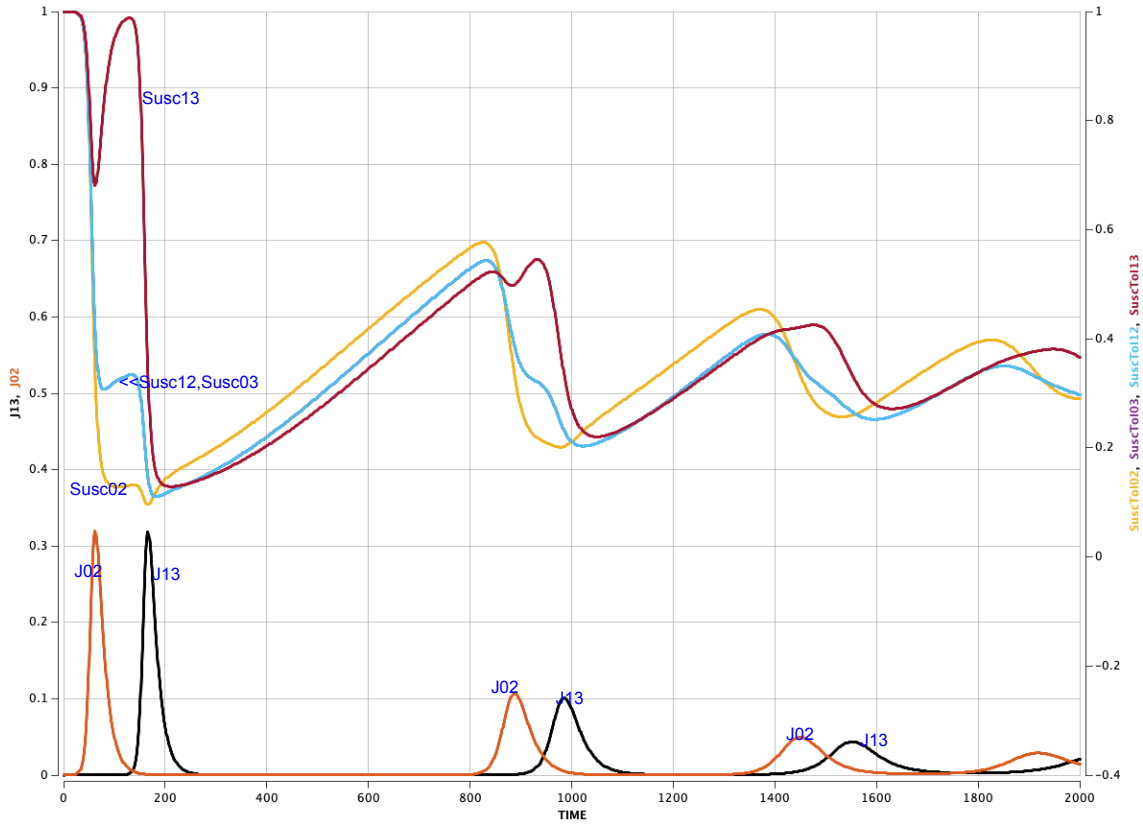

Figure 5. The strain prevalences and the susceptibility curves for the basic parameters.

Figure 5 includes both the four strain prevalences and the "four" susceptibility curves. The latter are pretty much as they are in figure 4, with the obvious exception of Susc13. We see strains 02 and 13 alternating as epidemic waves, with a moderately long gap of negligible infection after each 13 episode. Strains 03 and its complement 12 are barely ever present.

We present the dynamics in more detail:

- Prevalence of strain 02 first peaks at around 32% at t= 62, just as it did in Figure 4. By t=100, over 75% of the population has experienced and recovered from strain 02. (R2323 is at 75% at t=100.) Strain 02 prevalence then bottoms out at 0.000002% around t=425, before achieving its second wave with a peak of ~11% at t=883.
- Meanwhile its complementary strain 13 is invisible until t=52; it does take *two* mutations to generate a 13 from a 02. But at t=130, susceptibility for 13 reaches almost 100%, so it takes just a few 13s to generate a 13 epidemic wave. This 13-wave peaks at ~32% at

t=166, before strain 13 begins its decline, bottoming at 0.000004% around t=600. Then, once again its susceptibility rises as a result of recoveries from the second 02 wave. Strain 13 has its second peak at ~10% at t=985.

- The prevalences of strains 12 and 03 are identical for all time because they are governed by the exact same parameter values and initial conditions. Only one mutation away from the seeded 02, they appear early and peak at ~0.0007% at t=70. However, their susceptibilities do not have prior recoveries to build on, as strain 13 did, and so their prevalence times susceptibility product never reaches a threshold for epidemic takeoff. They never do re-achieve the ~0.0007% of their first "peak."

We call the model behind figure 5 our "basic model." We now ask how do various parameter changes affect the patterns in Figure 5. In particular, what is the effect of the appearance of strains with alleles whose parameter values differ from the default values? Will the appearance of alleles with very high transmissibility or immune-response values lead to the dominance of strains composed of those alleles? We will see: sometimes yes, and sometimes no.

4 Results: Effects of Parameter Changes

Parameter values play a role in which strains appear and in what order, during the course of the pandemic. We examine many kinds of parameter changes – single and multiple – partly to be thorough and partly because some of the multiple changes have surprising effects on the evolution of the pandemic. For the more interesting effects we include a figure with a graph of the new prevalences and susceptibility curves, for comparison with the graph in Figure 5 of the dynamic with our default parameters. For some parameter changes, we present the relevant graph in the Supplementary Material in figures with a prefix S.

Our simulations illustrate four different mechanisms that interact to increase or decrease the strength of a given strain during the pandemic. We list them here with illustrative examples:

Direct parameter effect: increasing transmissibility b_1 of allele 1 should foster the appearance and strength of strains 12 and 13;

Complementarity effect: Recoveries from strain 02 infection have high immunity to new infections involving alleles 0 and 2, and relatively low immunity and high susceptibility to infections involving alleles 1 and 3, leading to a strong presence of strain 13.

Founder (Seeding) Effect: Seeded strain 02 has a built-in advantage, one that can spread to its complementary strain 13;

Strain adjacency effect: Single allele mutation is more likely than multiple allele mutation, so that strain 02 is more likely to mutate to strain 12 or 03 that to strain 13.

Transmissibility

I. Direct increase to the transmission parameters

We focus on the transmissibility (contagiousness) parameter b_i to start our analyses of the effects of parameter changes. We will systematically adjust different subsets of the b_is and will see that the resulting dynamics illustrate all four of the above mechanisms.

We begin with an overall summary, presenting figures 6 through 10 – for comparison with the baseline Figure 5 --- to illustrate and summarize the diverse patterns that arise when we modify various b_i's from their baseline values $b_0 = b_1 = b_2 = b_3 = 0.3$. We follow these Figures with a more detailed discussion of the dynamics behind each of them.

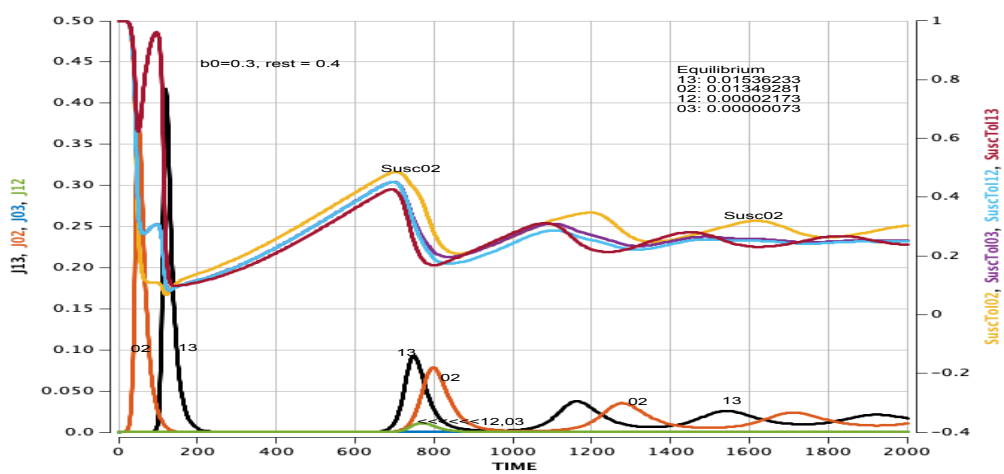

Figure 6: $b_0 = 0.3$, $b_1 = b_2 = b_3 = 0.4$. Despite a lower transmissibility for one of its alleles, seeded strain 02 maintains its presence throughout the waves of epidemics, with higher prevalence than the stronger (higher contagiousness) strain 12.

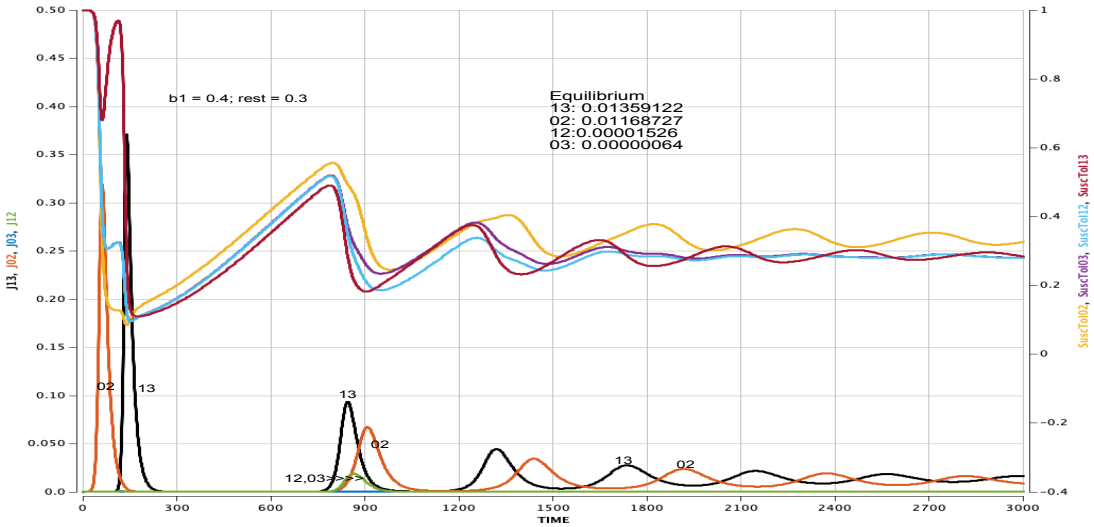

Figure 7: $b_1 = 0.4$, $b_0 = b_2 = b_3 = 0.3$. Increasing the transmissibility of non-seeded allele 1 aids the complement 13 of seeded strain 02 and therefore of seeded strain 02, with little effect on strain 12.

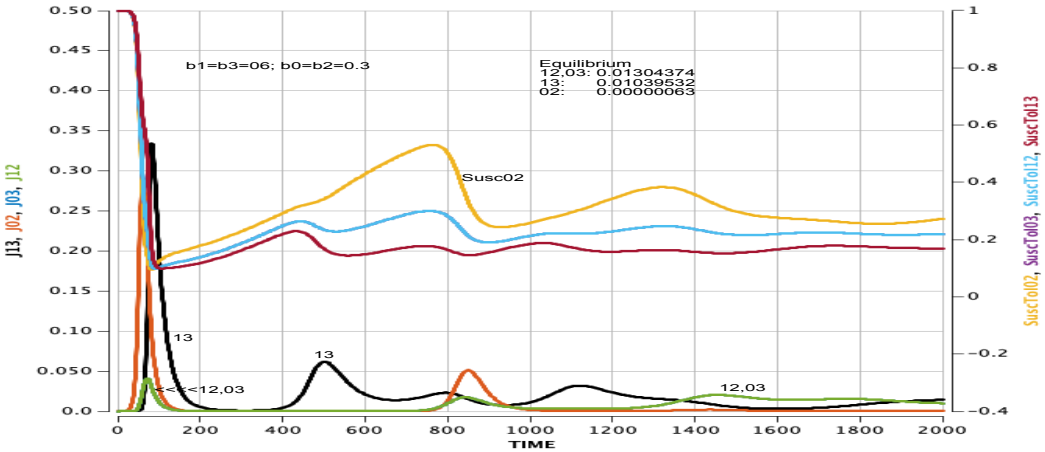

Figure 8. $b_0 = b_2 = 0.3$, $b_1 = b_3 = 0.6$. Increasing transmissibility of non-seeded alleles 1 and 3 eventually decreases the prevalence of seeded strain 02.

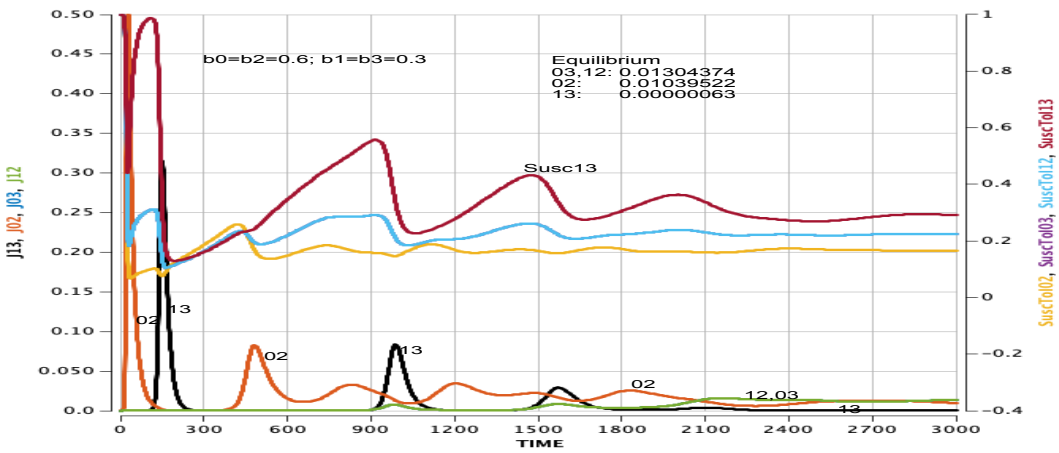

Figure 9. $b_0 = b_2 = 0.6$; $b_1 = b_3 = 0.3$. *Doubling the transmissibility of the seeded alleles 0 and 2 diminishes the prevalence of strain 02 because its complement 13 is so weak.*

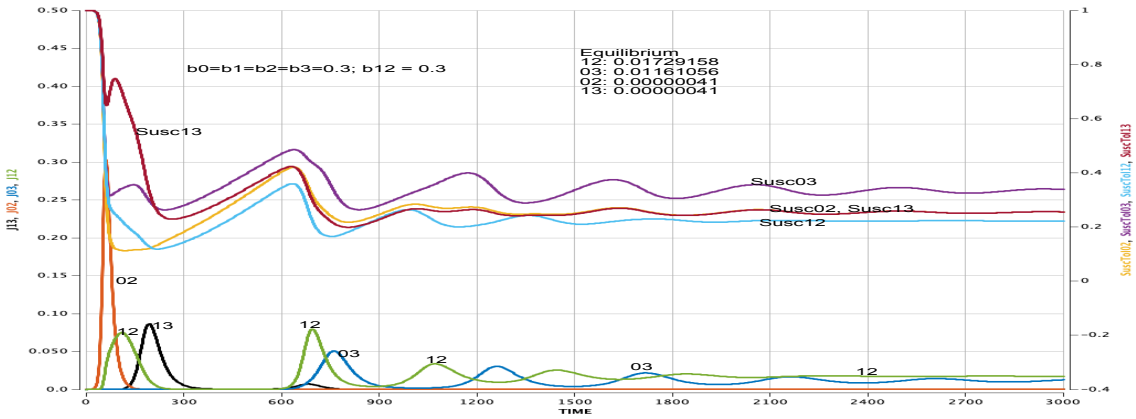

Figure 10. $b_0 = b_1 = b_2 = b_3 = 0.3$; $b_{12} = 0.3$. *Increasing* ***joint*** *transmissibility* b_{12} *increases the prevalence of 12 and its complement 03*

I.A. Increased transmissibility across all alleles leads to faster, higher epidemics.

If we increase all four b_is equally, all that happens is that the peaks of the 02 and 13 strain epidemics increase and the intervals between the epidemics decrease. Compare Figure S1 in the Supplementary Material with Figure 5.

I.B. Slightly lower transmissibility for one of the seeded alleles (Figure 6: $b_0 = 0.3$, $b_1 = b_2 = b_3 = 0.4$)

If we increase all the b_is except b_0, we'd expect that the less transmissible seeded strain would suffer. As Figure 6 illustrates, its seeding keeps strain 02 prominent in the course of the pandemic. Strain 12, in spite of its higher transmissibility than strain 02, has one small epidemic wave, but its equilibrium value is three orders of magnitude lower than that of seeded strain 02. The power of the seed!

All the parameters for strains 13 and 12 are the same. Yet, strain 13 dominates, while strain 12 virtually disappears. Strain 13 takes advantage of its complementary susceptibility to seed 02, while strain 12 does not take advantage of its strain adjacency advantage to seed 02 because of the increasing immunity to allele 2 from recovering 02 infections.

I.C. Much lower transmissibility for one of the seeded alleles (Figure S2: $b_0 = 0.3$, $b_1 = b_2 = b_3 = 0.4$)

The phenomenon of the preceding paragraph persists even if we double the advantage b_1, b_2, and b_3 over b_0. Compare Figure S2 to Figure 6. Similar patterns occur if we keep $b_1 = b_2 = b_3$ at 0.3 and lower b_0 to 0.2 or 0.15, or if we vary only b_2 instead of b_0.

I.D. Increasing transmissibility of a single non-seeded allele (Figure 7: $b_1 = 0.4$, $b_0 = b_2 = b_3 = 0.3$)

What happens when we change only b_1 or b_3, transmissibility of one of the non-seeded alleles? We work with b_1; analogous results hold for varying only b_3.

Who gains most from any change in allele 1 transmissibility: strain 12 or strain 13? And how does the seeded strain 02 fare? As expected, strain 02 continues to thrive, even though it has

lower transmissibility: seeding effect. With this benefit, strain 13 does much better than 12 because of the complementary susceptibility effect. (Strain 13 even surpasses 02 as the third epidemic wave.) In the long run, seeded strain 02 is five orders of magnitude larger than strain 03, despite strain 03's transmissibility advantage.

I.E. Increasing transmissibility of both non-seeded alleles (Figure 8: $b_0 = b_2 = 0.3$, $b_1 = b_3 = 0.6$):

What if we increase both b_1 and b_3? With low transmissibility for both of its alleles, strain 02 finally loses its seed advantage. It has enough strength for a second wave, but soon virtually disappears. As a result, strain 13 loses its complementary susceptibility advantage; although it has the highest total transmissibility, 13 loses out to the complementary strains 12 and 03 in the long run. The adjacency effect helps strains 12 and 03 mount an epidemic wave even before strain 13 does.

I.F. Increasing transmissibility of both seeded alleles (Figure 9: $b_0 = b_2 = 0.6$, $b_1 = b_3 = 0.3$):

Next, we let both b_0 and b_2 dominate. With the strong parameter effect and the seeding effect, one would expect strain 02 to truly dominate. As Figure 8 illustrates, it does not. Strains 12 and 03 do better in the short run, and especially in the long run. Parameter effects diminish the strength of strain 13; so strain 02 cannot benefit from any complementary susceptibility effect. Strains 12 and 03 show no adjacency benefit, but they have a high parameter benefit and are mutual complements.

Summarizing for b_i changes: We would expect strong direct parameter effects when we increase a transmissibility parameter b_i. Noteworthy are the persistence of the founder/seeding effect and the weakness of the strain adjacency effect. Most noteworthy is the strength of the complementary susceptibility effect in determining which strains thrive and which do not in the pandemic. This latter effect is most strongly tied to immunity issues, while the others more or less align with transmissibility issues.

I.G. Increasing a joint transmissibility (Figure 10: $b_0 = b_1 = b_2 = b_3 = 0.3$; $b_{12} = 0.3$): Finally, what happens if we vary one of the *joint* transmissibility parameters instead, say b_{12}? Figure 10 shows the result when b_{12} varies. Seeded strain 02 and its complement 13 start the pandemic, but the parameter and adjacency effects take over and strain 12 dominates the rest of the way. Its complement 03 with its resulting high susceptibility also does well, while strains 02 and 13 virtually disappear after their initial waves.

We achieve the exact same dynamic with only joint transmissibilities:

$b_0 = b_1 = b_2 = b_3 = 0$; $b_{02} = b_{13} = b_{03} = 0.6$, $b_{12} = 0.9$.

II. Increase contact rate c.

The basic reproduction number R_0 for the spread of an infectious disease is essentially:

(transmissibility) x (contact rate) x (infection duration).

We turn now from transmissibility b_i to contact rate and duration.

The contact rate appears in the model equations only as a multiplier of the transmission rates. Thus, increasing or decreasing the contact rate has the same effect as increasing or decreasing the transmission rate b, the changes portrayed in Figure S1.

III. Changing the recovery rate v_{ij}, or equivalently the duration of infection v_{ij}^{-1}

If we lower the infection recovery rate v_{ij} for all strains from our basic value of 0.1 to, say 0.05 – equivalently doubling the duration of infection for all strains, then the new strains arising via mutation make their appearance, while the pre-mutation strains are still active. The result, as seen in Figure 11, is the disappearance of any epidemic gap, and a four-fold increase in long-run values.

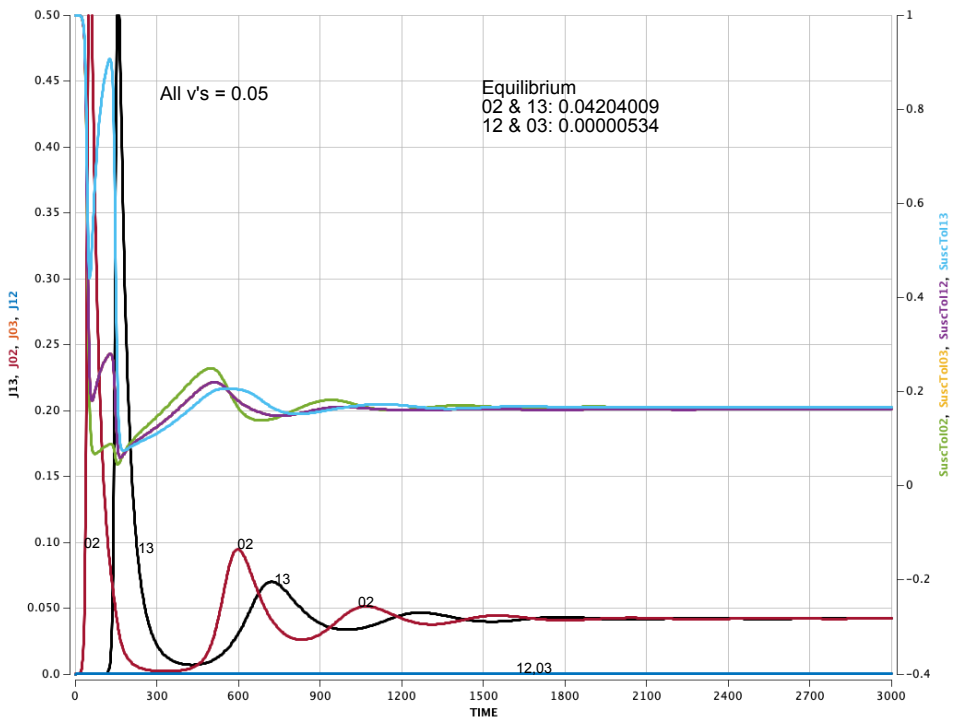

Figure 11. Slowing all recovery rates from 0.1 to 0.05 strongly increases prevalences.

Doubling all the recovery rates to 0.2, and thus halving the infection durations, substantially increases the epidemic gaps and leads to a five-fold decrease in long run infection levels.

What if we increase only one duration, say that of strain 03? We change v_{03} from 0.1 to 0.05, keeping the other recovery rates at 0.1. As seen in Figure 12, strain 03 gains strength impressively. Strain 02 and its complement still start the pandemic, of course. But they quickly lose their advantage as seed and seed complement, and virtually disappear from the graph. Strain 03 with its longer duration dominates ("direct parameter effect") complemented by regular epidemic waves from strain 12 ("complement effect").

Smaller increases in duration of 03 lead to similar, but slightly less dramatic effects. A strain with a longer duration of infection can dominate the pandemic in our model.

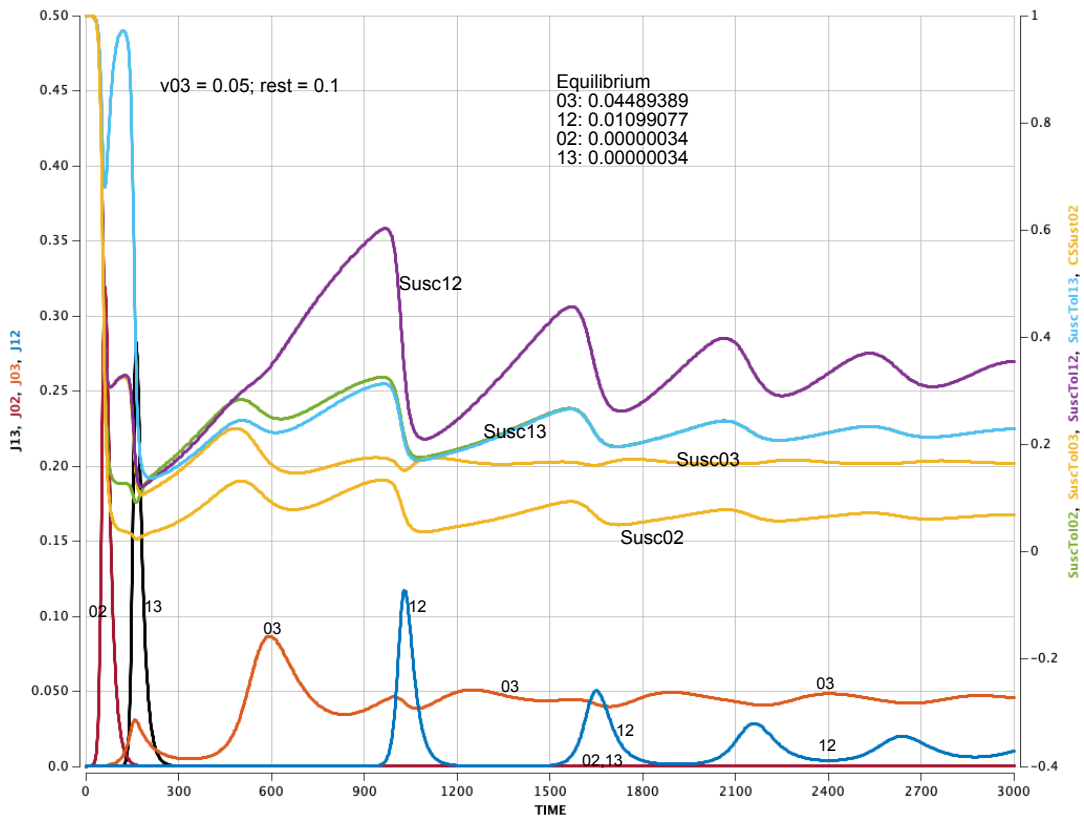

Figure 12. Strain 03 with the longest duration dominates the prevalences (v_{03}=0.05, other v_{ij}s = 0.1).

Immunity

Transmissibility and immunity are complementary components of infection spread. We have seen that the results of changing only transmissibility parameters can be effected by immunity considerations. We turn to an examination of the effects of changing immunity parameters – of which there are three:

- m_i: the average relative strength of immunity to new infections by strains with allele i (m_i =1 for full immunity, m_i = 0 for no immunity);
- w_{ih}: the average rate of immunity waning from immunity state h (=2 or 1) for allele i;
- nu_i: residual immunity to allele i strains at the end of the waning process.

As Saad-Roy, Levin, Grenfell et al. [Sa24a] note (and as we will see in this model as well), changes or uncertainty in the duration and strength of these immunity parameters can affect the outcomes in a wide range of ways.

IV. Increasing the waning rate w_{ih} leads to lower epidemic peaks but higher endemic prevalence

We start with the waning parameter w_{ih}, the rate at which allele i moves from immunity state K_h to immunity state K_{h-1}. Just as the recovery rate v_{ij} parameterizes the duration of infection, the waning rate parameterizes the duration of immunity.

If all the w_{ih} are from 0.001 to the base rate of 0.003 to 0.005. Now, steeper susceptibility curves dramatically shrink the intervals between pairs of epidemic waves; no other changes from basic model. If they are increased to 0.01, the gaps disappear. (Figure S3 in the Supplementary Material)

What if we change the waning rate for just one allele, say allele 1? We increased w_{11} and w_{12} from 0.003 to 0.03. Of course, the initial epidemic 02/13 wave pair is not affected by the waning rate. But after these, strains 12 and 13 dominate, as Figure 13 shows. Figure 13 also illustrates the high susceptibility to strain 12 early in the pandemic.

This surge of strains 12 and 13 that results from increased allele-1 waning rates shows the strong "direct parameter influence" of an increased waning rate. As [Sa24a, To24] point out, waning rate makes a big difference.

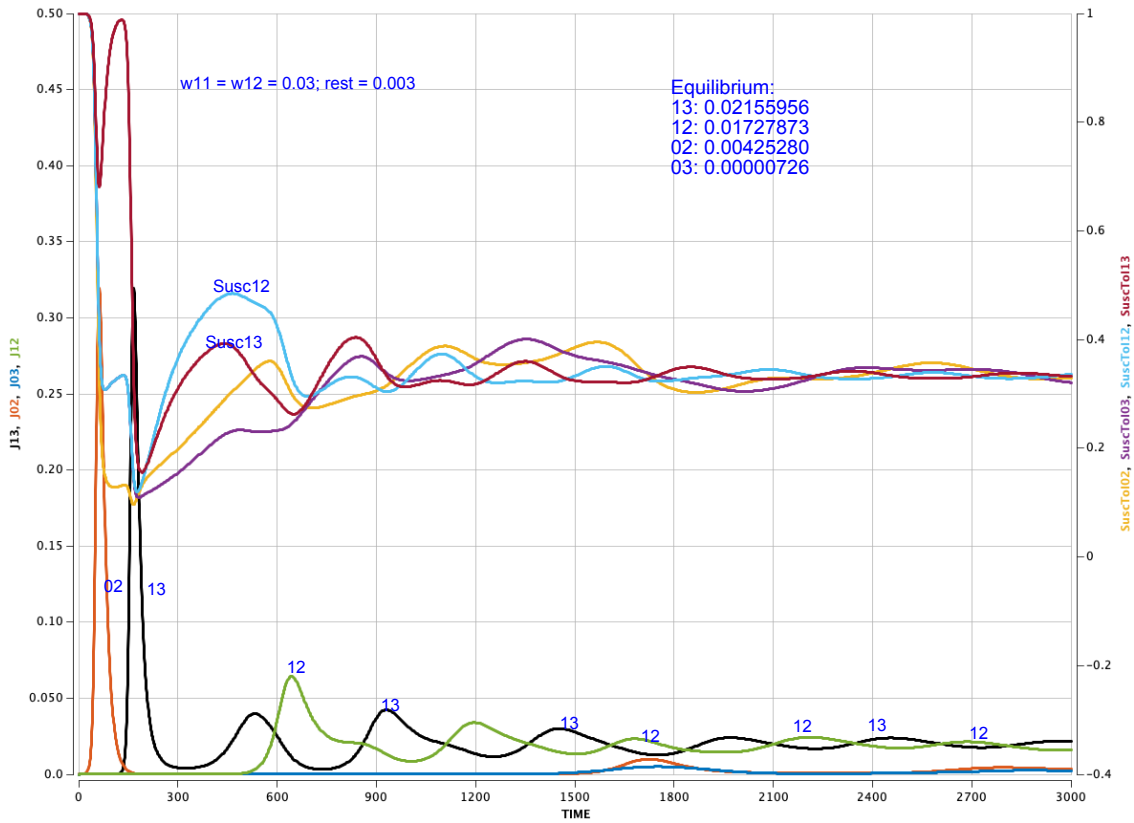

Figure 13. Increasing the waning rate of allele 1 leads to high prevalences of the strains with allele 1(w_{11}= w_{12}=0.03, other w's = 0.003).

V. Increasing the level of fully-waned immunity nu_i leads to longer inter-epidemic intervals and lower levels for the affected strain(s)

More residual immunity after waning leads to more immunity in the system, which naturally leads to lower epidemic peaks and longer inter-epidemic intervals compared to the baseline model (Figure 5) with all the nu_i's set at 0.05. We examine three cases**.** The corresponding graphs are straightforward modifications of the baseline case of Figure 5 and are presented in the Supplementary Material.

- Equally increasing all four nu_i's leads all around more immunity in the system and thus to longer intervals between epidemic waves. Fig. S4.

- If we increase only nu_0, strains 02 and 13 still dominate (no appearance of strain 03), but strain 02 undergoes lower peaks, lower long-run equilibria, and longer gaps between epidemic waves. Little effect on complementary strain 13. Fig. S5.
- If we increase nu_1, now strain 13 has the lower peaks, lower long-run equilibria, and longer gaps between epidemic waves. Fig. S6.

VI. Changing relative immunity m_i

The parameter m_i represents a susceptible's relative maximal immunity to allele i infections. If m_i=1, a susceptible with top level immunity (level K=1 or 2) cannot be infected by any strain that includes allele i. On the other hand, if m_i=0, the susceptible has no immunity to allele i, even at level K= 1 or 2.

VI.A Small uniform decreases in all m_i's leads to higher peaks and shorter gaps.

Decreasing all the immunity parameter m_i from 0.9 to 0.7 to 0.5 yields graphs similar to Figure 5 but with higher epidemic peaks, shorter epidemic gaps, and increasing equilibrium levels of infection. In all these cases, strain 02 and its complement 13 dominate. Strains 12 and 03 are barely visible but begin to show for all m_i = 0.5. See Figures S7 and S8 in the Supplementary Material for a detailed exposition.

VI.B. Weak population-wide immunity leads to the appearance of strains 12 and 03.

Behaviors begin changing for lower m_i's. Figure 14 shows the prevalence and susceptibility curves when all the m_is are 0.25. Infections with the seeded 02 strain still take off and recover to compartment R2323. But that (level K=2) immunity to alleles 0 and 2 is now weak and short-lived. Strain 02 re-infections occur. Furthermore, the infections of strains 12 and 03 that mutate from the 02s (long before the 13s do) are not slowed down as much by immunities to allele 0 and allele 2; they begin to take off before the 13s do. However, that *complementarity effect* is still

strong; there is still enough immunity in the system to allow the 13s to eventually surpass the 12s and 03s.

When population immunity is weak (m_i's < 0.4), strains 12 and 03 – just one mutation from 02 – gain a foothold in the population. A strong complementarity factor requires sufficiently strong immunity in the population.

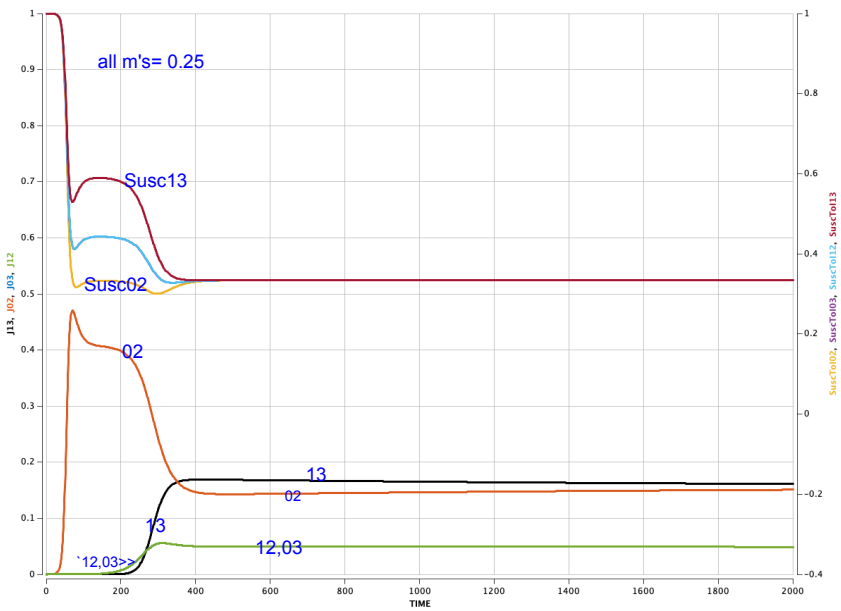

Figure 14. When all immunities are low (m_i's = 0.25), gaps disappear; all four strains appear and stabilize quickly; the seed and its complement still dominate.

VI.C Changing individual m_i's

We now allow for variation among the alleles' immune responses, allowing different m_i for different alleles i. After all, higher transmissibility and more successful immunity evasion are the key concerns as the virus evolves to new strains.

We first examine what happens to prevalence patterns as we drop one m_i for 0.7 to 0.5 or even 0.3.

Slightly weaker immunity to seeded allele 0 temporarily empowers Strain 03 (m_0 = 0.5, other m_is = 0.7)

We start by dropping m_0 to 0.5. This lower immunity should aid strains 02 and 03. As Figure 15 shows, it does –- to an extent. Epidemic waves of 02 remain at roughly the same height, but the gaps between them are shorter. Strain 03 is a bit stronger and strain 13 a bit weaker; but its role as complement to strain 02 allows strain 13 to soon dominate strain 03, despite 03's immunity advantage. (Figure 15. Seeding effect).

Much weaker immunity to seeded allele 0 permanently empowers strain 03 (m_0 = 0.3, other m_is = 0.7)

Lowering the immunity m_0 to allele 0 even more, say to 0.3, allows strain 03 to overcome this strain 13 complementarity advantage. In the long run, strains 02 and 03 have the highest prevalences. And they do not need strong complements to flourish. (Figure 16. Direct parameter effect).

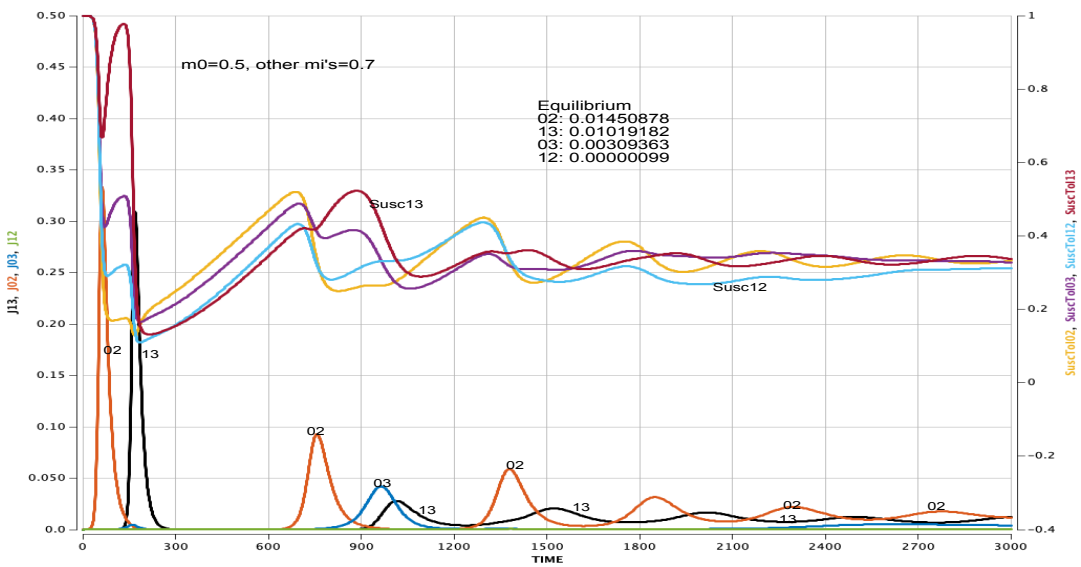

Figure 15. If we lower immunity m_0 to allele 0 a little, strains 02 and 03 do well for a while, but seeded strain 02 and its complement 13 eventually dominate (m_0 = 0.5, other m_i's = 0.7).

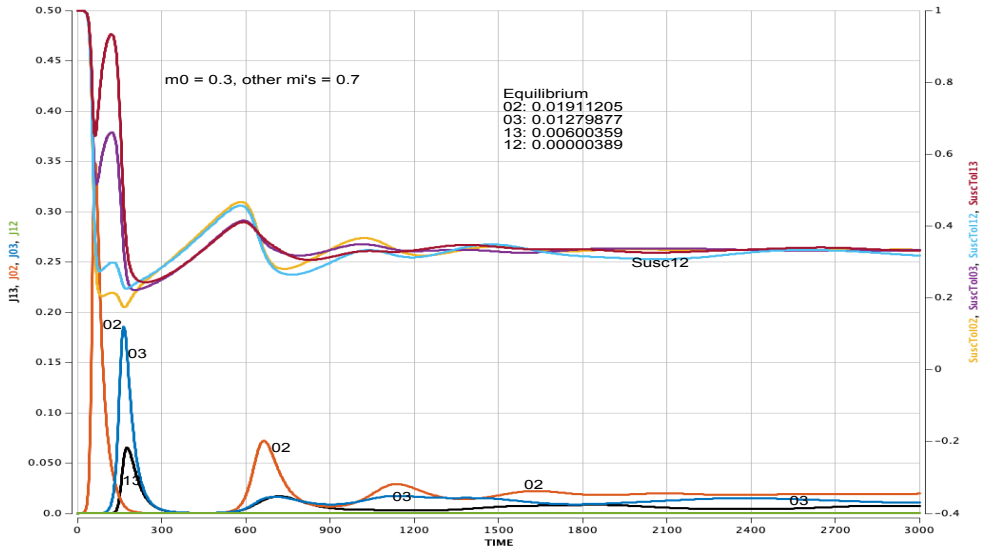

Figure 16. If we lower immunity m_0 to allele 0 a lot, strains 02 and 03 soon dominate (m_0 = 0.3, other m_i's = 0.7).

Decreasing immunity to non-seeded allele 1

Parallel changes occur if we unilaterally lower any of the other m_i's. For example, if we decrease m_1 unilaterally to 0.5, strain 12 is stronger than it was in the basic model. It even has a small epidemic wave just before strain 02's second wave. but it falls behind 13 and 02 in the long run. (Figure S9) If we decrease m_1 further to 0.3, then strain 02 disappears after its initial epidemic wave, and strains 12 and 13 dominate, bolstered by the low immunity to allele 1. See Figure S10.

Reducing two m_i's

If *two* allele-based inherent immunity levels are lower, say at 0.3 rather than 0.7, whether or not the two alleles are at the same virus site makes a big difference.

When both alleles with reduced immunity are at the same site, each of the four strains is equally affected, and no strain gains an advantage from the reduced immune response. Even though reduced immunities affect all strains the same, strains 02 and/or 13 still dominate -- another example of the power of the seeded strain. (Figure S11)

Figure 17 shows the epidemic patterns when the two alleles with reduced immunity are at different sites. If, for example, m_1 and m_3 are reduced, the strain composed of those alleles truly

dominates. Over 31% of the population is infected by strain 13 in the long run. If m_o and m_2 are decreased, the strain 02 strongly dominates. (Direct parameter effect)

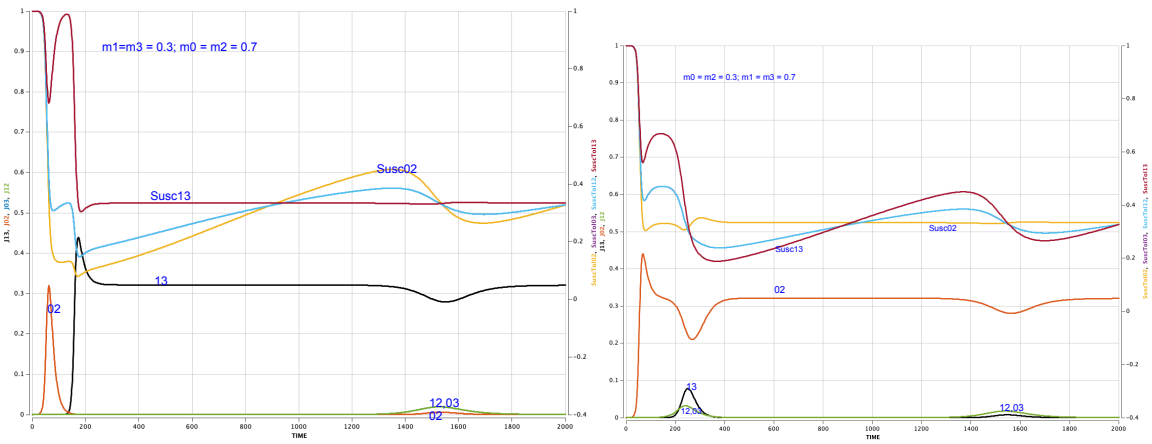

Figure 17. If we lower immunities for two alleles at different sites ($m_1 = m_3 = 0.3$, $m_0 = m_2 = 0.7$ (left panel); $m_0 = m_2 = 0.3$, $m_1 = m_3 = 0.7$ (right panel)), the strain composed of those two alleles strongly dominates.

VI.D. No immune pressure

Finally, we look at the case where there is little or no immunity to the seeded strain 02. We set m_0 and $m_2 = 0$. The result has shown in Figure 18: no successful mutation. As [Sa22] puts it: "if there is no immune pressure, then viral abundance may be high but selection for immune escape is absent."

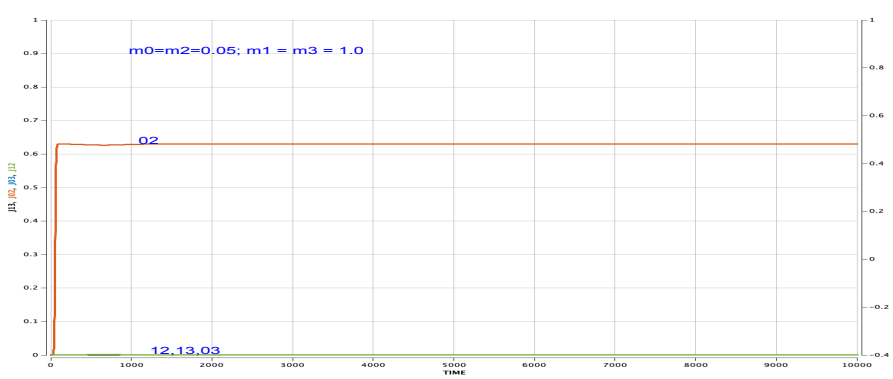

Figure 18. If there is no immunity for the seeded alleles ($m_0 = m_2 = 0$), seeded strain 02 quickly becomes the only positive strain in the pandemic.

- In fact, as long as m_0 and m_2 are =0, strain 02 remains the only positive strain for all time, for *any* values of m_1 and m_3.
- But if m_0, immunity to allele 0, increases just to 0.05, while the other allele-immunity levels m_i remain at 0, then strain 02 decreases and strain 12 (with $m_1 = m_2 = 0$) replaces it as the dramatically dominant strain (Figure 19, left). (Direct parameter effect)
- Similarly, if only m_2 increases from 0, then strain 03 replaces strain 02.
- If both m_0 and m_2 increase just to 0.05, while m_1 and m_3 remain at 0, then strain 13 replaces strain 02 as the sole positive strain (Figure 19, center). In this case strains 03 and 12 have a single epidemic wave before strain 13 takes over. (Direct parameter effect.)
- If all four m_is increase to 0.05, then strain 02 loses some of its prevalence, while the other three strains increase a bit over time (Figure 19, right). (Seeding effect)

For low levels of immunity, small changes in immunity lead to very different prevalence patterns, including which strain plays the role of the strongly dominating strain.

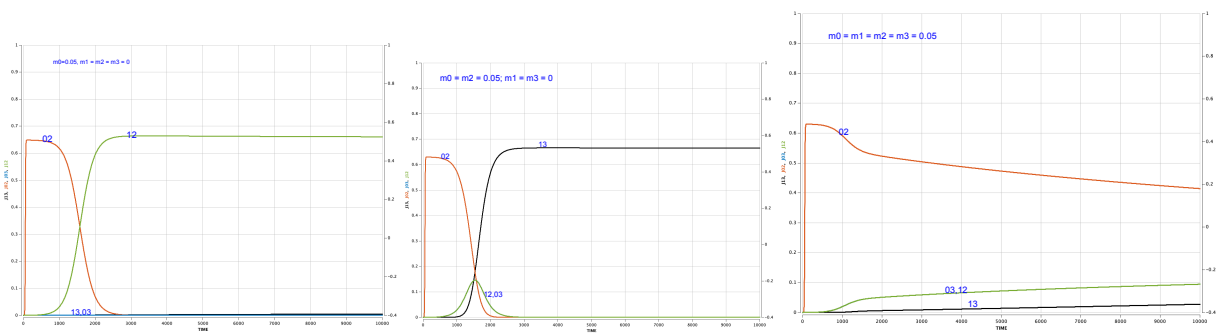

Figure 19. Left: Starting with all m_i=0 as in Figure 18, a 0.05 increase only to m_0 leads to stain12 replacing 02 as dominant strain. Center: a 0.05 increase to both m_0 and m_2 leads to domination by strain 13. Right: a 0.05 increase in all four m_is leads to a decrease in strain 02 and small increases in the other strains.

This discussion about the immunity parameter m_j highlights three of our main points:

- the key role that immunity plays in the population spread of infection,
- the advantage of focusing on alleles instead of the usual focus on strains, and
- the persistent strength of the initial seed, especially for homogenous parameter choices.

VII. Adding a joint immunity factor has little effect when allele-based immunity is strong, but a large effect when allele-based immunity is weak.

In this Section, we examine the effects of including a *joint* immunity factor m_{ij} along with allele immunity factors m_i. The effect of this change depends dramatically on the value for the single allele immunity level m_i.

If the m_i's are reasonably large, adding some joint immunity effects has little effect (Fig. S9 in the Supplementary Material). Differences do appear when the joint effects are added to the low allele-based immunity $m_i = 0.3$. Figure 20 displays prevalences and susceptibilities when the joint immunities m_{ij} increase from 0 to 0.2 to 0.4, while the allele-relative immunities remain low at 0.3. As the joint immunities increase from 0 to 0.4:

- The sizes of the peaks of the initial 02 and 13 waves decrease.
- However, the prevalences of (non-seeded) strains 12 and 03 *increase.*
- When the joint m_{ij}'s reach 0.4,
    - strains 12 and 03 dominate seeded strain 02 and its complement 13 in the short run, and
    - all four strains have the same long run equilibrium.

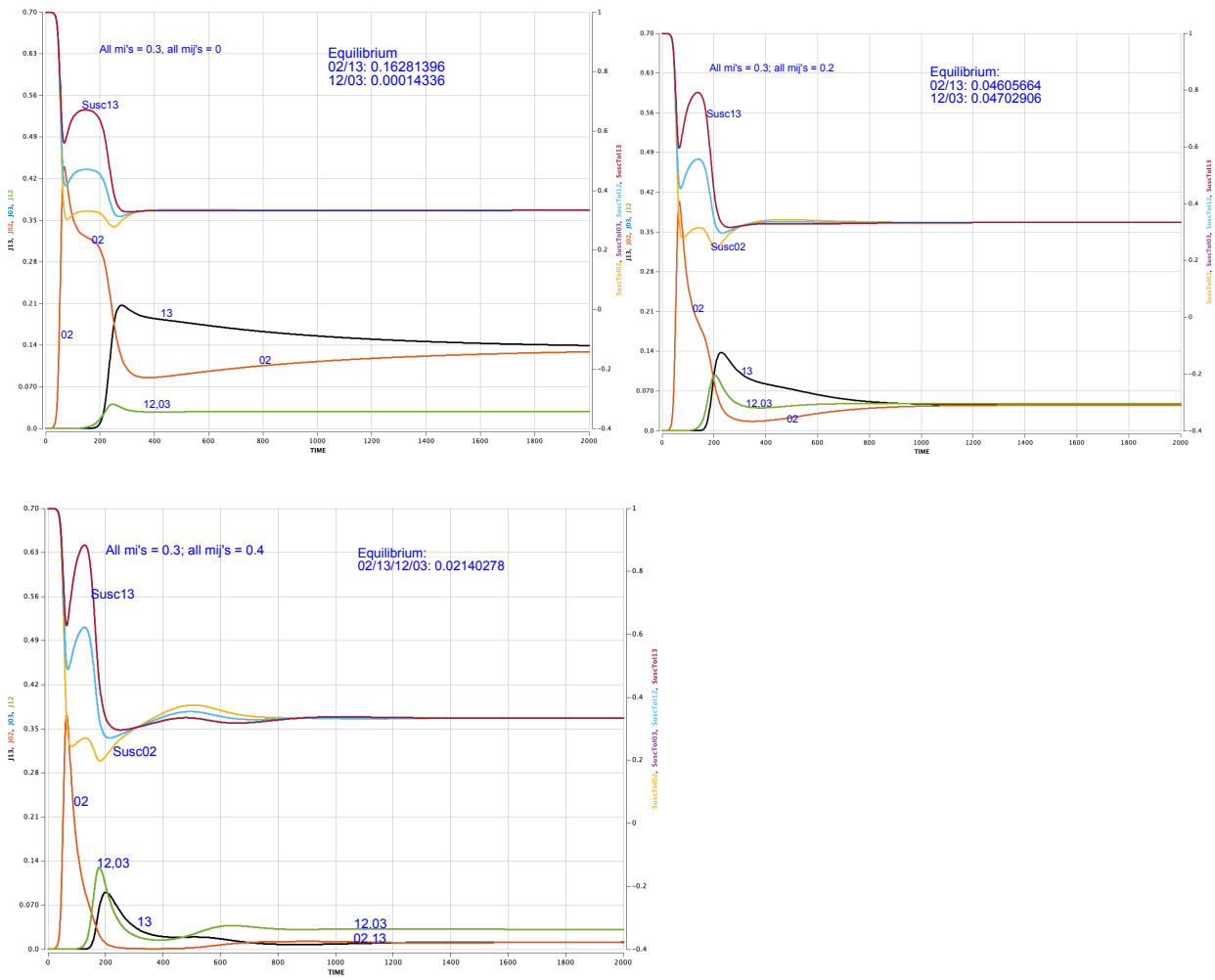

Figure 20. For low values of immunity response m_i to all alleles i, the advantages of mutant proximity outweigh the advantages of complementarity as the joint m_{ij}'s increase (all m_i's=0.3, with all m_{ij}'s = 0 (top left), all m_{ij}'s = 0.2 (top right), all m_{ij}'s = 0.4 (lower)).

VIII. Comparing strain-based immunity to allele-based immunity

VIII.A An Example

Earlier, we examined in detail the allele-based situation in which

- Case 1: each relative immunity level m_i to allele i was 0.5 and there was no joint immunity (all m_{ij}'s =0).

We now contrast this to the all joint-immunity case where

- Case 2: *all* the allele-based immunity levels m_i equal 0 and all the joint-immunity levels are set at 0.75.

We chose m_{ij}'s = 0.75 so that both cases have the same initial strain 02 epidemic wave, peaking at t=64 at prevalence 35.6%. The resulting graphs are displayed in Figures 21 and 22.

Recall that the susceptibility multiplier for transmission when an infecting strain with alleles i and j (i=0 or 1, j = 2 or 3) contacts a susceptible in $R(\kappa_0, \kappa_1, \kappa_2, \kappa_3)$ is

$$S = (1-m_i\,\kappa_i)(1-m_j\,\kappa_j)(1-m_{ij}\,\kappa_i\,\kappa_j).$$

This factor S modifies the pure transmission rate $b_i + b_j$.

In our allele-based -immunity case 1, all $m_i = m_j = 0.5$ and all $m_{ij} = 0$.
In our strain-based-immunity case 2, all $m_i = m_j = 0$ and all $m_{ij} = 0.75$.
In the allele-based case 1, when an infectee with strain 02 contacts a (recently recovered) susceptible in R2323, the susceptibility factor is:

$$(1-(0.5*1))*(1-(0.5*1))*(1-0) = 0.5 * 0.5 = 0.25.$$

In the strain-based case 2, the corresponding susceptibility is:

$$(1-0)*(1-0)*(1-(0.75*1*1)) = 0.25.$$

This gives a second reason for our choice of $m_{ij} = 0.75$ for this comparison.

As usual, in both cases, the seeded strain 02 defines the first epidemic wave, with recoveries entering compartment R2323. The first mutations from strain 02 are strains 12 and 03. How do these strains fare with the recently recovered in compartment R2323 who dominate the population early?

In the allele-based case 1, when an infective with strain 12 encounters a member of R2323, the sustainability factor is: $(1-(0.5*0))*(1-(0.5*1)) = 0.5$,
which cuts the pure transmission rate (b_1+b_2) in half.

In the strain-based case 2, the corresponding factor is: $(1-(0.75*1*0)) = 1$,
yielding no reduction in transmission rate (b_1+b_2).

Transition to a strain 12 infection is twice more likely in Case 2 than in Case1. Basically, the recovering 02s have some immunity to allele 0 and allele 2 in case 1, but not in case 2.

In allele-based Case 1, the stifling of strains 12 and 03 -- that are just one mutation away from the seeded strain 02 -- allows strain 13, which is two mutations away from 02 but very transmissible to those in R2323, to flourish. So, a strain 13 epidemic wave arises and thrives, and then the recovered in compartment R3232 are easily infected by strain 02-infectives and the cycle continues.

In the strain-based Case 2, the higher susceptibility of those in R2323 to strains 12 and 03 leads to 12 and 03 epidemic waves. This wave is strengthened by the fact that strains 12 and 03 complement each other and have identical prevalences. This transmissibility strength keeps the pair 12/03 dominant for quite a while, so that no further epidemic waves occur.

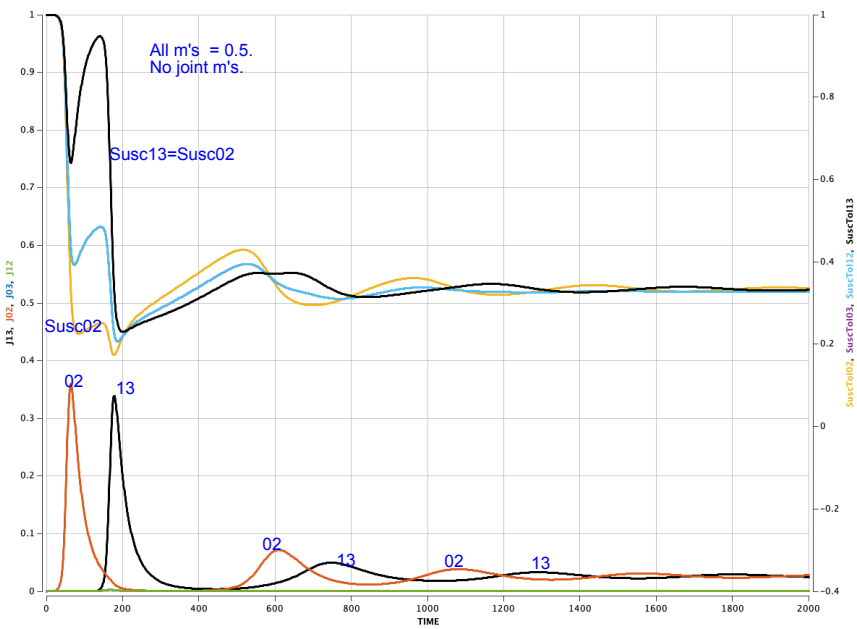

Figure 21. Case 1. All m_i's = 0.5, all m_{ij}'s =0. With independent allele-based immunities, the recovery-induced immunities for alleles 0 and 2 prevent strains 12 and 03 from gaining a foothold before complementary strain 13 arrives and takes off.

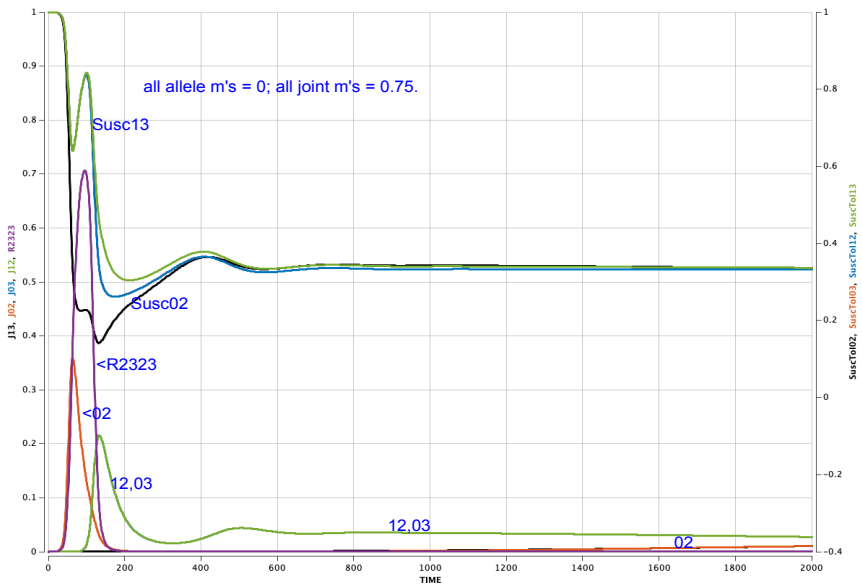

Figure 22. Case 2. All m_i's = 0, all m_{ij}'s = 0.75. In the strain-associated case 2, early forming strains 12 and 03 take off and dominate the later forming 13 strain, and even the seeded 02 strain.

VIII.B More General Comparisons of all joint immunities vs. all allele-based immunities

High to Moderate Immunity levels

Earlier, when we examined the allele-based case for all m_i's between 0.5 and 0.9 and no joint immunity, seeded strain 02 and its complement 13 dominate dramatically. If we reverse this by keeping the m_i's at 0 and let all the m_{ij}'s range from 0.5 to 0.9, we find the same prevalences, as in Figure 22 above. The initial 02 wave is succeeded not by a wave of its complementary 13 strain, but by waves of strains 12 and 03, both only one mutation away from 02. See Figure 23.

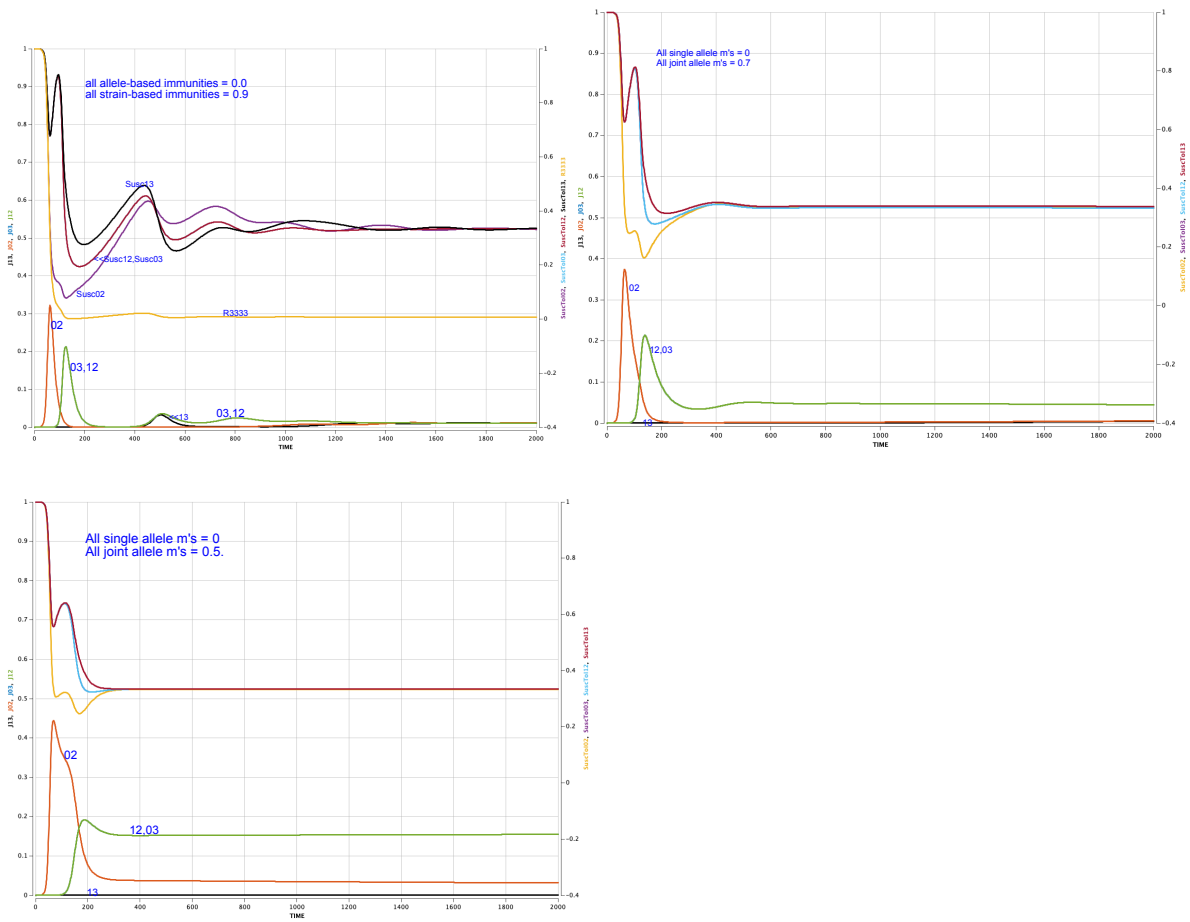

Figure 23. When all allele-based immunities are set to 0 and the stain-based immunities vary from 0.9 (top left) to 0.7 (top right) to 0.5 (bottom), strains 12 and 03 quickly dominate.

Low levels of immunity

We next examine the case where all the m_i's and m_{ij}'s are small. If $m_{02} = 0$, we find as we did for $m_0 = m_2 = 0$ in Figure 18, only the seeded strain 02 is present and it stabilizes at prevalence 67%, no matter the values of m_{12}, m_{13}, or m_{03}. Everyone becomes infected at these low immunities.

There are significant differences, however, between:

- Allele-based case 1: $m_0 = m_2 = 0.05$, $m_1 = m_3 = 0$, all $m_{ij} = 0$, and
- Strain-based case 2: $m_{02} = 0.05$, other m_{ij}'s = 0, all $m_i = 0$.

In Case 1, as seen in Figure 19b, strains 12 and 03 rise and fall after the initial 02 wave and strain 13 becomes the dominant strain. Figure 24 presents the strain-based immunities in Case 2: strains 12 and 03 replace the seeded strain 02; strain 13 has no presence.

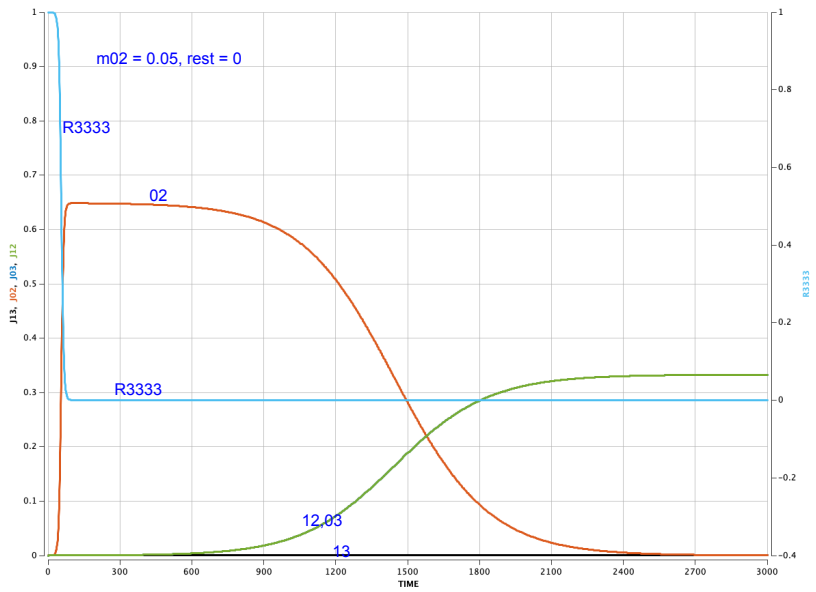

Figure 24. When $m_{02} = 0.05$, $m_{12} = m_{13} = m_{03} = 0$, strains 12 and 03 replace the seeded strain 02. The path of R3333 shows that all are soon infected.

Comparison Conclusion

In our comparisons of allele-based immunity models with strain-based immunity models, we found that for the allele-based case (all m_{ij}'s = 0), the seeded strain 02 mutated to the strain furthest from it, the strain with the lowest cross-immunity to 02. In the strain-based immunity model, strain 02 mutated to the nearest strains, those just one allele-change away and prevalences stabilized there.

Mutations in the current CoVid-19 pandemic seem more like the former case than the latter. The transition to the Delta strain has nine mutations in the spike protein itself and thirteen mutations in the added regions. The transition from Delta variant to Omicron has more than 50 known mutations, 32 of which are in the spike protein. [Ch23, Ing21]

We conclude our study of parameter changes, by considering the impact of multiple strains as initial seeds and the possible impact of ADE.

IX. Multiple Initial Seeds

Extensive calculations suggest that the solutions of the differential equations in Part Two with the parameter values in Part Three converge to one of two locally stable equilibria:

- one with strains 02 and 13 at 0.01173624, and strains 03 and 12 at 0.00000109, which we call E_{02}*, and
- one with strains 12 and 03 at 0.01173624, and strains 02 and 13 at 0.00000109, which we call E_{03}*.

The positive orthant is divided into the basin of attraction B_{02}* of E_{02}* and the basin of attraction B_{03}* of E_{03}*, with a hypersurface separating the two. If the initial conditions ("seeds") are one or both of strains 02 and 13, the system converges to E_{02}*. If the initial conditions ("seeds") are one or both of strains 12 and 03, the system converges to E_{03}*. On the other hand, if the initial seeds are equal amounts of strains 02 and 03 or strains 12 and 13, the long run equilibrium lies on the separating hypersurface; any small change or rounding restriction will send the trajectory to E_{02}* or E_{03}*.

Suppose three of the four strains are equally seeded, say 02, 12, and 13, but not 03,
The excluded strain 03 has a significant epidemic wave around t=650 followed by an epidemic wave of its complement 12, but 03 and 12 fade away thereafter. The long run equilibrium is E_{02}*. See Figure 25, and the detailed description in the Supplementary Material.

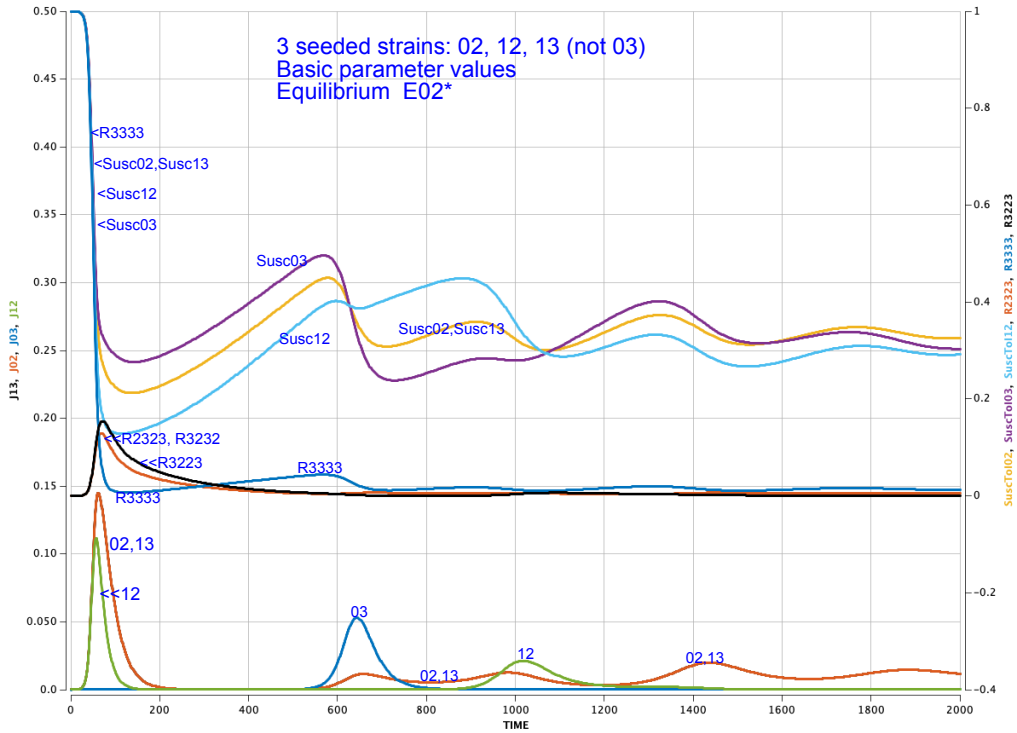

Figure 25. If all strains except 03 are initial strains, strain 03 still has a significant epidemic wave before petering out.

This is an interesting example because there are epidemic waves for all four strains, even though 02 and 13 dominated in the beginning and at the end. The waning of R3232 and R2323 lead to resurgences of the Susc02 and Susc13 curves.

Waning immunities play an important role in this case. If in the scenario of Figure 25 we slow all the waning rates from 0.003 to 0.0003, strains 12 and 03 dominate the pandemic after the initial waves (Figure 26) even though strain 03 started at a major disadvantage. Waning in an important phenomenon.

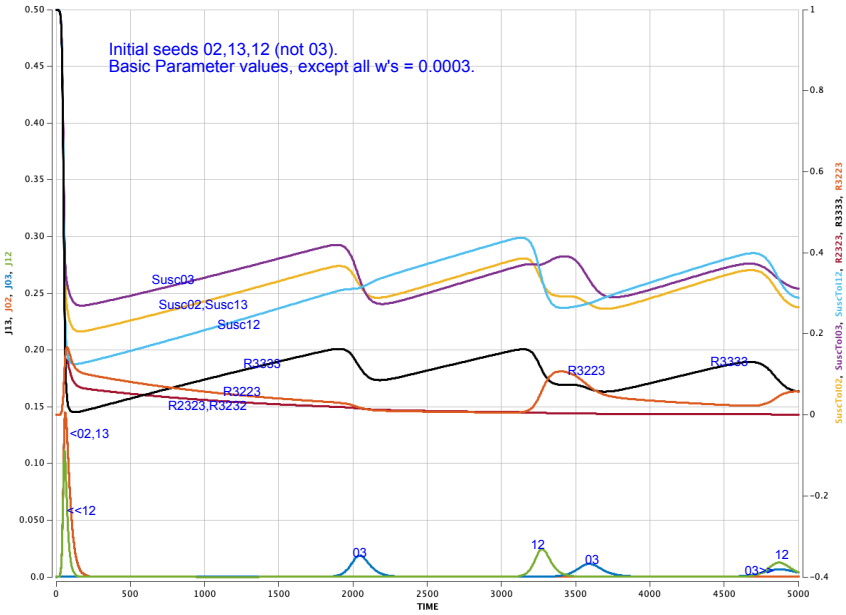

Figure 26. If all strains except 03 are initial strains and waning is slowed to 0.0003, strains 03 and 12 become the dominant strains (with the highest susceptibility curves).

X. Antibody Dependent Enhancement (ADE) d_j.

We conclude our study of parameter changes, by considering the possible impact of ADE.

Antibody-Dependent Enhancement (ADE) occurs when pre-existing non-neutralizing antibodies, which had arisen in response to previous infection, lead to enhanced new infection by facilitating the virus's cell entry. ADE has played a major role in epidemics of the RNA-virus dengue [Fe99] and has been postulated as a possible mechanism for SARS-CoV-2 cell entry [Ik23, Song23, Sa20b].

In our model, the ADE terms d_i are similar in the infection dynamics to the joint transmissibility terms b_{ij} in that both increase transmissibility in an additive fashion. Unlike the joint transmissibility terms, an ADE term d_i only kicks in when the potential infectee has experienced an infection with a strain containing allele i. For example, when an infective with strain 02 encounters a susceptible in R2311, the transmissibility rate is:

$$((b_0+b_2+b_{02}) + (d_0+0+d_2+d_3));$$

those in R2311 have never been infected by a strain that includes allele 1.

In Figure 27, we maintain the four b_is at 0.2 and the b_{ij}s at 0. We increase the ADE factor d_i from 0 to 0.07. To compare curve heights better, we increased the scale of the left vertical axis. Additional Figures in the Supplementary Material.

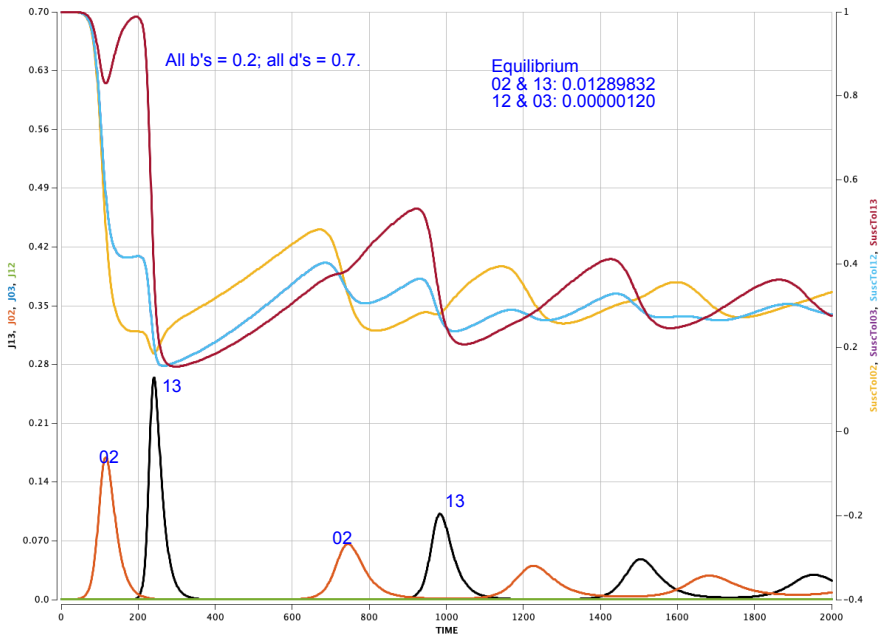

Figure 27. Strain 13 has especially high peaks if an ADE factor(d_i = 0.7) is added to the basic model.

The inclusion of ADE terms has little effect on the first strain 02 epidemic wave, but it does increase the levels of later 13 epidemic waves. With increasing d_i, not only does the peak of the first 13 wave increase, but so does the peak of the second and third 13 waves. Higher levels of infections in the gaps means that the Susc13 curves reach their threshold sooner, and the gaps become shorter.

This leads to the rare situation in which the height of the initial 02 epidemic is smaller than the height of the succeeding 13 epidemic. It opens the possibility that with more realistic models one can use population infection patterns to detect the presence of an ADE effect.

It is natural to ask what happens if we decrease the b_is even further, say to 0.1, while keeping the d_is high. Figure S17 in the Supplementary Material shows the answer: basically, nothing

happens, no infection at least for the first 6,000 time-steps. The higher values of the b_is are needed to achieve $R_0 > 1$ so the epidemic can get started.

Part 5 Discussion

A. Summary

Our default model had the seeded strain 02 and its complement 13 dominating the pandemic, with an 02 wave quickly followed by a 13 wave and then a gap before the next 02 wave appeared. We asked how the appearance of strains with alleles with different properties would affect this pattern. In some cases, barely; for example: changes in the mutation rate, changes in the residual immunity after waning, small changes in transmissibility or in background population immunity.

On the other hand, some changes allowed new strains to dominate; for example, strains with a longer infection duration, strains with an allele with more rapidly waning immunity, strains with an allele for which the population has significantly lower immunity. And some changes in transmissibility or in immune-response led to more nuanced outcomes.

We present a summary here ("J02" stands for "strain 02"):

SUMMARY OF EFFECTS OF PARAMETER CHANGES

Parameter(s) to increase	Effect
all b_i's	Higher peaks and shorter (between wave) intervals
Small increase in b_0 or b_2	Higher peaks and shorter intervals
Large increase in b_0 or b_2	**J12 and J03 replace J02 and J13 in LR**
b_1 or b_3	J02 and J13 continue to dominate
b_0 & b_2 or b_1 & b_3	J12 and J03 eventually dominate

b_0 & b_1 & b_3	J02 and J13 dominate
contact rate c	Higher peaks and shorter intervals
all durations ($1/v_{hj}$)	Higher peaks and shorter intervals
one duration ($1/v_{hj}$)	**strain J(h,j) dominates**
waning rate w_{i*} for all i	Shorter intervals
waning rate w_{i*} for allele i	Strains with allele i dominate
all nu_i	Lower peaks and **longer** intervals
nu_0 or nu_2	Longer intervals between J02 waves
small Δ in drift rate f	Little change
f large (>0.01)	**Strains equalize in strength**
initial seed change	J02 & J13 dominate or J12 & J03 dominate
Seed all but J03, and increase allele 0 waning rate	**J03 strengthens**
Decrease all m_i	Higher peaks and shorter intervals
All m_i very low	All 4 strains active and stable; J12 & J03 stronger
Small decrease only in m_0	Shorter J02 intervals, small J03 increase
Big decrease only in m_0	**J02 and J03 dominate**
Big decrease only in m_1	J12 and J13 dominate
Decrease m_1 & m_3	J13 dominates
Decrease m_0&m_1 or m_2&m_3	seeded J02 still dominates

Set $m_0 = m_2 = 0$	**Only J02 present; no mutation**
High m_{ij}'s, low m_i's	**Mutation more likely to nearby strain (1 allele away) than to complementary strain (2 alleles away)**

B. **Advantages of an allele-based model**

This paper addresses the challenge of Saad-Roy, Metcalf, and Grenfell [Sa22] for "epidemiological studies that simultaneously quantify immunity and transmission across populations and at different biological scales … along with the development of cross-scale modeling approaches." Models that focus on strains or variants without going deeper into viral structures do not meet this call for "phylodynamic models." This paper carries out this cross-scale approach by using the alleles at virus sites as the basic building blocks.

There are many good reasons to prefer an allele-based approach over a strain-based approach.

1) In a strain-based approach, a mutation is an ad-hoc transformation from one strain to another, possibly determined by some distance criterion or a link on an ad-hoc network of strains. In an allele-based approach, a mutation is the replacement of one allele by another at the same site.
2) In a strain-based approach, the distance between two strains is some measure of distance on that network of strains. In an allele-based approach, one can measure the "distance" by counting how many allele flips occur to get from one strain to the other.
3) We want our modeling to learn from and hopefully give guidance to the laboratory work that has thus far examined and laid bare the very nature of coronaviruses. An allele-based approach is far closer to the nucleotides and other cellular and subcellular structures on which this lab work has focused.

Our model allows for epistasis in transmission rate and immune response through the use of joint-transmissibility and joint-immunity parameters. If we turn on these parameters, turn off the individual allele-based transmissibility and immunity parameters, and keep all the other parameters homogeneous across alleles, we can recapture a strain-based approach. We found very different results when using this strain-based approach compared to using the allele-based

approach. Most noticeably, the strain-based version had a strong proclivity to have the epidemic wave of the any strain directly followed by an epidemic wave of a strain just one mutation away – even when parameters were set to favor distant strains with lower cross-immunity. In the allele-based approach, an epidemic wave was usually followed by a wave of the strain with least cross-immunity. In the SARS-CoV-2 pandemic, the mutations from the alpha-variant to the delta-variant and from the delta-variant to the omicron-variant consisted of many allele changes, over twenty in the first case and fifty in the second [Ch23, Ing21, Ko22].

C. The Seeding Advantage

Allele-specific immunity generates an imprinting of initially introduced alleles in a population, an imprinting that sustains those initially introduced alleles in circulation for a period of time. It leads to an ongoing dominance of the founding strain even though that strain does not have characteristics that make it more transmissible or more immunity-resistant than other strains. This temporarily fixes strains in a population in a way that is different from how high transmissibility or how high immunity-escape temporarily fixes strains at high levels. This fixation acts at an allele level through immunity actions that are allele-specific.

We found many instances of the Seeding Effect, parameter changes for which seeded strain 02 remained dominant as the pandemic evolved – sometimes in the face of parameter changes that should have favored other strains.

D. ADE

Our approach has also suggested populations patterns of virus evolution that may indicate the existence of ADE in the transmission process.

E. **The Critical Role of Immunity**

A major take-home from our allele-based approach is the critical role of immunity in the prevalence patterns.

At the beginning of the previous section, we identified four different ways that a new strain can prosper after a parameter change: the direct parameter effect, complementarity effect, seeding effect, and adjacency effect. Focusing on these effects allowed us to begin a program to carry out the challenge in [Sa22] to "simultaneously quantify immunity and transmission across populations."

We found, for example, that the waning rate, infection duration, and strong loss of immunity led to the most impressive direct parameter effects. However, the complementary effect played big in our allele-based approach. Since this effect is built on an immune response, it requires immunity to play a role at least as important as the transmissibility role in the evolution of the pandemic. The central role of immunity in our model underlies the observation [Sa24a] that "the strength of immunity shapes immune-epidemiological dynamics at multiple resolutions."

F. Future Work

Our model encompasses complexities affecting virus transmissibility, host immunity, population susceptibility, antibody dependent enhancement of infection, joint effects across specific alleles on transmissibility, joint effects across alleles on immunity, waning immunity, recovery rate, and other details. We model this diversity of forces not because their contributions to evolution are clearly understood, but because their interactions need to be clarified. In order to deal with these phenomena analytically and to understand more systematically the drivers of the SARS-CoV-2 pandemic, we constrained our dynamical model to the unrealistic situation of two mutation sites with two mutating alleles at each site.

An obvious next step will be to increase the number of such sites in our model. But even small increases lead to unwieldy numbers. Consideration of just ten sites with two alleles at each site yields a system with 20 alleles and 1,024 strains. Keeping track of immune history, say with five levels of waning in a compartmental model like ours, leads to roughly 10^{14} compartments of susceptibles. It may be more reasonable to aim for an in-between goal of four sites with four immunity levels; that would yield only $\sim 4^8 = 65{,}536$ compartments of susceptibles.

This, our first allele-based model, focused on biological processes and avoided sociological concerns since we are especially interested in relating our model to the work of laboratory scientists. Future projects will incorporate population effects, such as age structure, and heterogeneous contact rates and immune responses. Many references have stressed the importance of including contact structures [So23, El20], population behavior [Es24, Sa24a], and vaccination effects [Dys21, Sa20, Sa21a, Sa23, Sa24a, Sa21b] in models of the spread of SARS-CoV-2. Relatedly, this work sets the stage for exploring how insights from more realistic and complete allele-based models can help to inform seasonal vaccine design for respiratory illnesses such as COVID-19 but also influenza.

Finally, in line again with the challenge in [Sa22], we are especially eager to elaborate the processes in the model to integrate with knowledge generated by laboratory studies, including ultimately a better understanding of how the RNA dependent RNA polymerase (RdRp) of coronaviruses influences coronavirus evolution.

Three-quarters of the SARS-CoV-2 genome encodes how genetic sequences are translated into specific sequence codes in the virus. The size and function of this translation site are specific to coronaviruses. This is the region in which the enzyme RdRp plays a key role. Our allele-based model does not formulate any effects of RdRp, but it can help discover such effects and indicate how large they may be by formulating processes at the level of allele specific effects.

REFERENCES

[Bu25] James Bull, Katia Koelle, Rustom Antia. Waning immunity drives respiratory virus evolution and reinfection. Evolution, Medicine, and Public Health, 2025.

[Ch23] S Chatterjee et al. "A detailed overview of SARS-CoC-2 Omicron: Its subvariants, mutations and pathophysiology, clinical characteristics, immunological landscape, immune escape, and therapies." Viruses 15, 167-. 2023.

[Da24] Bernadeta Dadonaite, Jack Brown, Teagan E. McMahon, Ariana G. Farrell, Marlin D. Figgins, Daniel Asarnow, Cameron Stewart, Jimin Lee, Jenni Logue, Trevor Bedford, Ben Murrell, Helen Y. Chu, David Veesler & Jesse D. Bloom
Spike deep mutational scanning helps predict success of SARS-CoV-2 clades
Nature Vol 631, 617-625. 18 July 2024.

[Dys21] Louise Dyson Edward M. Hill , Sam Moore Jacob Curran-Sebastian, Michael J. Tildesley, Katrina A. Lythgoe, Thomas House , Lorenzo Pellis, and Matt J. Keeling. Possible future waves of SARS-CoV-2 infection generated by variants of concern with a range of characteristics. Nature Communications (2021) 12:5730.

[El20] Rashad Eletreby, Yong Zhuang, Kathleen M. Carley, Osman Yagan, and H. Vincent Poor. The effects of evolutionary adaptations on spreading processes in complex networks. PNAS 117, 5664-5670. 2020.

[Es24] Baltazar Espinoza, Chadi Saad-Roy, Bryan Grenfell, Simon Levin, Madhav Marathe. Adaptive human behaviour modulates the impact of immune life history and vaccination on long-term epidemic dynamics. Proc. R. Soc. B 291: 20241772. 2024

[Fe99] N. Ferguson, R. Anderson, S. Gupta. The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens.
Proc. Natl. Acad. Sci. U.S.A. 96, 790-794 (1999). doi:10.1073/pnas.96.2.790

[F18] Fisher, R.A. The correlation between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinburgh. 52,399–433, 1918.

[Getz21] WM Getz, R Salter, L Luisa Vissat, JS Koopman, CP Simon. A runtime alterable epidemic model with genetic drift, waning immunity and vaccinations.
Journal of the Royal Society Interface 18 (184), 20210648, 2021.

[Gi20] Giulia Giordano, Franco Blanchini, Raffaele Bruno, Patrizio Colaneri, Alessandro Di Filippo, Angela Di Matteo and Marta Colaneri.
Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nature Medicine VOL 26 | 855–860, June 2020. www.nature.com/naturemedicine

[H08] Hardy, G. H. "Mendelian Proportions in a Mixed Population." Science. 28 (706): 49–50. (July 1908).

[Ik23] Nobunao Ikewaki, Gene Kurosawa, Gary A. Levy, Senthilkumar Preethy , Samuel J.K. Abraham. “Antibody dependent disease enhancement (ADE) after COVID-19 vaccination and beta glucans as a safer strategy in management.”
Vaccine 41, 2427-2429. (2023).

[Ing21] Nicholas E. Ingraham, David H. Ingbar . 2021. The omicron variant of SARS-CoV-2: Understanding the known and living with unknowns.
Clin. Transl. Med. 2021 https://doi.org/10.1002/ctm2.685

[J-S] JA Jacquez, CP Simon. Qualitative Theory of Compartmental Systems
S.I.A.M. Review, 35, 43—79 (1993).

[Ko22] Katia Koelle, Michael A. Martin, Rustom Antia, Ben Lopman, Natalie E. Dean. The changing epidemiology of SARS-CoV-2. Science 375, 2022.

[K-S21] JS Koopman, CP Simon, WM Getz, R Salter. Modeling the population effects of escape mutations in SARS-CoV-2 to guide vaccination strategies Epidemics 36, 100484, 2021.

[Li16] Fang Li. Structure, Function, and Evolution of Coronavirus Spike Proteins
Annu Rev Virol. 2016. 237-261.

[Me23] Meijers et al. Population immunity predicts evolutionary trajectories of SARS-CoV-2, Cell. (2023). https:// doi.org/10.1016/j.cell.2023.09.022

[Mo21] Sam Moore, Edward M Hill, Michael J Tildesley, Louise Dyson, Matt J Keeling. “Vaccination and non-pharmaceutical interventions for COVID-19: a mathematical modelling study.” Lancet Infect Dis 2021; 21: 793–802.

[Sa20] Chadi M. Saad-Roy1, Caroline E. Wagner, Rachel E. Baker, Sinead E. Morris, Jeremy Farrar6, Andrea L. Graham, Simon A. Levin, Michael J. Mina, C. Jessica E. Metcalf, Bryan T. Grenfell. Immune life history, vaccination, and the dynamics of SARS-CoV-2 over the next 5 years. Science 370, 811–818 (2020)

[Sa21a] Saad-Roy CM, Levin SA, Metcalf CJE, Grenfell BT. Trajectory of individual immunity and vaccination required for SARS-CoV-2 community immunity: a conceptual investigation. J. R. Soc. Interface 18: 20200683. 2021.

[Sa21b] Chadi M. Saad-Roy, Sinead E. Morris, C. Jessica E. Metcalf, Michael Mina, Rachel E. Baker, Jeremy Farrar, Edward Holmes, Oliver Pybus, Andrea L. Graham, Simon A. Levin, Bryan T. Grenfell, Caroline E. Wagner.
Epidemiological and evolutionary considerations of SARS-CoV-2 vaccine dosing regimes. Science 372 eabg8663. 2021.

[Sa22] Chadi M Saad-Roy, Jessica E. Metcalf, Bryan T. Grenfell. “Immuno-epidemiology and the predictability of viral evolution.” Science 376, 1161-1162. 2022.

[Sa23] Chadi M. Saad-Roy, Sinead E. Morris, Rachel E. Baker, Jeremy Farrar, Andrea L. Graham, Simon A. Levin, Caroline E. Wagner, C. Jessica E. Metcalf, Bryan T. Grenfell. Medium-term scenarios of COVID-19 as a function of immune uncertainties and chronic disease. Royal Society Interface (2023) 20:20230247.

[Sa24a] Chadi M. Saad-Roy, Sinead E. Morris, Mike Boots, Rachel E. Baker, Bryan L. Lewis, Jeremy Farrar, Madhav V. Marathe, Andrea L. Graham, Simon A. Levin, Caroline E. Wagner, C. Jessica E. Metcalf, Bryan T. Grenfell. Impact of waning immunity against SARS-CoV-2 severity exacerbated by vaccine hesitancy. PLOS Computational Biology. August 5, 2024.
https://doi.org/10.1371/journal.pcbi.1012211

[Sa24b] Chadi M. Saad-Roy, Andy White, Mike Boots. The evolution of post-infection mortal. Proc. R. Soc. B 291: 20241854.

[Song23] Song, H., Yuan, Z., Liu, S. et al. Mathematical modeling the dynamics of SARS-CoV-2 infection with antibody-dependent enhancement.
Nonlinear Dyn 111, 2943–2958 (2023).
https://doi-org.proxy.lib.umich.edu/10.1007/s11071-022-07939-w

[So23] Marisi Sood, A. Sridhar, R. Eletreby, CW Wu, Simon Levin, O. Yagan, H Poor. Spreading processes with mutations over multilayer networks. PNAS 120 (24) 2023.

[Te20] J. A. Tetro. Is COVID-19 receiving ADE from other coronaviruses?
Microbes Infect. 22, 72-73 (2020). doi'l0 1016/j micinf 2020 02.006

[To24] Xin Tong, Benjamin Kellman, Maria-Jose Avendano, Maanasa Mendu, Jeff C. Hsiao, Eileen Serrano, et al. 2024. Humoral waning kinetics against SARS-CoV-2 is dictated by disease severity and vaccine platform. medRxiv preprint doi: https://doi.org/10.1101/2024.10.17.24315607.

[Top24] E Topol. “Why We Need an Immunome”
(July 21, 2024) Ground Truths. https://erictopol.substack.com/

[Wa21] Caroline E Wagner, Chadi M Saad-Roy , Sinead E Morris, Rachel E Baker, Michael J Mina, Jeremy Farrar, Edward C Holmes, Oliver G Pybus, Andrea L Graham, Ezekiel J Emanuel, Simon A Levin, C Jessica E Metcalf , Bryan T Grenfell . Vaccine nationalism and the dynamics and control of SARS-CoV-2
Science 2021 Sep 24;373(6562):eabj7364. 2021.

[W08] Weinberg, W. "Über den Nachweis der Vererbung beim Menschen". Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg. 64: 368–382. (1908).

[Xu15] GJ Xu, et al. Comprehensive serological profiling of human populations using a synthetic human virome. Science 348 (5 Jun 2015). DOI: 10.1126/science.aaa0698

[Yip14] M. S. Yip, N. H. Leung, C. Y Cheung, P.H. Li, H. H. Lee, M. Daeron, J. S. Peiris, R. Bruzzone, M. Jaume. Antibody-dependent infection of human macrophages by severe acute respiratory syndrome coronavirus. Viral. J. 11, 82 (2014)