Immunity-driven evolution of virulence and diversity in respiratory diseases

Johan A.J. Metz and Barbara Boldin

DOI:
https://doi.org/10.1093/evolut/qpad145

Evolution, 2023, **77(11)**, 2392–2408
https://doi.org/10.1093/evolut/qpad145
Advance access publication 18 August 2023
Original Article

OXFORD

Immunity-driven evolution of virulence and diversity in respiratory diseases

Johan A.J. Metz[1,2,3,4,5] **and Barbara Boldin**[6]

[1]Advancing Systems Analysis Program, International Institute of Applied Systems Analysis (IIASA), Laxenburg, Austria
[2]Mathematical Institute, Leiden University, Leiden, The Netherlands
[3]Institute of Biology, Leiden University, Leiden, The Netherlands
[4]Netherlands Centre for Biodiversity, Naturalis, Leiden, The Netherlands
[5]Complexity Science and Evolution Unit, Okinawa Institute of Science and Technology Graduate University (OIST), Onna, Japan
[6]Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Koper, Slovenia

Corresponding author: Barbara Boldin, Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Glagoljaška 8, SI-6000 Koper, Slovenia. Email: barbara.boldin@upr.si

Abstract

The time-honored paradigm in the theory of virulence evolution assumes a positive relation between infectivity and harmfulness. However, the etiology of respiratory diseases yields a negative relation, with diseases of the lower respiratory tract being less infective and more harmful. We explore the evolutionary consequences in a simple model incorporating cross-immunity between disease strains that diminishes with their distance in the respiratory tract, assuming that docking rate follows the match between the local mix of cell surface types and the pathogen's surface and cross-immunity the similarity of the pathogens' surfaces. The assumed relation between fitness components causes virulent strains infecting the lower airways to evolve to milder more transmissible variants. Limited cross-immunity, generally, causes a readiness to diversify that increases with host population density. In respiratory diseases that diversity will be highest in the upper respiratory tract. More tentatively, emerging respiratory diseases are likely to start low and virulent, to evolve up, and become milder. Our results extend to a panoply of realistic generalizations of the disease's ecology to including additional epitope axes. These extensions allow us to apply our results quantitatively to elucidate the differences in diversification between rhino- and coronavirus caused common colds.

Keywords: respiratory tract infection, cross-immunity, trade-off, virulence evolution, disease diversity, emerging diseases, COVID-19

Introduction

It has long been acknowledged that pathogen virulence (broadly defined as infection-caused damage to the host) should evolve in response to selection pressures brought about by the disease's ecology. The earlier common wisdom that avirulence should be the inevitable outcome of pathogen evolution gave way in the 1980s to the so-called trade-off hypothesis, that is, the idea that pathogen transmissibility and virulence are coupled, with increased transmission being linked to an increase in virulence (Alizon et al., 2009; Anderson & May, 1982). In the wake of this hypothesis, combined with the assumption (usually left implicit) of full cross-immunity between pathogen strains, theoretical explorations of virulence evolution flourished, giving novel insights into how, for instance, co- and super-infections, multiple transmission modes, host heterogeneity, density-dependent mortality and spatial structure shape pathogen evolution (see Lion & Metz, 2018 for references). Here, we attempt to broaden this tradition by concentrating on alternatives to the two aforementioned assumptions: cross-immunities that increase with the similarity between pathogen strains and virulence that decreases with an increase in transmission, with all these parameters being smooth functions of some underlying quantitative traits.

The specific biological systems motivating our deliberations and used for their illustration are respiratory diseases. For didactical reasons, we first formulate a simple core model, taking the respiratory depth on which a disease strain specializes as the main trait underlying the between-hosts part of a pathogen's life-history and with the similarity of strains decreasing with their distance in the respiratory tract. We explore the evolutionary consequences of the interplay between location-dependent infectiousness, host mortality, recovery rates and limited cross-immunity in two stages: first, a detailed analysis of the monomorphic case up to the potential onset of diversification followed by numerical analyses of the polymorphic dynamics (note that polymorphisms here should be interpreted as long-term polymorphisms, not the evolutionarily ephemeral transient polymorphisms that occur in the course of directional selection). Next, we consider a variety of more complex scenarios in order to establish to which extent the results from the core model extend to realistic generalizations of the disease's ecology and to including additional trait axes co-determining the fitness of pathogen strains. Among other things, we find that, when more trait axes are brought into play, our main conclusions from the core model even go through in the extreme case that cross-immunity fails to wane along the respiratory depth axis. As an

Received April 19, 2022; revisions received July 19, 2023; accepted July 30, 2023
Associate Editor: Ben Ashby; Handling Editor: Tracey Chapman

application, we estimate the life-history parameters for rhino- and corona-virus caused common colds to find first that the observed patterns of diversification in these two groups cannot be understood unless the viruses under consideration have a very different ratio of the tendency to diversify caused by the partial cross-immunity between strains and the counteracting stabilizing tendency coming from the pressure to maximize their capacity for within-host multiplication, and then provide a possible explanation for this difference in terms of the different morphologies of rhino- and coronaviruses. We round off by considering what our approach adds to earlier explorations of the dynamical effects of limited cross-immunity, why the host density (through a term $R_0 - 1$) appears in the threshold criterion for diversification while the standard resource competition models of the ecological literature have a diversification threshold that is independent of resource availability, discuss the pathogen-oriented view on the standard host-centered mathematical approach to disease dynamics, consider how our simplifying assumptions might limit the reach of our conclusions (almost not, provided one reads them appropriately), and discuss how our conclusions accord with what befell to COVID-19 as the latest arrival among our respiratory diseases.

On trade-off assumptions and respiratory diseases

The classical assumption of infectivity increasing with virulence is valid in, for example, leaf diseases (where both quantities increase with infected leaf area) and more generally in host–pathogen systems where virulence and infectivity are both determined solely by per-host pathogen load. For organisms where the disease unfolds mainly internally, the latter assumption appears to have been implicitly based on seeing the host's body as well-mixed (cf. Acevedo et al., 2019). However, many animal bodies are heavily compartmentalized, and mechanistically both virulence and infectivity are to a considerable extent determined by how the pathogens distribute themselves over those compartments (cf. de Jong & Janss, 2002). Respiratory diseases are an example where the classical trade-off assumption is flouted since diseases making their home lower down the respiratory tract are less infectious and more harmful. The latter derives from clinical experience and mechanistic considerations: tissues lower down the respiratory tract are less resilient to damage by the pathogen as well as the elicited immune reaction, while the resulting debris is not cleared away as effectively as higher up and then forms a breeding ground for secondary bacterial pneumonias. Moreover, higher up, like elsewhere in the contact zones of the body with the outside world, the innate immune system is better prepared for action. For corroboration of the inferred overall effect on the level of the host individual see, for example, Erkkola et al. (2020); Feldman et al. (2015); Lee et al. (2012); and Monto and Cavalarro (1971). On the infectivity side, infectious particles produced higher up have a better chance of getting to a next host, and once there face fewer obstacles. For an individual level corroboration of the implied trade-off see, for example, Reperant et al. (2012). Furthermore, we expect the specialization of diseases on different parts of the respiratory tract to be caused by the surface proteins of the pathogen matching a position-dependent mix of receptors, thus potentially diminishing cross-immunity between differently specialized types.

Ecological dynamics

Prologue: the classical SIR model

The starting point for our considerations is the classical SIR model for the dynamics of a single pathogen strain in a homogeneous host population with a constant influx of susceptibles (the latter on the time-honoured assumption that the host population density is mainly determined by other ecological factors; cf. Anderson & May, 1982, text after Equation (12)),

$$\begin{aligned}\frac{dS}{dt} &= B - (\delta + \beta I)S\\ \frac{dI}{dt} &= (\beta S - \delta - \alpha - \gamma)I\\ \frac{dR}{dt} &= \gamma I - \delta R.\end{aligned} \tag{1}$$

In this model, the host population is divided into three classes, namely, susceptible, infected (and infectious), and recovered (and immune), occurring in densities S, I, and R, respectively. The host population has a fixed birth rate B and all newborns are susceptible. Infection occurs via mass-action kinetics with transmission rate constant β. Individual hosts have a baseline mortality rate δ, with infected hosts suffering an additional mortality rate α (in the theoretical literature customarily called virulence). Infected hosts recover at rate γ, upon which they are immune to further infections. Although this model may be considered badly oversimplified, it has the considerable advantage that, with the additional assumption that the immunity extends to all other variants of the disease (an assumption that in the wake of Anderson and May (1982) usually is left implicit), it has only three parameters that may be considered as potentially under evolutionary control. Not only that, any mutants that may arise will invade and take over if and only if they have a larger value of the average lifetime infectivity

$$Q = \frac{\beta}{\delta + \alpha + \gamma}, \tag{2}$$

and evolution thus in the end produces a disease variant that has the maximal feasible value of Q. The latter incidentally corresponds with the largest possible value of

$$R_0 = QN_0, \quad \text{where} \quad N_0 = \frac{B}{\delta}. \tag{3}$$

Here, N_0 is the equilibrium density of an uninfected host population and R_0 is defined, whatever the disease life-history (cf. Appendix D), as the expected number of infections caused by freshly infected individuals present in negligible numbers in an otherwise infection-free community (Diekmann et al., 1990; Lion & Metz, 2018).

Limited cross-immunity between strains

Without the assumption of full cross-immunity, we must account for the possibility that evolution, or invasions by strains coming from different hosts, may cause more than one disease strain to co-circulate (like the plethora of rhino-virus serotypes, see van Regenmortel et al., 2000, pp. 666–667). Therefore, we extend the basic SIR model by assuming an arbitrary number of co-circulating disease strains. We denote the set of currently co-circulating strains as $\mathcal{N} = \{1, \ldots, n\}$. In order to keep the model parameter-sparse and numerically manageable, we furthermore assume, in accordance with Andreasen et al. (1997), that the life cycles of the host and

disease are as simple as possible and that the birth rate into the host population is constant, that immunity only affects the hosts' susceptibilities, is independent of the order of infections and does not wane (but see "More complex ecological scenarios" below). The latter two assumptions together imply that the immune status of a host individual corresponds to the set $P \subseteq \mathcal{N}$ of all strains in the population from which that individual has recovered sometime in the past.

For $P \subseteq \mathcal{N}$, we denote by S_P the density of uninfected hosts with immune status P (for $n = 1$ compartments $S_\varnothing$ and $S_{\mathcal{N}}$ correspond, respectively, to S and R of the standard SIR model). Let $I_{i,P}$ denotes the density of hosts infected by strain $i \in \mathcal{N}$ and with immune status $P \subseteq \mathcal{N}\backslash\{i\}$. We assume that, at any point in time, any infected individual belongs to exactly one of such infection classes, that is, we (1) neglect the possibility of multiply infected hosts (on the supposition that infectious periods are very short relative to the time scale of endemic population turnover[1].) and (2) assume that recovery from a strain confers complete immunity to that particular strain.

With this notation, the force of infection of strain i, as experienced by $\varnothing$- (i.e., never infected) susceptibles, becomes

$$\lambda_i := \beta_i \sum_{P \subseteq \mathcal{N}\backslash\{i\}} I_{i,P}. \tag{4}$$

For uninfected individuals with infection history P, cross-immunity reduces susceptibility to further infections according to the relative infection probabilities

$\theta_{i,P}$ = the probability that a susceptible with immune status P becomes infected following an encounter with an i-infected host, relative to the probability that a $\varnothing$–susceptible would have become infected.

We have $\theta_{i,\varnothing} = 1$. Furthermore, our assumption that recovery from a strain renders the host completely immune to that particular strain requires

1. If $i \in P$ then $\theta_{i,P} = 0$.

Additional natural restrictions on θ are:

2. If $i \in \mathcal{N}\backslash P$ then $0 < \theta_{i,P} < 1$. That is, recovery from strains in P provides some (but never full) immunity to novel strains.
3. $\theta_{i,P\cup\{j\}} \le \theta_{i,P}$. That is, a richer infection history can only enhance immunity.

We can then describe the dynamics of n co-circulating disease strains by way of

$$\begin{aligned}
\frac{dS_\varnothing}{dt} &= B - \Big(\delta + \sum_{i=1}^{n} \lambda_i\Big) S_\varnothing \\
\frac{dS_P}{dt} &= \sum_{i \in P} \gamma_i I_{i,P\backslash\{i\}} - \Big(\delta + \sum_{i=1}^{n} \lambda_i \theta_{i,P}\Big) S_P \quad \text{for} \quad P \subseteq \mathcal{N}, P \neq \varnothing \\
\frac{dI_{i,P}}{dt} &= \lambda_i \theta_{i,P} S_P - \Big(\delta + \alpha_i + \gamma_i\Big) I_{i,P} \qquad \text{for} \quad i \in \mathcal{N}, P \subseteq \mathcal{N} \setminus \{i\}.
\end{aligned} \tag{5}$$

[1] Even when such is not the case, medical experience tells that superinfections with related diseases tend to be rare, probably due to the quick activation of the less-specific innate immune system by the current infection (I. van der Sar, personal communication)

The model represented by Equations (5) and (4) generalizes that in Andreasen et al. (1997) by allowing for strain-dependent recovery rates γ_i and the addition of strain-dependent disease-induced mortalities α_i.

Evolution

Biological assumptions

For evolutionary considerations, we shall distinguish disease strains by the values of a trait vector X or Y in the case of a mutant. We denote by X_i the trait vector of strain i and assume α, β, and γ to be smooth functions of X_i, for example, $\alpha_i = \alpha(X_i)$. We, furthermore, assume that immunity can be represented by a smooth function φ such that

$$\theta_{i,P} = \prod_{j \in P} \big(1 - \varphi(X_i|X_j)\big), \tag{6}$$

$$\text{with} \qquad \varphi(X|X) = 1 \quad \text{and} \quad 0 \le \varphi(Y|X) < 1 \quad \text{for } Y \neq X$$

(for "|" read "given a history of infections including"), which may be interpreted as the result of a freshly transferred inoculant having to pass a number of independent stochastic barriers, set up by each type of infection experienced in the past.

In our concrete application, we shall first assume that the trait is a scalar $x \in [0,1]$ corresponding with depth in the respiratory tract with $x = 0$ at the lower end (Figure 1A); x may be thought of as the mode of the depths over which the pathogen thrives. Here, depth should not be interpreted literally, but seen as a transform of that depth through local elongation or shrinking to make the physiological and chemical conditions everywhere change at roughly the same rate. For the basic demographic parameters of the disease we assume:

- $\alpha(x)$ decreases with x (infections of the lower parts of the respiratory tract cause higher mortality than those of the upper part),
- $\gamma(x)$ increases with x (recovery from lower respiratory tract infections is more difficult)
- $\beta(x)$ increases with x (strains residing in upper parts of the respiratory tract are more easily transmitted), except that near the tip of the nose β has to decrease again as the pathogen there runs out of habitable space. Thus β attains a maximal value close to $x = 1$.
- For each x, $\varphi(y|x)$ decreases with $|y - x|$,

see Figure 1B–D. The decrease of β for x close to 1 has the same effect of guaranteeing the existence of an internal evolutionarily singular point (like an evolutionarily steady strategy (ESS) or a branching point) as the classical trade-off between the disease's demographic parameters, but emerges on a deeper mechanistic level. The final assumption reflects the supposition that the specialization of strains on different parts of the respiratory tract is caused by the surface proteins of the pathogen matching a position dependent mix of receptors, with the difference in surface proteins of differently specialized types diminishing their cross-immunity. However, as we shall discuss in subsection "Additional immune diversity," our main conclusions generally remain valid even when this assumption is relaxed.

Where for the numerical illustrations we have to become more specific, we use

- $\alpha(x) = \alpha_0(1 - \alpha_1 x^k)$ for some positive α_0, α_1, and k,
- $\gamma(x) = \gamma_0 x^l$ for some positive γ_0 and l,

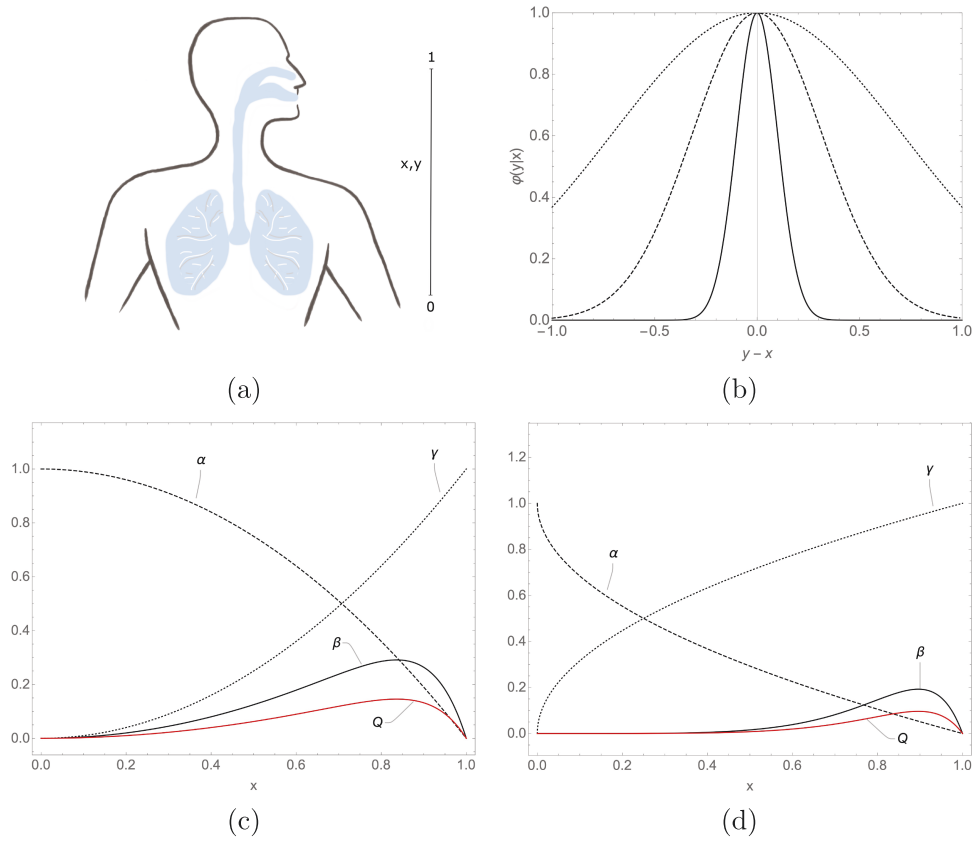

Figure 1. (A) Strains are characterized by a scalar variable x representing infection location, with $x = 0$ and $x = 1$ representing the lower and upper end of the respiratory tract, respectively. Panel (B) depicts the graph of $y \mapsto \varphi(y|x)$ for $c = 2$ (dotted), $c = 10$ (dashed) and $c = 100$ (full). Panels (C) and (D) depict the trade-offs and the expected lifetime infectivity: virulence $\alpha(x)$ (dashed), recovery rate $\gamma(x)$ (dotted), transmission rate $\beta(x)$ (full black), and $Q(x)$ (red) with $\alpha_0 = \alpha_1 = \gamma_0 = \delta = 1$, $\beta_0 = 0.5$ and (C) $k = l = m = 2, n = 12$ and (D) $k = l = 0.5, m = 5, n = 15$.

- $\beta(x) = \beta_0(x^m - x^n)$ for some positive β_0 and $0 < m < n$,
- $\varphi(y|x) = e^{-c/2(y-x)^2}$ for some positive c,

as simple calculation recipes producing shapes in accord with our qualitative assumptions.

Evolution in monomorphic pathogen populations

With only one strain circulating, the system Equation (5) with Equation (4) becomes

$$\begin{aligned}\frac{dS_\varnothing}{dt} &= B - (\delta + \lambda_1)S_\varnothing, \qquad \lambda_1 := \beta_1 I_{1,\varnothing},\\ \frac{dS_{\{1\}}}{dt} &= \gamma_1 I_{1,\varnothing} - \delta S_{\{1\}},\\ \frac{dI_{1,\varnothing}}{dt} &= \lambda_1 S_\varnothing - (\delta + \alpha_1 + \gamma_1)I_{1,\varnothing}.\end{aligned} \tag{7}$$

That is, we get back the SIR model Equation (1) in more elaborate notation (with $S_\varnothing = S$ and $S_{\{1\}} = R$).

Following the adaptive dynamics rulebook (e.g., Brännström et al., 2013; Diekmann, 2004; Geritz et al., 1998; Metz, 2012), we first calculate the equilibria of Equation (7) to determine the invasion fitness of new mutants. If we assume for the moment that the force of infection λ_1 is known, we can express the equilibrium densities as

$$\tilde{S}_\varnothing(\lambda_1) = \frac{B}{\delta + \lambda_1}, \quad \tilde{I}_{1,\varnothing}(\lambda_1) = \frac{\lambda_1 \tilde{S}_\varnothing(\lambda_1)}{\delta + \alpha_1 + \gamma_1}, \quad \tilde{S}_{\{1\}}(\lambda_1) = \frac{\gamma_1 \tilde{I}_{1,\varnothing}(\lambda_1)}{\delta}. \tag{8}$$

Substituting the expression for $\tilde{I}_{1,\varnothing}(\lambda_1)$ in the definition of λ_1 in Equation (7) gives a scalar equation for $\hat{\lambda}_1$

$$\hat{\lambda}_1 = \frac{\beta_1 \hat{\lambda}_1 B}{(\delta + \alpha_1 + \gamma_1)(\delta + \hat{\lambda}_1)},$$

and so

$$\hat{\lambda}_1 = 0 \quad \text{or} \quad \hat{\lambda}_1 = \frac{\beta_1 B}{\delta + \alpha_1 + \gamma_1} - \delta = \delta(R_{0,1} - 1). \tag{9}$$

In the latter case,

$$\begin{aligned}\hat{S}_\varnothing &= \frac{N_0}{R_{0,1}} = Q_1^{-1}, \quad \hat{I}_{1,\varnothing} = \frac{\delta(R_{0,1} - 1)Q_1^{-1}}{\delta + \alpha_1 + \gamma_1} = \frac{\delta(R_{0,1} - 1)}{\beta_1},\\ \hat{S}_{\{1\}} &= \frac{\gamma_1(R_{0,1} - 1)}{\beta_1}.\end{aligned} \tag{10}$$

For evolutionary considerations, it is crucial that the equilibrium values of the epidemiological state variables depend on the traits that are currently present. We shall, therefore, make that dependence explicit in the following part of the argument. Moreover, since in the case of only one circulating strain, the index gives no additional information, we shall in the remainder of this subsection write X instead of X_1.

From the model specification, or with rather more effort from Equations (4) and (5), it can be seen that the initial

growth rate, in this context called invasion fitness, of a new mutant Y given an X resident environment, is given by

$$s(Y|X) = \beta(Y)\Big(\hat{S}_{\varnothing}(X)+\big(1-\varphi(Y|X)\big)\hat{S}_{\{1\}}(X)\Big)-\big(\delta+\alpha(Y)+\gamma(Y)\big). \tag{11}$$

To calculate internal evolutionarily singular strategies, X^*, we have to set the derivative of s for Y at $Y = X$ (customarily called selection gradient, a measure for the local strength of directional selection) equal to 0.

$$\begin{aligned}\frac{\partial s(Y|X)}{\partial Y} &= \beta'(Y)\Big(\hat{S}_{\varnothing}(X) + \big(1-\varphi(Y|X)\big)\hat{S}_{\{1\}}(X)\Big)\\ &\quad -\beta(Y)\varphi_1'(Y|X)\hat{S}_{\{1\}}(X)-\alpha'(Y)-\gamma'(Y)\end{aligned} \tag{12}$$

where $\beta' := d\beta/dY = \big(\partial\beta/\partial y_1,\dots,\partial\beta/\partial y_k\big)$, and so on and $\varphi_1'(Y|X) := \partial\varphi(Y|X)/\partial Y$. On setting $Y = X$, the selection gradient can be simplified by using that $\varphi(X|X) = 1$ and $\varphi_1'(X|X) = 0$ (this follows from properties 1 and 2),

$$G(X) := \left.\frac{\partial s(Y|X)}{\partial Y}\right|_{Y=X} = \beta'(X)\hat{S}_{\varnothing}(X)-\alpha'(X)-\gamma'(X). \tag{13}$$

Now we have only one argument left we can hide it to remove clutter. By rewriting G as

$$G = (\delta+\alpha+\gamma)\Big(\frac{\beta'}{\beta}-\frac{\alpha'+\gamma'}{\delta+\alpha+\gamma}\Big) \tag{14}$$

we get

$$G = (\delta+\alpha+\gamma)\big(\ln(Q)\big)', \tag{15}$$

with Q the average lifetime infectivity (of the resident strain X) as in Equation (2). The biological conclusion is that for monomorphic gradual evolution X converges to a maximizer X^* of Q, just as in the classical SIR model. However, X^* need no longer be a local maximizer of the invasion fitness, since the latter now is co-determined by the cross-immunities. Hence, near X^* invasion need not imply substitution, which opens the possibility for the evolutionary trajectory to become polymorphic.

Here, we only derive the branching criterion for scalar traits, deferring the higher-dimensional trait spaces of the Section "Additional immune diversity" to Appendix A. A convergence stable evolutionarily singular strategy x^* is an ESS when $\partial^2 s(y|x)/\partial y^2|_{y=x=x^*} < 0$ (i.e., x^* maximizes $s(\cdot|x)$ and, therefore, locally selection is stabilizing) and a branching point when $\partial^2 s(y|x)/\partial y^2|_{y=x=x^*} > 0$ (i.e., x^* minimizes $s(\cdot|x)$ and, therefore, locally selection is disruptive) (Geritz et al., 1998; Metz et al., 1996). Using the same conditions on φ as before, gives,

$$\text{at } x = x^*: \quad \left.\frac{\partial^2 s(y|x)}{\partial y^2}\right|_{y=x} = \beta''\hat{S}_{\varnothing}-\alpha''-\gamma''-\beta\varphi_{11}''\hat{S}_{\{1\}}. \tag{16}$$

Hence, x^* is a branching point when (with all parameter functions evaluated at x^*)

$$\beta''\hat{S}_{\varnothing}-\alpha''-\gamma'' > \beta\varphi_{11}''\hat{S}_{\{1\}}, \tag{17}$$

which can be rewritten as

$$(\delta+\alpha+\gamma)\big(\ln(Q)\big)'' > \gamma(QN_0-1)\varphi_{11}'', \tag{18}$$

or, expressed in dimensionless parameter groups,

$$\frac{\gamma}{\delta+\alpha+\gamma}(QN_0-1)\frac{\varphi_{11}''}{[\ln(Q)]''} > 1. \tag{19}$$

(see Supplementary Material SI1). The advantage of inequality Equation (18) is that it directly extends (as shown in Appendix A) to the case with additional epitope axes, to be discussed later on. Inequality Equation (19), on the other hand, has the advantage that its terms allow a biological interpretation, which may help to guesstimate their values in concrete instances: the first of the three terms on its left is the host's probability of surviving an infection; the second term equals the maximum feasible value of $R_0 - 1$; the third term is the strength of disruptive selection caused by limited cross-immunity relative to that of the stabilizing selection coming from how effective the disease is in reaching new host individuals, all three at the maximizer of R_0.

The overall conclusion is that the currently dominant respiratory disease strain will keep increasing its R_0 on the evolutionary time scale, thereby evolving into strains residing higher and higher up in the respiratory tract until the strains start running out of habitable space close to the tip of the nose (neglecting the unlikely case of a disease that starts its career even closer to the tip and increases its habitable space till it likewise gets close to the maximal R_0). Any selectively driven diversification will only start in the neighborhood of the highest feasible value of R_0, and is facilitated by larger host densities. The less extensive the cross-immunity, the larger $|\varphi_{11}''|$ and the stronger this effect. We shall further on demonstrate the first, better empirically accessible, of these two effects in a series of numerical examples.

Polymorphic populations

For polymorphic populations, we no longer have explicit expressions for endemic equilibria. So we have to use numerical methods. The straightforward approach relies on running the system in Equation (5) for sufficiently long so as to guarantee satisfactory convergence to the resident equilibrium and then use the equilibrium values to derive a mutant's fitness. However, with increasing number of strains, the system size increases rapidly. Indeed, for n co-circulating strains, the system in Equation (5) has a whopping $2^{n-1}(n+2)$ equations! We thus propose an alternative method that relies on rewriting the $2^{n-1}(n+2)$ equilibrium equations as a fixed point problem with n equations for the equilibrium forces of infection. We argue that the obtained map (which we name the force-of-infection map) yields an evolutionarily equivalent system thus providing us with a more manageable alternative to studying polymorphic evolution. Furthermore, such a reformulation allows us to highlight deeper connections with ecological competition models. But first, we present the results of numerical investigations and their biological implications.

Numerical examples

Numerical examples depicted in Figure 2 show how the effect of N_0 on disease diversification extends to the polymorphic realm. The depicted evolutionary trajectories come from simulations of the evolutionary dynamics using the procedure described in Appendix B, starting with a single strain infecting the lower end of the respiratory tract. From there the strain evolves toward the depth x^* maximizing the life-time infectivity Q. At low host density (Panel A) x^* is an ESS and, therefore, the evolutionary end point. In accordance with Equation (18), at higher host densities (Panel B) x^* becomes a branching point, leading to dimorphic pathogen populations. Panel C, moreover, shows that still larger host densities lead to

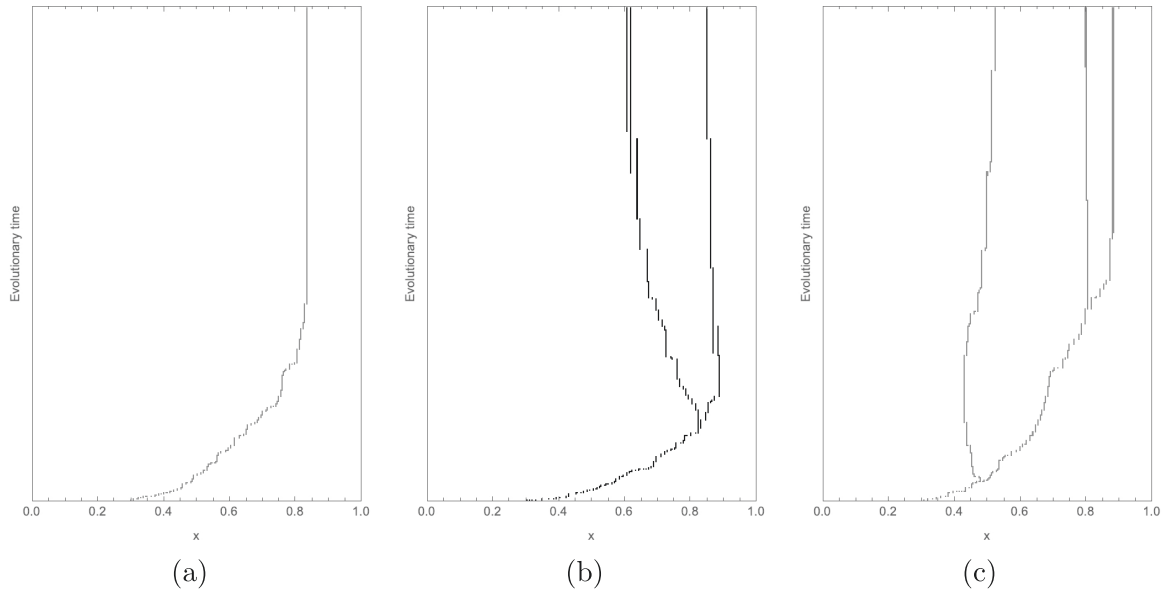

Figure 2. Simulated evolutionary trees for $N_0 = 10^2$ (A), $N_0 = 10^3$ (B), and $N_0 = 10^5$ (C). The parameters are $\alpha_0 = \alpha_1 = \gamma_0 = \delta = 1$, $\beta_0 = 0.5$, $c = 2$, $k = 2, l = 2, m = 2, n = 12$.

further branching and correspondingly richer strain diversity, harbored by the upper parts of the respiratory tract.

Figure 3 graphically represents the selection patterns underlying these phenomena. Panels A–C (called Pairwise Invasability Plots or PIPs in adaptive dynamics lingo) show how selection on mutants (on the vertical axis) changes with the resident trait (horizontal axis) by showing the sign of a mutant's fitness in dependence on its own trait and that of the resident. The evolutionarily singular trait value x^*, indicated with a black dot, occurs where the second fitness zero contour, separating the positive and negative signs, crosses the diagonal (which itself is also a zero contour). Going from Panel A–C, it can be seen how increasing the host density makes selection at the singular trait value turn disruptive (in Panel A, the fitness just above and below the dot is negative and in panels B and C, it is positive). Panels D–F (called TEPs by adaptive dynamics adepts) depict in the form of arrows the fitnesses of mutants with small effect for the case where there are two different residents. An upward arrow means that an upward mutant of the resident depicted on the vertical axis has positive fitness, while a downward arrow indicates the same for downward mutants. A right-pointing arrow means that an upward mutant of the resident depicted on the horizontal axis has positive fitness, whereas a left-pointing arrow indicates the same for downward mutants. The colored curves are so-called evolutionary isoclines, separating the regions with arrows in the different slant classes, while the white region characterizes pairs of strains that can coexist since, according to the PIPs in Panels A–C, each can invade into the other.

Up to tri- and tetra-morphisms, the patterns seen in Figure 2 can be inferred from the TEPs. To this end, consider a thought experiment in which one of the two lines of descent is forbidden to mutate, making its trait value, say x_1, a fixed parameter. This way the isocline for the other trait x_2 in the TEP gets an alternative interpretation as the (set of) loci where for fixed x_1, the evolution of x_2 would change direction, that is (the locus of), the singular points of those (monomorphic) evolutions in dependence on the value of the parameter x_1. These singular points may or may not be branching prone. In the former case, the local piece of isocline around that point connects to a region of trimorphisms in trait space tripled. A positive branching proneness is indicated in Figure 3 by representing the corresponding pieces of the isoclines interrupted. When a mutation lets the evolutionary trajectory jump into that region of trimorphisms, it may not stay there for long as evolution of the other, unbranched, line of descent may quickly lead the evolutionary trajectory again out of the trimorphisms. However, close to where the isoclines intersect, the movement of that other line slows down so far that branching really may take off, thus linking the occurrence of additional branching to places where a branching prone isocline intersects the other isocline. If at their intersection both isoclines are branching prone, the corresponding lines of descent usually branch shortly after each other. Figure 3D–F shows that in our particular case, in addition, evolution starting from a monomorphism has almost no chance to get close to a branching prone part of the isoclines except close to their intersection.

The force-of-infection map and its connection to ecological competition models

Here, we briefly describe how the $2^{n-1}(n+2)$ equilibrium equations of (5) can be rewritten as a fixed point problem with n equations for the equilibrium forces of infection and argue that the obtained force-of-infection map yields an evolutionarily equivalent system. We only give a heuristically argued description of the method, while deferring rigorous proofs to a companion paper geared to a mathematical audience Boldin & Metz (manuscript in preparation).

We use the strategy for calculating the equilibria exemplified in "Evolution in monomorphic pathogen populations" to reduce the $2^{n-1}(n+2)$ equilibrium equations of (5) to n equations for the equilibrium forces of infection $\hat{\lambda}_i$. Figure 4 shows for $n = 2$, the order in which for given (λ_1, λ_2), one

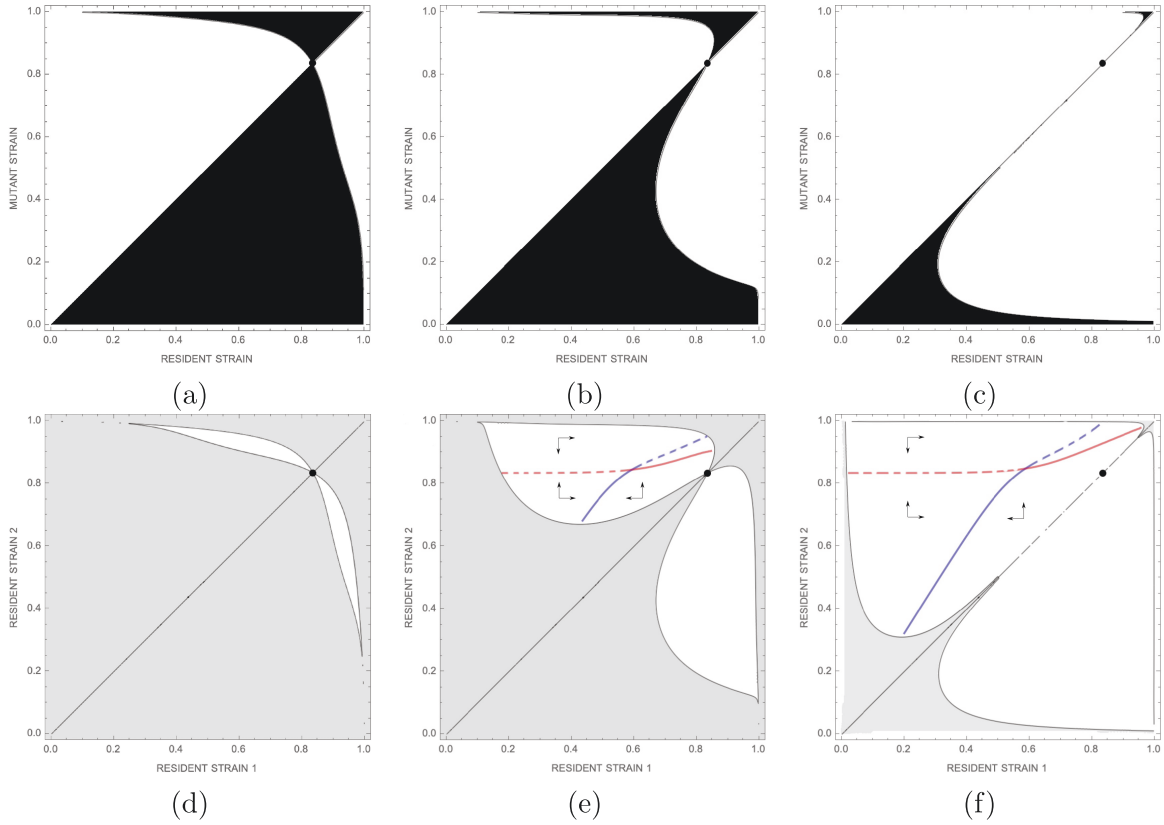

Figure 3. Pairwise invasibility plots and trait evolution plots (TEPs) for $N_0 = 10^2$ (A,D), $N_0 = 10^3$ (B,E), and $N_0 = 10^5$ (C,F). In the pairwise invasibility plots (PIPs) in Panels A, B, and C, the white and black represent the regions where the invasion fitness is positive and negative, respectively. The black dot on the diagonal of the PIP depicts the monomorphic singularity. In the TEPs in Panels D, E, and F, the white region corresponds to the area where the two residents can mutually invade each other. The blue and the red line depict the 1-, 2- isocline, respectively (parts of the isoclines where further branching is imminent are depicted with a dashed curve). The parameters are $\alpha_0 = \alpha_1 = \gamma_0 = \delta = 1$, $\beta_0 = 0.5$, $c = 2$, $k = 2, l = 2, m = 2, n = 12$.

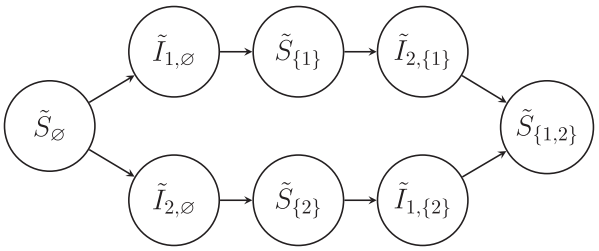

Figure 4. The scheme for calculating the epidemic equilibrium for given forces of infection in the case of two co-circulating strains.

can successively calculate the equilibrium values $\tilde{I}_{i,P}, \tilde{S}_P$ of Equation (5) in an easy manner. The resulting expressions are linear in N_0 due to the fact that prescribing the λ_i makes Equation (5) linear. Let $\Lambda = (\lambda_1, \ldots, \lambda_n)^\mathsf{T}$ be the vector of forces of infection. Then, in the last step, where we substitute the expressions for the $\tilde{I}_{i,P}$ in the definition of λ_i $(i = 1, \ldots, n)$, we end up with an equation of the form

$$\hat{\Lambda} = F(\hat{\Lambda}). \tag{20}$$

Hence, $\hat{\Lambda}$ is a fixed point of F (which we call the force-of infection map). Supplementary Material SI2 shows F for the case $n = 2$. It is also possible to find an explicit expression for general n. However, for numerical work, it is just as practical to write a recursive procedure for calculating the value of $F(\Lambda)$ for given Λ.

A simple means for solving Equation (20) derives from the observation that it also is the equilibrium equation of the recurrence

$$\Lambda_{t+1} = F(\Lambda_t), \tag{21}$$

so that we could calculate $\hat{\Lambda}$ as $\Lambda_\infty = \lim_{t\to\infty} \Lambda_t$, where in the numerical procedure we replace Λ_∞ by Λ_t for some suitably large t. A closer look at F, moreover, shows that this recurrence allows an alternative interpretation as the representation of an ecological competition model (in the phenomenological sense that for each species increasing its density leads to a decrease of the per capita growth ratios of the other species). A direct calculation shows that this competition model mimics its parent epidemic model in that, as long as both have only point attractors, the signs of its invasion fitness in dependence on $(Y|\{X_1, \ldots, X_n\})$ match those of its

parent (where the invasion fitness for an unstructured discrete-time model is defined as the logarithm of the invasion reproduction ratio), corroborating the intuition that the evolutionary diversification resulting from limited cross-immunity is caused by apparent competition (Holt & Bonsall, 2017).

So far we only for the cases $n = 1, 2$ have a rigorous proof that the interior of the state space of Equation (21) has a unique attractor, which is either an interior or a boundary equilibrium. However, our extensive numerical work for $n > 2$ has not brought to light any cases of untoward behavior of either the original model or the recurrence. (Neither can the examples of oscillatory behavior described in Andreasen et al. (1997) and Lin et al. (1999) occur for the trait dependence assumed in our models.) The upshot is that, in the case that both models have only point equilibria for their community dynamical attractors, the evolutionarily steady coalitions of both models coincide, as do their domains of evolutionary attraction (if we specify the stochastic structure of the individual-based model associated with the recurrence Equation (21) by assuming the individual offspring distributions to be geometric, to accord with those of the parent model, as described in Appendix B). We express these relations by saying that the two models are evolutionarily equivalent.

More complex ecological scenarios

For didactical purposes, we started with the simplest possible model describing the dynamics of multiple strains with limited cross-immunity. Here, we discuss how the conclusions of previous section carry over to various generalizations.

Additional immune diversity

Of course, the assumption that the degree of cross-immunity between strains depends only on the difference in their preferred depth in the respiratory tract is a gross oversimplification. In this section, we explain why we expect that the results of the previous sections will only be strengthened when there are additional factors co-determining cross-immunity. For the sake of the argument, we shall represent these factors by just a single scalar that we draw orthogonal to the depth axis. As it seems reasonable to assume that the additional factors are not there just for their influence on the immune reaction, but arise as a side effect of other physiological functions and thus have also other, more straightforward, fitness effects, we scale that axis such that strains with the highest fitness are located centrally. (This in order to make the respiratory-depth axis attract evolutionarily while making the fitness landscape locally flat in directions away from this axis, thus creating a perspective that puts the trait specific for respiratory diseases center stage in our mathematics.) Again for the sake of the argument, we assume that these other factors and depth interact roughly multiplicatively in determining Q and thereby R_0. (This in accordance with our intuition about the qualitative picture of the variation of R_0 shown in Figure 5; see Supplementary Material SI3.)

From a fuller perspective, the single variable "respiratory depth" should be seen as coming from a restriction of our attention to the principal dimension in a multivariate numerical representation of a space of characteristics of the disease agent that can act as predictors for its ecological roles. The present section should thus be seen as restoring the remaining dimensions.

The conclusions that can be drawn from that fuller picture are graphically represented in Figure 5. Since the numerical examples tell that the increase of the readiness to diversify with R_0 also extends to the polymorphic case, we feel justified to draw in the hypothetical results of this diversifying process in the form of the locations of strains present after a fair period of evolution. The experience with other multivariate competition models (e.g., Zhang et al., 2023) is that even though the eventual evolutionary end result might be more regular, simulations tend to lead to a lot of turmoil in which a larger number of roughly competitively matched strains keep dancing around, slowly filling up the accessible areas of the trait space in patterns suggestive of how one expects a

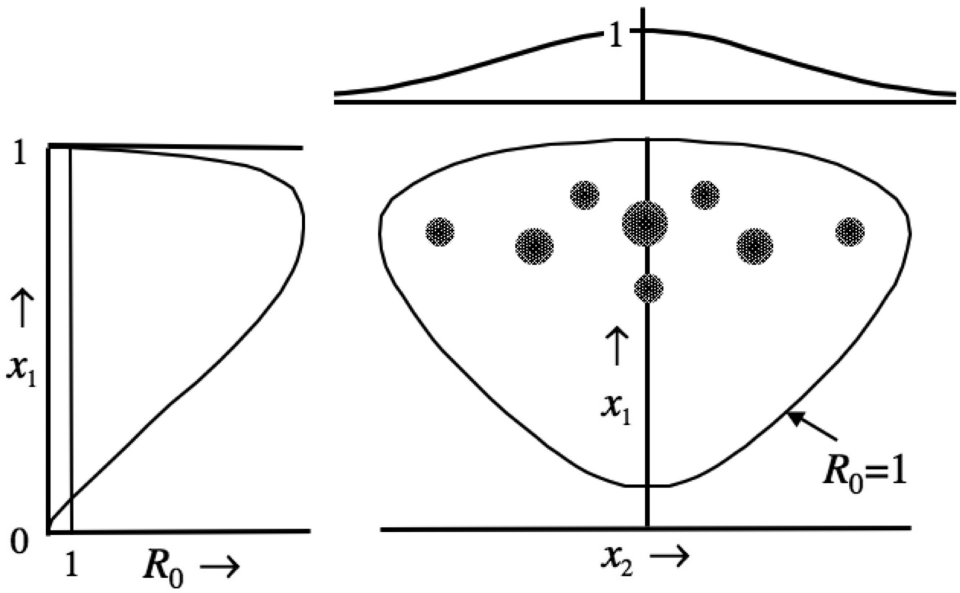

Figure 5. Stylized representation of how the interplay between respiratory tract position, x_1, and additional factors, x_2, is expected to determine how strain diversity depends on the respiratory tract position. Width of the discs symbolizes the relative population sizes of the strains.

potential evolutionary end result to look, but far more irregular. This is because in higher dimensions, the greater freedom of movement makes the effects of stochasticity of the mutational process more pronounced, while the remnants thereof disappear with the contraction of the probability cloud during the approach to an attracting Evolutionarily Steady Coalition. Figure 5 shows how we envision such an end result. Figure 5 should make clear that, in line with the extension of the diversification criterion Equation (18) to higher-dimensional trait spaces derived in Appendix A, the diversification along any additional axes can be expected to be higher near the nose than in the recesses of the lungs.

The above arguments also allow dropping the assumption that cross-immunity decreases with distance along the respiratory depth axis, and still to get all the qualitative predictions put forward till now (for a proof see the last part of Appendix A).

Waning immunity and other generalizations

The assumption that the disease's influence on the host population dynamics can be neglected, makes that our conclusions qualitatively extend to cases with density dependent encounter rates: If we replace β with $\beta g(N)$, where N denotes the total host population density, $g(0) = 1$ and $g(N)$ decreases sufficiently slowly that $Ng(N)$ increases, then our epidemic differential equations only change by having everywhere β replaced by $\beta g(N)$, for which we, by assumption, may substitute $\beta g(N_0)$. Hence, R_0 becomes $Qg(N_0)N_0$, instead of just QN_0, with similar changes ($N_0 \to g(N_0)N_0$) occurring everywhere else (see Supplementary Material SI4). So the readiness to diversify is still modulated by $R_0 - 1$, but with R_0 proportional to $N_0 g(N_0)$. (Note that our assumptions about g rule out the trendy extreme assumption of frequency-dependent transmission, $g(N) = N^{-1}$, which we do not see as a great loss since we cannot believe effective encounter rates to be the same in the remote countryside, small villages, or densely built up cities (with comparable scenarios for other host species).)

In our model formulation, we have so far assumed that cross-immunity affects only susceptibility. In models with cross-immunity affecting both susceptibility and infectivity, the two occur as a product in the expressions for the monomorphic invasion fitnesses (Supplementary Material SI5). Hence, for such models, all our evolutionary conclusions about monomorphic disease evolution, including those about the initiation of diversification, stay intact. Unfortunately, this quantitative evolutionary equivalence of the two model classes does not extend to the polymorphic realm. However, we have no reason to suppose that the two model classes may show any qualitatively different behavior either. In fact, we show in Supplementary Material SI5 that in diseases that last short relative to the lifetime of their hosts, the change in the fitness landscapes relative to our core model amounts to adding or subtracting a small amount of the order of the ratio of the disease duration and the host lifetime, $\delta/(\delta + \alpha + \gamma)$.

On instigation of the reviewers we, in Supplementary Material SI6, analyze the evolutionary effects of an effect of immunity on virulence. We only considered the monomorphic case. Notwithstanding the heavier mathematics needed to reach them, the results are surprisingly simple. The only effect is that the criterion for diversification has to be adapted in that the factor $\delta + \alpha + \gamma$, that is, the inverse of the average disease duration at the evolutionarily singular strategy has to be replaced by the inverse of the average disease duration in immune hosts (infected by a mutant close to the ESS, so that immune residents still have a small chance to get infected). (Note that, as phrased, this result also holds true for our core model, except that the average disease durations of mutant-infected immune and nonimmune hosts are equal, so that mathematically we may express the result in terms of either.)

We also have assumed the simplest possible life cycle of the disease, that is, with only one, Markovian, disease state, infected as well as infective. However, since we only deal with population equilibria, our main results extend seamlessly to more general scenarios where the within-individual disease progression is independent of the population state (c.f. Appendix C; see e.g., Diekmann et al., 2003; Heesterbeek & Metz, 1993; Metz & de Kovel, 2013.) This point is especially relevant in applications where we should estimate the relevant ecological parameters. Moreover, by the results in Metz (1978), this result even extends to structured host populations, provided all host classes experience the same forces of infection λ_i (so that the epidemic does not change the composition of the susceptible population) while Trapman et al. (2016) show that the effect of fine scale differences in contact structure is generally small.

The extensions discussed above are those for which we were able to devise hard proofs. The fact that our results extend over so many different variations of the basic model structure suggests that those results also extend, at least approximately, to even more variations on the deliberately simplified scenario that we focused on for a start. However, there are also exceptions, the main one being when immunity wanes rather than keeps up for ever. This adds additional terms to the epidemiological differential equations (5) to account for the transfer of hosts from the different immune classes to classes with one of the elements of P removed (cf. Appendix C and Supplementary Material SI7). This so far precluded proving that F in the analog of Equation (21) can be reinterpreted as representing a discrete-time competition model. However, the invasion fitness for the monomorphic case can still be explicitly expressed, leading to

$$\frac{\delta/\varepsilon}{\delta/\varepsilon + (\delta+\alpha)/(\delta+\alpha+\gamma)} \frac{\gamma}{\delta+\alpha+\gamma} (R_0 - 1) \frac{\varphi''}{\left(\ln(Q)\right)''} > 1, \tag{22}$$

with ε the rate of immunity loss, as criterion for the occurrence of selectively driven diversification. Here, δ/ε equals the average immune period relative to the average lifetime of the host, and $(\delta + \alpha)/(\delta + \alpha + \gamma)$ the probability of dying when ill.

Application to rhino- and coronaviruses

Although the preceding formulas had the form of quantitative expressions, we used them only for conceptual elucidation and reaching qualitative conclusions. With Equation (22) in place, we can explore the possibilities for becoming more quantitative. To this end, we focus on the common colds caused by rhinoviruses, mainly since, although we have not been able to find tailored parameter estimates (Appendix D discusses tools and Appendix E data), we can combine what estimates there are with common experience to arrive at fair guesstimates of at least their order of magnitude. The 209 tentative species of rhinovirus mentioned in van Regenmortel et al. (2000) no doubt also owe their existence to a worldwide spatial structure, with only the coexistence on a more local scale governed by the mechanism put forward above. Yet, on local scales, there

definitely is a good amount of diversity as well: Monto et al. (1987) found over a five-year period in Tecumseh, Michigan, 52 of the then known 89 rhinovirus serotypes. The lore of the 50s was that colds either conferred no immunity or it waned very fast, such in contradiction to the observation that while snotty noses were ubiquitous in Kindergarten their incidence became considerably less toward the end of secondary school (for modern numbers see Heikkinen & Järvinen, 2003). The explanation, already proposed by Andrewes (1950), is that there exists a large number of co-circulating viruses all conferring at least some longer lasting immunity, but little or no cross-immunity. The first and only phenomenological level measurements of immunity conferred by rhinoviruses were done by Jackson and co-workers, in particular, Jackson and Dowling (1959) and Jackson et al. (1962), who found no cross-immunity between the tested strains, and an immunity that rose over the first year after recovery and then started to decrease again. Finally, although it takes two or three days before the symptoms start, virus shedding starts already after one day, peaks on the second and third day, to end one to two weeks later together with the ceasing of the symptoms (Lorber, 1996; Thompson et al., 2013). In Appendix E, we end up with the estimate for the equivalent average duration of the immune period $\varepsilon^{-1} \approx 3$ years. This makes the first two terms on the left of Inequality Equation (22) both approximately equal to one (since the chance of not surviving a common cold, $(\delta+\alpha)/(\delta+\alpha+\gamma)$, is far smaller than $\delta/\varepsilon \approx 80^{-1}/3^{-1} \approx 0.04$). Next, consider the one but last term in Equation (22). Given the large number of coexisting rhinoviruses each of them need not have a large prevalence. Starting from a rough lifetime average of 4 colds per person per year, of which 40% caused by rhinoviruses and some 50 circulating strains, the use of Formula (E1) got us $R_0 \approx 1.1$, which in view of the humongous observed diversity requires the last term to be a good deal larger than one. That is, the disruptive selection caused by the downward curvature φ'' of the resident-elicited immunity to close by mutants should far outstrip the stabilizing selection coming from the downward curvature $\big(\ln(\mathcal{Q})\big)''$ at the top of $\mathcal{Q}$. (For more detail see Supplementary Material SI9.) In the following paragraphs, we argue that it follows from all data together that the term $\varphi''/\big(\ln(\mathcal{Q})\big)''$ should be very different in rhino and corona viruses and give an explanation for that difference in the terms of their morphology.

It should be noted here that that last number should not be interpreted as referring to the changes of immunity and lifetime infectivity in the direction of the respiratory axis. By the results in Appendix A, for higher-dimensional epitope spaces, the direction with the largest value of $\varphi''/\big(\ln(\mathcal{Q})\big)''$ is the one that counts. This suggests that $|\varphi''| \gg |\big(\ln(\mathcal{Q})\big)''|$ along at least one, and probably more directions in the space of additional epitope axes. The inequality should then be interpreted as the change of the cross-immunity away from 1 being far larger than the detrimental effect that this change may have on the virus' ability to spread due to its lesser reproductive efficiency within its hosts, an observation corroborated by the seeming lack of cross-immunity of the various rhinovirus strains observed by Jackson and Dowling (1959).

For coronavirus-caused common colds the data are far worse, but if we try to do the same exercise, all parameters come out in the same order of magnitude except for R_0 where, with only four co-occurring species (recently increased to five, see, e.g., Cui et al. (2002)),we find $R_0 \gtrsim 2.5$. (The reason for the $\gtrsim$ instead of an $\approx$ is that the partial cross-immunity between the serotypes decreases the frequency of corona caused colds below the value for independently infecting ones.) So given the fact that coronaviruses so far show little diversification, and that the new SARS-CoV-2 has dramatically shown that this cannot be due to their slow speed of evolution, they should have a dramatically lower value of $\varphi''/\big(\ln(\mathcal{Q})\big)''$ than we surmised for rhinoviruses. The molecular data even point at separate animal origins of the corona-caused common colds (Forni et al., 2017), implying that there has been no cross-immunity driven diversification at all.[2] This would agree with the anecdotal evidence that the various coronaviruses afflicting humans might still have non-negligible cross-immunity.

The question remains how this difference in diversification potential between the two kinds of viruses relates to their different operations as a consequence of their different structures: where corona viruses do their initial interfacing with the within-host environment through their spike proteins, anchored in a lipid bilayer envelope encasing their protein capsid, the "naked" rhinoviruses do this through their capsid-proteins VP1, VP2, and VP3, that cause their antigenic diversity, while a canyon in VP1 attaches to the cell surface receptors of their hosts (Jacobs et al., 2013). The latter suggests a mutational decoupling of the docking abilities and the interactions with the hosts' adaptive immunity, causing many directions in epitope space where $|\big(\ln(\mathcal{Q})\big)''|$ is very small while $|\varphi''|$ is not, making $\varphi''/\big(\ln(\mathcal{Q})\big)''$ large. (For a more elaborated version of this argument see Supplementary Material SI9.)

Discussion

Connections to older models

We begin the discussion by briefly pointing to earlier modeling efforts on cross-immunity between co-circulating disease strains, only mentioning the references that started up a particular research tradition. Previous modeling efforts have investigated the population dynamical patterns engendered by relatively few strains (typically up to four) (Andreasen et al., 1997; Bhattacharyya et al., 2015; Lin et al., 1999) considered a single evolutionary step (Restif & Grenfell, 2006), or studied long-term evolutionary dynamics assuming a discrete linear strain space and the simpler case of polarized immunity in which partial cross-immunity renders some of the hosts totally immune (Gog & Grenfell, 2002) (the practical advantage of the latter assumption being that the dimensionality of the multi-strain model increases linearly with the number of circulating strains, rather than exponentially, but at the cost of only allowing directional immune escape without diversification). Here, we study immune-driven pathogen evolution in the context of a more general multi-strain model with partial cross-immunity coming from individual hosts' infection histories, while linking the similarity-induced strength of cross-immunity and the strain-specific epidemiological parameters by means of underlying continuous quantitative traits. The latter assumption is crucial as it gives access to the geometric structures of the fitness landscapes that underlie our general

[2] Although ultimately correct, there is still a tiny technical gap in this argument which is plugged in Supplementary Material SI8.

conclusions (albeit only technically in that those conclusions should extend to cases where a discrete epitope space can be embedded in $\mathbb{R}^n$ for some n in a manner that accords with the other assumptions of the model). The only other paper so far following this strategy is Best and Hoyle (2013), with as differences that we base our assumptions in concrete biology, and for this reason also look into the effect of higher-dimensional epitope spaces, and that our assumption that the influence of the disease on the host birth rate is so small that it for all practical purposes can be neglected led us to more and stronger biological conclusions. (Note that while a few percent additional mortality caused by a novel human disease is perceived as unacceptably large, this has a non-negligible influence on evolutionary predictions only in the special case that its influence on Q differs manifestly among disease strains.) Like ours, all these models stayed on the phenotypic level. A model exploring the underlying genetic level is Koelle et al. (2006). There, evolution goes in fits and starts, where the fits correspond to the oozing out in a subset of genetic states that are phenotypically roughly equivalent as measured by their ecological effects, and the starts to what we in our phenotypically oriented language call substitution events.

Where does the factor R_0–1 come from?

The first two conclusions from our model are that highly virulent strains infecting the lower airways are bound to evolve toward the nose thereby becoming more benign, and that the upper airways are expected to be infected by a richer diversity of pathogen strains than the lower parts of the respiratory tract. This first conclusion is a result of the assumed negative, instead of the usually assumed positive, relation between infectivity and virulence, coming from mechanistic considerations and in accord with clinical experience. The second conclusion comes from the similarity of the effect of limited cross-immunity to ecological competition. We fleshed out this intuition by establishing the evolutionary equivalence of our epidemic model and a discrete-time competition model. In this model, the usual branching criterion in the form of an inequality in terms of the curvature of the carrying capacity at its maximum and the curvature of the competition effect for near-equal competitors turns out to be modulated by $R_0 - 1$ leading to a dependence of the condition for first branching on N_0, with branching becoming easier for higher N_0. The natural conjecture that this effect extends to any subsequent branching as well was corroborated by our numerical studies dealing with the further diversification in preferred respiratory depth. Finally, the extension of the monomorphic branching condition to higher-dimensional trait spaces worked out in Appendix A indicates that branching in other directions in epitope space, not directly related to respiratory depth, will be similarly promoted and will thus also be more common close to the nose where R_0 is highest. Mechanistically, readiness to diversify can be regarded as selection for escaping from mutual cross-immunity, with the factor $R_0 - 1$ deriving from the fact that in the endemic situation a higher $R_0 - 1$ entails a higher density of partially immune individuals, and therewith an increasing advantage for mutants able to infect them (as seen by tracing the term $QN_0 - 1$ in Equation (18) back to $\hat{S}_{\{1\}}$ in Equation (17) and Equation (10)). Seen from a resource competition perspective, the two pathogen strains can mutually invade and undergo disruptive selection since in addition to a common resource they exploit alternative resources, and do so better when they differ more, in the form of immune individuals produced by the other strain at a rate proportional to its density, with the latter proportional to its $R_0 - 1$. For the initiation of that branching, we focus on the close neighborhood of the prospective branching point, so that to first order of approximation these $R_0 - 1$s equal the $R_0 - 1$ at that point.

Continuing in this vein, we speculate that an extension of the principle that a higher host density leads to higher disease diversity may also be behind the shift from one to two co-circulating influenza strains from 1997 onwards (see Earn et al., 2002).

What about within-host evolution?

In the wake of Anderson and May (1982), we have followed the by now established tradition that treats disease evolution fully on the level of diseased hosts. In reality, we, of course, are dealing with the evolution of a structured meta-population of pathogen individuals. This meta-population differs from the usual structured meta-populations in ecology in that we disregard the possibility of secondary immigrants into a patch (cf. Metz & Gyllenberg, 2001) on the supposition that these are both rare and their establishment is prevented by the action of the innate immune system elicited by the initial infection. This is also the context in which we should view the recent attention in the media to within-host evolution of SARS-CoV-2, in particular, in immunocompromised patients: What in our between host picture we refer to as a mutation is actually the presence of a mutant, arisen within the producer of an infective droplet, containing a sample from the pathogen population harbored by it. An invading mutant is, moreover, supposed to take over the new host from among the variety, if any, of genotypes in the droplet. (This should not be taken literally. This assumption is only supposed to lead to a fair approximation at the level of the evolutionary predictions. Moreover, it may well be less contrived than it seems, given the large chance effects and concomitant genetic drift during the initial stages of the infection process (an indication is that after experimental intranasal infections with a dose way above that in an aerosol droplet, Jackson et al. (1958) found that out of 872 volunteers only less than 40% caught a cold).) From this perspective, within-host evolution is accounted for by the phenomenological host level, as opposed to pathogen level, mutation probability. For the respiratory diseases we consider, the disease duration is, moreover, so short that we expect (except perhaps in immunocompromised hosts) the within-host evolution to be relatively simple, producing on the between-host level only a relatively negligible "gene drive." (Note, moreover, that in respiratory pathogens this "gene drive" and the between-host selection will act in line in causing an upward evolution since mutants adapted to a higher position will also be overrepresented in an infective droplet.) Although for immunocompromised hosts that "gene drive" generally will be larger, we believe that even with present day medical attitudes such hosts are sufficiently rare that their contribution can safely be neglected when dealing with the longer between-individual time scales considered in this paper.

To where do our conclusions reach?

The prediction about the increase of the tendency for diversification with the host density, although inspired by the analysis of the evolution of respiratory diseases, should apply to just any diversification caused by limited cross-immunity. This

increase is linear for mass-action-style infections and starts to saturate when behavioral limitations on the hosts' encounter rate start to dominate. The result even extends beyond the restriction of small disease caused mortality if we interpret "host density" as the density realized in the endemic situation, although with the proviso that the host density then is no longer determined by the host on its own but by the dynamics of host and disease together, and thus co-depends on which disease traits maximize Q (Supplementary Material SI4).

The large generality above contrasts starkly with the strictness with which the term "respiratory diseases" should be interpreted, to wit, diseases living in and affecting only the respiratory system. It is for such diseases that the mechanistic arguments apply that lead to the trade-off underlying our evolutionary deductions. COVID-19, for example, does not fully fall into this category since it also has effects outside the respiratory tract, i.a., in the form of cytokine storms and thrombocytopenia. (See Crespi (2020) for evolutionary explanations of these effects in terms of differences between the life histories and consequent defence mechanisms of bats and humans.) We hypothesize though that the latter effect also has to do with immune disregulation (in the form of an autoimmune reaction affecting part of the clotting cascade) and that both effects occur only in a genetically delimited segment of the host population, so that without medical interventions, the theoretical considerations would still apply in the longer run, after selection has reduced these subgroups to a negligible proportion of the population. Another scenario with the same result is that those side effects only occur when the disease is first contracted later in life, which is bound to happen far more rarely in the expected endemic than in its present epidemic stage.

Generalizing from the COVID-19 example, we reiterate the general biological rule that specific cases are always more complex than can be accounted for in any remotely tractable overarching model. We, therefore, saw our model predictions primarily as theoretically emerging expectations about comparative trends, rather than as telling what will happen in specific concrete instances, and this paper primarily as an attempt to create an overarching picture for immune-driven disease diversification and thus provide a reference point for the study of specific cases.

Implications for emerging respiratory diseases

To conclude, we put forward an additional, more tentative, message that stems from the connection between epidemic models and competition models. The experience with competition models is that evolutionary simulations after an initial spate of diversification around the maximum carrying capacity subsequently start to fill up the inhabitable part of the trait space, but do so ever more slowly. The reason is that the changing fitness landscape remains "anchored" at zero where there are residents. More residents make for more zeros, and thus flatter fitness landscapes, smaller fitness gradients and slower evolution. (As residents by definition stay around on an evolutionarily relevant time-scale, their trait package should have invasion fitness - that is, longer term average per capita rate of growth or decline - zero.) In higher-dimensional trait spaces, the slowing down is less, but where it comes to the filling up of the outer reaches of the inhabitable trait space is replaced by a turmoil in which a largish number of roughly competitively matched strains dance around while the boundary of the swarm through branching slowly expands toward the outer reaches of the inhabitable trait space. This makes us expect that it will take relatively long for the deeper reaches of the lungs to become colonized by strains evolving back from higher up, thus leaving those reaches open to invaders from outside. This then leads to the expectation that emerging respiratory diseases more often than not will have low R_0 and high virulence.

A cautionary note is that in the earlier introduced high-dimensional trait space spanned by the respiratory depth axis along with additional axes affecting other aspects of its functioning (with some of them as side effect eliciting an immune reaction), a new entrant in the game more often than not initially will engage in a climb toward the fitness ridge located at the respiratory depth axis. Think, for example, of a virus coming equipped with tricks for coping with its former host, which still can considerably improve its manipulative activity inside the cells of its new one. That additional movement through trait space generally will not be without side effects (see Metz, 2011, subsection 3.1 for one reason why the latter are almost bound to occur). Therefore, the simple story told above will, in general, be, as COVID-19 has amply shown, only a gross summary of a more extensive one that is studded with less predictable detail depending on as yet unknown facts about the within-host part of the ecology of the pathogen.

As a final note, we point to the fact that the Omicron variant(s) of SARS-CoV-2 behaves exactly in line with our tentative predictions: it lives higher up in the respiratory tract and is less virulent, more infective and only partially immune to the earlier lower-living strains, see, for example, Bentley et al. (2021); McMahan et al. (2022); Meng et al. (2021); and Peacock et al. (2022). We are, of course, aware that this is only one data point and that there may exist also other mechanisms causing the same combination of properties. Yet, it is gratifying to see that both its position in the respiratory tract relative to that of the older strains and their replacement fits in well with our adaptive dynamics inspired predictions.[3] A technical objection might be that at the time, the COVID-19 epidemic had not yet relaxed to the endemic state, contrary to what is assumed in adaptive dynamics, and for that matter ESS reasoning. However, the prediction of an upward movement together with a decrease in virulence is, as is often the case, quite robust against deviations from the model details. In this case, because it essentially follows from the mechanistically derived trade-off, where virulence largely hitchhikes with the increase in infectivity as the main determinant of the selection gradient, independent of the details of the selection process. That we yet approached this problem through the simplified framework of adaptive dynamics is because the latter has the advantage of providing a coherent accessible picture of the associated complex of evolutionary phenomena (which only in the end motivated us to dive into the attendant concrete details).

Code availability

Numerical analysis was performed with Wolfram Mathematica 12. Computer codes are available at https://github.com/bboldin/evolution-respiratory-diseases.

[3] These predictions were actually made long before COVID-19 invaded the human population (with a first public appearance in a conference poster in 2012), and appear in print only now due to a long meandering route toward finally reaching the press.

Acknowledgements

Hans Metz benefited from the support from the "Chaire Modélisation Mathématique et Biodiversité of Veolia Environnement-École Polytechnique-Museum National d'Histoire Naturelle-Fondation X," from the marvellous programming of Kevin Kleine during the start-up of the reported research in 2010, from seeing the discrete time Lotka-Volterra competition simulations done by Mattias Siljestam, and, last but not least, his wife Ineke van der Sar who educated him on the clinicians' view about respiratory diseases and thus initiated the reported research. Barbara Boldin acknowledges the support of the Slovenian Research Agency (I0-0035, research program P1-0285 and research projects J1-9186, J1-1715). We are grateful to three anonymous reviewers whose comments considerably improved the manuscript.

Conflict of interest

The authors declare there is no conflict of interest.

Appendix A. The extension of the monomorphic branching criterion to higher-dimensional trait spaces

For higher-dimensional epitope spaces, the first change to Equation (18) is that the second derivatives have to be interpreted as Hessian matrices. In addition, the inequality sign should be interpreted as indicating that the left-hand side minus the right-hand side has at least one positive eigenvalue.

In the following, we use that $[\ln(Q)]''$ and φ''_{11} are symmetric matrices, that symmetric matrices, generically to be called A and B, are negative (positive) definite if and only if $X^{\mathrm{T}}AX <$ $(>)$ 0 for all $X \neq 0$, that $X^{\mathrm{T}}(A+B)X = X^{\mathrm{T}}AX + X^{\mathrm{T}}BX$ and that the highest and lowest eigenvalues of such matrices are equal to, respectively, the maximum and minimum of $\{X^{\mathrm{T}}AX\,;\ X^{\mathrm{T}}X = 1\}$.

Since $[\ln(Q)]''$ is negative definite, at $R_0 = QN_0 = 1$, the difference between the right-hand and the left-hand side in Equation (18) has only negative eigenvalues, and since φ''_{11} also is negative definite, increasing R_0 increases the eigenvalues, eventually making them all positive.

Note that, in general, branching in higher-dimensional trait spaces is a more complicated affair than may be expected from the one-dimensional case. In particular invadability of the singular point does not imply, as in the monomorphic case, that there exist nearby dimorphisms, or if such dimorphisms do exist, it may be that the diverging branches evolve out of the region where they can coexist in trait-space squared so that the evolutionary trajectory falls back to monomorphism again, and so on for ever (see Geritz et al., 2016), a phenomenon dubbed "trapping" by Zhang et al. (2023). Luckily, for the present model, we can exclude the occurrence of such trapping by falling back on a theorem from the same manuscript (see Supplementary Material SI10). One further result from Geritz et al. (2016) is that even in cases where diversification can initially start in more directions, as is the case when there is more than one positive eigenvalue, usually only the direction that goes with the largest eigenvalue remains. Simulations with discrete-time Lotka-Volterra models, which in Durinx et al. (2008) are proven close to evolutionarily singular points to be evolutionarily universal, in addition, it indicates that any higher-order branches in the case of a single positive eigenvalue occur in the direction of their parent branch, but otherwise do so transversally (personal communication Mattias Siljestam).

In the extreme case that the distance along the respiratory axis does not affect cross-immunity, the assumed multiplicativity of the fitness components coming from the (ecology modulated effect of) respiratory depth and from the (ecology-independent acting) additional epitope axes, makes that diversification will start orthogonal to this axis, but only after R_0 has become sufficiently large. Thereafter, all branches will eventually move on to the respiratory height maximizing R_0, which we by extrapolation of the earlier discovered regularities expect also to be the height supporting the largest diversity.

Appendix B. Simulating the adaptive dynamics

In the usual mutation-limitation approximation of adaptive dynamics, new mutants come one at a time in a Poisson process with a local rate that, after scaling out the system size and the mutation probability per infection event, at time t equals

$$\hat{B}_+ = \sum_{i=1}^{n} \hat{B}_i, \quad \text{where} \quad \hat{B}_i = \hat{\lambda}_i \hat{S}_i^+, \quad \hat{\lambda}_i = \beta_i \sum_{P\subseteq\mathcal{N}\setminus\{i\}} \hat{I}_{i,P},$$
$$\hat{S}_i^+ = \sum_{P\subseteq\mathcal{N}\setminus\{i\}} \theta_{i,P}\hat{S}_P \tag{B1}$$

where the quantities in Equation (B1) have to be calculated for the strains present at time $\tau(t)+$, where $\tau(t)$ is the time of the last mutant invasion before t. This means that after a mutation event, we can produce the waiting time till the next mutation by sampling from a standard exponential distribution and dividing the result by $\hat{B}_+$. The probability that the new mutation happens in strain i is then $\hat{B}_i/\hat{B}_+$. The mutant trait is generated by adding a small random number, distributed symmetrically around 0, to the trait value of its parent population. (Of course, close to 0 or 1 this distribution has to be adapted to keep the mutant between these numbers.) The next point is that as a result of it arriving singly the mutant may by chance fail to invade. To calculate its invasion probability, we observe that our assumption that immunity only affects the probability of infection makes that the mutant population initially grows according to a linear birth and death process, with per capita birth rate $b = \beta(Y)\sum_{P\subseteq\mathcal{N}} \theta\big(Y|\{X_j; j \in P\}\big)\hat{S}_P(X_1,\ldots,X_n)$ and per capita death rate $d = \delta+\alpha(Y)+\gamma(Y)$. Hence, its probability to make it to the deterministic realm is $1 - d/b = (b-d)/b = s/b = s/(d+s)$, where s is the invasion fitness of Y. If the mutant fails to invade, we just repeat the previous steps till a mutant comes that does. When this has happened, we have found our next $\tau(t)$ and we use Equation (21) starting from the old equilibrium values with a small admixture of $X_{n+1} = Y$ mutants to generate the corresponding population composition.

If we are only interested in the ESCs, we can, of course, replace all the population composition depended exponential random variables by a single constant, and if Equation (21) has only a single attractor, we can also replace the invasion probability by 1. (We conjecture that Equation (21) has at most a single internal attractor. However, when there is no internal attractor, there could be more than one boundary attractor, in which case the order of invasions and hence the probabilistic nature of the invasion process might matter for the evolutionary end result.)

Appendix C. Waning immunity

Here, we compare the basic model with the waning immunity model. When immunity wanes at a rate ε, the single-strain disease dynamics follows

$$\begin{aligned}\frac{dS_{\varnothing}}{dt} &= B + \varepsilon S_{\{1\}} - (\delta + \beta I_{\{1\}})S_{\varnothing}\\ \frac{dI_{\{1\}}}{dt} &= (\beta S_{\varnothing} - \alpha - \gamma - \delta)I_{\{1\}}\\ \frac{dS_{\{1\}}}{dt} &= \gamma I_{\{1\}} - (\delta + \varepsilon)S_{\{1\}}.\end{aligned} \tag{C1}$$

Note that this type of immunity decay at the population level, although seemingly coming from individuals randomly fully loosing their immunity in one go, while till that time it stays constant, can also come from a purely within-host exponential decay (see Appendix D, cf. Diekmann et al., 2020a,b).

At the non-trivial equilibrium

$$\begin{aligned}\hat{S}_{\varnothing} &= Q^{-1}\\ \hat{S}_{\{1\}} &= \frac{\delta/\varepsilon}{(\delta/\varepsilon + (\alpha+\delta)/(\alpha+\gamma+\delta))}\,\frac{\gamma}{\beta}(R_0 - 1)\\ \hat{I}_{\{1\}} &= \frac{\delta/\varepsilon}{(\delta/\varepsilon + (\alpha+\delta)/(\alpha+\gamma+\delta))}\,\frac{\delta+\varepsilon}{\beta}(R_0 - 1).\end{aligned} \tag{C2}$$

The invasion fitness is

$$s(Y|X) = \beta(Y)\Big(\hat{S}_{\varnothing}(X) + \big(1-\varphi(Y|X)\big)\hat{S}_{\{1\}}(X)\Big) - \big(\delta + \alpha(Y) + \gamma(Y)\big).$$

At $Y = X$ we have

$$G(X) = \beta'(X)\hat{S}_{\varnothing}(X) - \alpha'(X) - \gamma'(X).$$

The selection gradient is independent of ε and hence as before

$$G = (\delta + \alpha + \gamma)\big(\ln(Q)\big)'.$$

So the story for the monomorphic evolutionary dynamics up to reaching the evolutionarily singular point is exactly the same as in the case of permanent immunity. However, things start to change when we consider the diversification tendency. At $X = X^*$

$$\frac{\partial^2 s(Y|X)}{\partial Y^2}\Big|_{Y=X} = \beta''\hat{S}_{\varnothing} - \alpha'' - \gamma'' - \beta\varphi''_{11}\hat{S}_{\{1\}}.$$

Once again using $(\delta + \alpha + \gamma)\big(\ln(Q)\big)'' = Q^{-1}\beta'' - \alpha'' - \gamma''$ gives that X^* is a branching point if

$$\frac{\delta/\varepsilon}{\delta/\varepsilon + (\delta+\alpha)/(\delta+\alpha+\gamma)}\,\frac{\gamma}{\delta+\alpha+\gamma}\,(R_0 - 1)\,\frac{\varphi''}{\big(\ln(Q)\big)''} > 1, \tag{C3}$$

an expression similar to that for the case of permanent immunity, except that in the latter case $\varepsilon = 0$. More, in particular, the dependence of the readiness to diversify depends in the same manner on $R_0 - 1$, only the pre-factor differs.

Appendix D. Parameter estimation

For the quantitative fitting of any eco-evolutionary model, one also needs properties, and, therefore, parameter values, of organisms that could exist in principle but are not actually around. Although in the best imaginable world such parameters should be obtainable from mechanistic insights, for models for disease evolution this is still far from being a practical option. However, luckily we can do a little better if we focus on the branching criterion. There we need the disease parameters primarily for the disease strain that maximizes Q. Moreover, although this strain may not be present at this particular moment, the evolutionary trajectory will, in general, linger for some while in its neighborhood, and when branching takes place mainly along the additional epitope axes the kinetic parameters of our simplest model may be expected to change little with further evolution. If, in addition, the estimated kinetic parameters lie in a region where most of the parameter groups that occur in Formula (C3) depend only little on the precise values of the constituting parameters we are in business. This last property was actually the main reason for choosing the common cold as an example.

Another point is that most diseases are more complicated at the individual level than the simple Markov chain model that we have adopted for convenience. However, often such more complicated processes can where the equilibria are concerned be mimicked by simple Markov chain models, provided we chose for their parameters appropriately chosen statistics of the processes that they should mimic. Below we first discuss the estimation of the parameters of simple Markov chains, followed by a discussion of the statistics that in more complicated cases can be used for arriving at evolutionarily equivalent simple Markov chains. The concrete application to rhino- and corona-virus caused common colds can be found in Appendix E.

Many of the rate parameters of the differential equation model correspond to the parameters of a continuous time Markov chain model for the infected individuals. These parameters can thus be estimated from clinical data, according to the usual procedures for such models, by the following moment (and also maximum likelihood) estimators: the rate of leaving a Markovian state equals the inverse of the mean holding time, while the transition rates between states equal the leaving rate times the probabilities of those transitions. (For details like estimates of variances and covariances of the estimators see, e.g., the appendix of Dienske et al. (1980): reprinted in Metz (1981), downloadable at http://webarchive.iiasa.ac.at/Research/ADN/Metz1Book.html.) If we use the same estimator for semi-Markov chain models, that is, models for which the holding times are not exponentially distributed, while semi-state transitions happen in the same manner as in ordinary Markov chains, then the mean fraction of the time spent in each semi-state is the same as calculated for an ordinary Markov chain parameterized with the same statistics. This applies to (semi-)Markov chains allowing stationary states as well as to defective (semi-)Markov chains that are "born" at a given stationary rate, like the different disease semi-states in the endemic equilibrium of an infectious disease.

Things become a little more complicated when we want to consider properties like infectivity and (partial) immunity. If the average infectivity is not constant but depends on the age in a disease semi-state, we can just calculate the average total infectivity, while in that semi-state (i.e., the average area under the infectivity curve), multiply that with the rate at which that semi-state is "born" and add over all semi-states. This then gives the equilibrium value of λ_i. To get at the equivalent of β, we should divide the total infectivity per individual by the mean duration of the period between infection and the end of the illness, that is, the start of the period where we want to consider the host as being recovered (or dead, in any case noninfective). However, unfortunately infectivity as defined is not strictly an individual level quantity, since

it is only proportional to the rate of production of infective particles, or the rate of virus shedding, which are the individual level quantities. To get infectivity from these, they have to be multiplied with the probability that a shed inoculum reaches a new host times the probability that a fully susceptible host would then get infected by it. However, neither does infectivity appear separately in Formula (C3), but only as a component of R_0, which can be estimated in many different manners as described by, for example, Anderson and May (1992), Britton and Scalia Tomba (2019), and Diekmann et al. (2013), and in the elusive $[\ln(Q)]''$, of which we only discuss its relation to φ''.

In the case of waning immunity, it is rarely so that immunity wanes in a Markovian manner, that is, abruptly changes from 1 to 0 in a random manner with a fixed rate constant, as we seemingly assumed in Appendix C. Rather, the immunity status of a single individual changes in a continuous fashion taking values between zero and one. We shall assume here that the immunity can be decomposed as $\varphi(Y|X)\psi(\tau)$, where $\varphi(Y|X)$ is the specificity to a challenge Y of the antibodies elicited by X, satisfying the same restrictions as in the Markovian case, and $\psi(\tau)$ is the immune strength in dependence of the time since recovery, caused by the changing size of the antibody carrying cell population. Only when $\psi(\tau)$ decays exponentially starting from 1 the differential equations in Appendix C still describe the population immunity in the sense that in the monomorphic case partially immune individuals with immune strength $\psi(\tau)$ can be attributed to $S_{\{1\}}$ for a fraction $\psi(\tau)$ and for a fraction $1-\psi(\tau)$ to $S_{\varnothing}$, with similar but more complicated expressions in the polymorphic case. If immunity does not decay in this manner this allocation rule no longer leads to the traditional differential equations, but still applies in the community equilibrium with $\psi(\tau)$ replaced by $\varepsilon^{-1} = \int_0^\infty \psi(\tau)\mathrm{d}\tau$. The reason is that only the probabilities that newly received Y inocula infect randomly chosen hosts that once were infected with X, matter both for the calculation of the equilibrium and for the invasion of new mutants. (This on the assumption that Formula (6) holds good; when the immunity due to more than one previous infection is determined differently, the previous argument only applies to the monomorphic case.)

For the invasion dynamics of new mutants the simple time averaging argument used for the equilibria suffices only for the determination of the rate at which they can infect, but this is not enough for the calculation of the initial growth rate of the mutant population. For that purpose, we also need the relation $s \approx \ln(R_i)/T$, where R_i is the equivalent of R_0 but now for an invader in the environment created by the currently circulating strains, and T is the mean waiting time since infection to producing a daughter infection, where without changing the order of the approximation we may substitute the value for the "parent" of the mutant. This mean waiting time equals the mean latent period plus the mean of the normalized infection kernel g. Since g equals the normalized curve of infective particle shedding, this is again a fully individual level quantity. (For the estimation of T from epidemic as opposed to experimental data see Britton & Scalia Tomba , 2019; Scalia Tomba et al., 2010.) In Appendix B of Durinx et al. (2008), it is shown that we can use this approximate expression even for calculating the second derivatives of s near an evolutionary singularity. Now define the quantities α, β, γ, δ, and Q for semi-Markov individual behavior in the same manner as before. Then $R_i = Q(Y)\left(\hat{S}_{\varnothing}(X) + \left(1-\varphi(Y|X)\right)\hat{S}_{\{1\}}(X)\right)$ just as before. Not only that, if we take first and second derivatives of $s(Y)$ for Y, we get exactly the same expressions as we did for the simple differential equation model. So the branching criterion Equation (22) is also valid in this more general case.

Appendix E. Common colds

Lets first again consider condition (C3),

$$\frac{\delta/\varepsilon}{\delta/\varepsilon + (\delta+\alpha)/(\delta+\alpha+\gamma)}\,\frac{\gamma}{\delta+\alpha+\gamma}\,(R_0-1)\,\frac{\varphi''}{\left(\ln(Q)\right)''} > 1,$$

to see on which parameters we should concentrate. The terms $\gamma/(\delta+\alpha+\gamma)$ and $(\delta+\alpha)/(\delta+\alpha+\gamma)$ can be interpreted as the probabilities to survive and to die respectively during a cold. We all know from experience that the latter probability is very small, say less than 0.01%, making $\gamma/(\delta+\alpha+\gamma) \approx 1$. To see what the effect of $(\delta+\alpha)/(\delta+\alpha+\gamma)$ is on the first term of Equation (22) we have to find an estimate of δ/ε. For δ, we may without erring too much substitute $\delta \approx 1/80$. For the estimation of ε, we had to go back to the experiments in Jackson et al. (1962) as later data sets only give immune titers, not immunity. Jackson et al. (1962) could not yet put a name to the virus they used as initial infection as well as later challenges. The reason for believing that this common cold was caused by a rhinovirus are threefold. Firstly, common colds are most often caused by rhinoviruses (30–40%) with corona viruses as second (10–20%) (Lorber, 1996). Secondly, the shape of the antibody response against corona in Callow et al. (1990) differs greatly from the shape seen in Jackson et al. (1962), whereas, thirdly, the shape of the antibody response against a rhinovirus shown in Barclay et al. (1989) grossly agrees. The area under the curve in Figure 4 of Jackson et al. (1962) amounts to ca 0.86 immunity years. However, that curve ends after two years at an immunity of ca 0.16. Linearly extrapolating the curve from there would amount to about 0.14 additional immunity years, but this is only a lower bound, neglecting the suggested upward curvature of the curve. The age dependence of the frequency of rhinovirus infections in the population even suggests that this lower bound is far from sharp. So we guesstimate the immunity years conferred by a rhinovirus infection at around 3. (Of course, different rhinovirus strains could induce very different immunity time courses. However, the total of the data in Callow et al. (1990) and Jackson et al. (1962) suggests the different values of ε to be at least in the same ballpark.)

To estimate R_0-1, we use that the number of new infection per year equals the number of ill hosts divided by the average duration of the illness, $\hat{I}_{\{1\}}(\delta+\alpha+\gamma)$. Dividing this by the density of hosts gives (see Appendix C)

$$\begin{aligned}\frac{(\delta+\alpha+\gamma)\hat{I}_{\{1\}}}{N_0} &= \frac{\delta/\varepsilon}{\delta/\varepsilon + (\delta+\alpha)/(\delta+\alpha+\gamma)}(\delta+\varepsilon)(1-R_0^{-1}) \\ &\approx \varepsilon(1-R_0^{-1}). \qquad \text{(E1)}\end{aligned}$$

Let our lifetime average of common colds per year be 4, of which only 40% are caused by rhino viruses of which there circulate, say, 50 fully compatible ones, then for a single strain the number of infections per year per head is ca 0.03. This then gives $R_0 \approx 1.1$.

For corona caused colds the data are less complete. The best data on immunity, by Callow et al. (1990), are in the form of log geometric mean concentrations (over volunteers) of specific antibodies, not of frequencies in previously infected

relative to uninfected volunteers of staving of a new challenge, which would be the appropriate estimate of the ψ from Appendix D. To obtain at least some numbers we assumed ψ to be proportional to the concentration of serum specific IgG (for coronavirus 229E) reported in Callow et al. (1990) with a maximum of 0.8. This then gave 0.25 as an initial estimate for ε^{-1}. However, in the second half of the considered period, the immunity seemed to wane slower than that of the rhinoviruses considered by Jackson et al. (1962), which would in the end bring our estimate for ε in the ballpark of our earlier estimate for rhinoviruses. Moreover, we have to account for the fact that there appears to be some cross-immunity between corona strains, lowering their incidence. As we have too little data on that cross-immunity, we decided to go for an inequality only. With 20% of the cases and only four serotypes this then gives $R_0 \gtrsim 2.5$.

References

Acevedo, M. A., Dillemuth, F. P., Flick, A. J., Faldyn, M. J., & Elderd, B. D. (2019). Virulence-driven trade-offs in disease transmission: A meta-analysis. *Evolution*, *73*(4):636–647.

Alizon, S., Hurford, A., Mideo, N., & Van Baalen, M. (2009). Virulence evolution and the trade-off hypothesis: History, current state of affairs and the future. *Journal of Evolutionary Biology*, *22*(2):245–259.

Anderson, R. M. & May, R. M. (1982). Coevolution of hosts and parasites. *Parasitology*, *85*:411–426.

Anderson, R. M., & May, R. M. (1992). *Infectious diseases of humans: Dynamics and control*. Oxford University Press.

Andreasen, V., Lin, J., & Levin, S. A. (1997). The dynamics of cocirculating influenza strains conferring partial cross-immunity. *Journal of Mathematical Biology*, *35*(7):825–842.

Andrewes, C. H. (1950). Adventures among viruses. III. The puzzle of the common cold. *The New England Journal of Medicine*, *242*:235–240.

Barclay, W. S., al Nakib, W., Higgins, P. G., & Tyrrell, D. A. J. (1989). The time course of the humoral immune response to rhinovirus infection. *Epidemiology and Infection*, *103*(3):659–669.

Bentley, E. G., Kirby, A., Sharma, P., Kipar, A., Mega, D. F., Bramwell, C., Penrice-Randal, R., Prince, T., Brown, J. C., Zhou, J., et al. (2021). SARS-Cov-2 Omicron-b. 1.1. 529 variant leads to less severe disease than pango b and delta variants strains in a mouse model of severe COVID-19. *BioRxiv*. doi: https://doi.org/10.1101/2021.12.26.474085

Best, A., & Hoyle, A. (2013). A limited host immune range facilitates the creation and maintenance of diversity in parasite virulence. *Interface Focus*, *3*:20130024.

Bhattacharyya, S., Gesteland, P. H., Korgenski, K., Bjørnstad, O. N., & Adler, F. R. (2015). Cross-immunity between strains explains the dynamical pattern of paramyxoviruses. *Proceedings of the National Academy of Sciences*, *112*(43):13396–13400.

Boldin, B. and Metz, J. A. J. (2023). The evolution of disease diversity with an application to respiratory diseases. In preparation.

Brännström, Å., Johansson, J., & Von Festenberg, N. (2013). The hitchhiker's guide to adaptive dynamics. *Games*, *4*(3):304–328.

Britton, T., & Scalia Tomba, G. (2019). Estimation in emerging epidemics and remedies. *Journal Royal Society Interface*, *16*:20180670.

Callow, K. A., Parry, H. F., Sergeant, M., & Tyrrell, D. A. J. (1990). The time course of the immune response to experimental coronavirus infection of man. *Epidemiology and Infection*, *105*(2):435–446.

Crespi, B. (2020). Evolutionary medical insights into the SARS-CoV-2 pandemic. *Evolution, Medicine, and Public Health*, *2020*(1):314–322.

Cui, J., Li, F., & Shi, Z.-L. (2019). Origin and evolution of pathogenic coronaviruses, *Nature Reviews Microbiology* **17**, 181–192.

de Jong, M. C. M., & Janss, L. L. G. (2002). Virulence management in veterinary epidemiology. In U. Dieckmann, J. A. J. Metz, M. W. Sabelis, & K. Sigmund (Eds.), *Adaptive dynamics of infectious diseases: In Pursuit of virulence management* (pp. 425–435). Cambridge University Press.

Diekmann, O. (2004). A beginner's guide to adaptive dynamics. In *Mathematical modelling of population dynamics* (Vol. 63, pp. 47–86). Polish Academy of Sciences.

Diekmann, O., Gyllenberg, M., & Metz, J. A. J. (2003). Steady-state analysis of structured population models. *Theoretical Population Biology*, *63*(4):309–338.

Diekmann, O., Gyllenberg, M., & Metz, J. A. J. (2020a). Correction to: Finite dimensional state representation of physiologically structured populations. *Journal of Mathematical Biology*, *81*:905–906.

Diekmann, O., Gyllenberg, M., & Metz, J. A. J. (2020b). Finite dimensional state representation of physiologically structured populations. *Journal of Mathematical Biology*, *80*(1–2):205–273.

Diekmann, O., Heesterbeek, H., & Britton, T. (2013). *Mathematical tools for understanding infectious disease dynamics*. Princeton University Press.

Diekmann, O., Heesterbeek, J. A. P., & Metz, J. A. J. (1990). On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, *28*(4):365–382.

Dienske, H., Metz, J. A. J., van Luxemburg, P. J. C. M., & de Jonge, G. (1980). Mother-infant body contact in macaques II: Further steps towards a representation as a continuous time Markov chain. *Biology of Behaviour*, *5*:51–94.

Durinx, M., Metz, J. A. J., & Meszéna, G. (2008). Adaptive dynamics for physiologically structured models. *Journal of Mathematical Biology*, *56*(5):673–742.

Earn, D. J. D., Dushoff, J., & Levin, S. A. (2002). Ecology and evolution of the flu. *Trends in Ecology & Evolution*, *17*(7):334–340.

Erkkola, R., Turunen, R., Räisänen, K., Waris, M., Vuorinen, T., Laine, M., Tähtinen, P., Gern, J. E., Bochkov, Y. A., Ruohola, A., et al. (2020). Rhinovirus C is associated with severe wheezing and febrile respiratory illness in young children. *The Pediatric Infectious Disease Journal*, *39*(4):283.

Feldman, A. S., Hartert, T. V., Gebretsadik, T., Carroll, K. N., Minton, P. A., Woodward, K. B., Larkin, E. K., Miller, E. K., & Valet, R. S. (2015). Respiratory severity score separates upper versus lower respiratory tract infections and predicts measures of disease severity. *Pediatric Allergy, Immunology, and Pulmonology*, *28*(2):117–120.

Forni, D., Cagliani, R., Clerici, M., & Sironi, M. (2017). Molecular evolution of human coronavirus genomes. *Theoretical Population Biology*, *25*(1):35–46.

Geritz, S. A. H., Kisdi, É., Meszéna, G., & Metz, J. A. J. (1998). Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evolutionary Ecology*, *12*(1):35–57.

Geritz, S. A. H., Metz, J. A. J., & Rueffler, C. (2016). Mutual invadability near evolutionarily singular strategies for multivariate traits, with special reference to the strongly convergence stable case. *Journal of Mathematical Biology*, *72*(4):1081–1099.

Gog, J. R., & Grenfell, B. T. (2002). Dynamics and selection of many-strain pathogens. *Proceedings of the National Academy of Sciences*, *99*(26):17209–17214.

Heesterbeek, J. A. P., & Metz, J. A. J. (1993). The saturating contact rate in marriage-and epidemic models. *Journal of Mathematical Biology*, *31*(5):447–471.

Heikkinen, T., & Järvinen, A. (2003). The common cold. *The Lancet*, *361*(9351):51–59.

Holt, R. D., & Bonsall, M. B. (2017). Apparent competition. *Annual Review of Ecology Evolution, and Systematics*, *48*:529–539.

Jackson, G. G., & Dowling, H. F. (1959). Transmission of common cold to volunteers under controlled conditions. IV. Specific immunity to common cold. *Journal of Clinical Investigation*, *38*(5):762–769.

Jackson, G. G., Dowling, H. F., Akers, L. W., Muldoon, R. L., Van Dyke, A., & Johnson, G. C. (1962). Immunity to the common cold from protective serum antibody: Time of appearance,

persistence and relation to reinfection. *The New England Journal of Medicine*, *266*(16):791–796.

Jackson, G. G., Dowling, H. F., Spiesman, I. G., & Boand, A. V. (1958). Transmission of the common cold to volunteers under controlled conditions. I. The common cold as a clinical entity. *AMA Archives of Internal Medicine*, *101*(2):267–278.

Jacobs, S., Lamzon, D. M., St.George, K., & Walsh, T. J. (2013). Human rhinoviruses. *Clinical Microbiology Reviews*, *26*(1):135–162.

Koelle, K., Cobey, S., Grenfell, B., & Pascual, M. (2006). Epochal evolution shapes the phylodynamics of interpandemic influenza a (h3n2) in humans. *Science*, *314*(5807):1898–1903.

Lee, R. J., Xiong, G., Kofonow, J. M., Chen, B., Lysenko, A., Jiang, P., Abraham, V., Doghramji, L., Adappa, N. D., Palmer, J. N., et al. (2012). T2R38 taste receptor polymorphisms underlie susceptibility to upper respiratory infection. *The Journal of Clinical Investigation*, *122*(11):4145–4159.

Lin, J., Andreasen, V., & Levin, S. A. (1999). Dynamics of influenza A drift: The linear three-strain model. *Mathematical Biosciences*, *162*(1–2):33–51.

Lion, S., & Metz, J. A. J. (2018). Beyond R_0 maximisation: On pathogen evolution and environmental dimensions. *Trends in Ecology & Evolution*, *33*(6):458–473.

Lorber, B. (1996). The common cold. *Journal of General Internal Medicine*, *11*(4):229–236.

McMahan et al. (2022). Reduced pathogenicity of the SARS-CoV-2 omicron variant in hamsters. *Med* **3**, 262–268 .

Meng, B., Ferreira, I., Abdullahi, A., Kemp, S. A., Goonawardane, N., Papa, G., Fatihi, S., Charles, O., Collier, D., Choi, J., et al. (2021). SARS-Cov-2 Omicron spike mediated immune escape, infectivity and cell-cell fusion. *BioRxiv*. doi: https://doi.org/10.1101/2021.12.17.473248

Metz, J. A. J. (1978). The epidemic in a closed population with all susceptibles equally vulnerable: Some results for large susceptible populations and small initial infections. *Acta Biotheoretica*, *27*(1):75–123.

Metz, J. A. J. (1981). *Mathematical representations of the dynamics of animal behaviour: An expository survey*. Mathematisch Centrum.

Metz, J. A. J. (2011). Thoughts on the geometry of meso-evolution: Collecting mathematical elements for a postmodern synthesis. In *The mathematics of Darwin's legacy* (pp. 193–231). Springer.

Metz, J. A. J. (2012). Adaptive dynamics. In A. Hastings, & L. Gross, editors, *Encyclopedia of theoretical ecology* (pp. 7–17). California University Press.

Metz, J. A. J., & de Kovel, C. G. F. (2013). The canonical equation of adaptive dynamics for mendelian diploids and haplo-diploids. *Interface Focus*, *3*(6):20130025.

Metz, J. A. J., Geritz, S. A. H., Meszéna, G., Jacobs, F. J. A., & Van Heerwaarden, J. S. (1996). Adaptive dynamics: A geometrical study of the consequences of nearly faithful reproduction. In *Stochastic and spatial structures ofdynamical systems* (pp. 183–231). North-Holland.

Metz, J. A. J., & Gyllenberg, M. (2001). How should we define fitness in structured metapopulation models? Including an application to the calculation of evolutionarily stable dispersal strategies. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1466):499–508.

Monto, A. S., Bryan, E. R., & Ohmit, S. (1987). Rhinovirus infections in Tecumseh, Michigan: Frequency of illness and number of serotypes. *Journal of Infectious Diseases*, *156*(1):43–49.

Monto, A. S., & Cavalarro, J. J. (1971). The Tecumseh study of respiratory illness, II. Patterns of occurrence of infection with respiratory pathogens, 1965–1969. *American Journal of Epidemiology*, *94*(3):280–289.

Peacock, T. P., Brown, J. C., Zhou, J., Thakur, N., Newman, J., Kugathasan, R., Sukhova, K., Kaforou, M., Bailey, D., & Barclay, W. S. (2022). The SARS-Cov-2 variant, Omicron, shows rapid replication in human primary nasal epithelial cultures and efficiently uses the endosomal route of entry. *BioRxiv*, pages 2021–12. doi: https://doi.org/10.1101/2021.12.31.474653

Reperant, L. A., Kuiken, T., Grenfell, B., Osterhaus, D. M. E., & Dobson, A. P. (2012). Linking influenza virus tissue tropism to population-level reproductive fitness. *PLoS One*, *7*(8): e43115.

Restif, O., & Grenfell, B. T. (2006). Integrating life history and cross-immunity into the evolutionary dynamics of pathogens. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1585):409–416.

Scalia Tomba, G., Svensson, A., Asikainen, T., & Giesecke, J. (2010). Some model based considerations on observing generation times for communicable diseases. *Mathematical Biosciences*, *223*: 24–31.

Thompson, M., Vodicka, T. A., Blair, P. S., Buckley, D. I., Heneghan, C., & Hay, A. D. (2013). Duration of symptoms of respiratory tract infections in children: Systematic review. *BMJ*, *347*:f7027.

Trapman, P., Ball, F., Dhersin, J. S., Tran, V. C., Wallinga, J., & Britton, T. (2016). Inferring R_0 in emerging epidemics: the effect of common population structure is small. *Journal Royal Society Interface*, *13*:20160288.

van Regenmortel, M. H. V., Fauquet, C. M., Bishop, D. H. L., Carstens, E. B., Estes, M. K., Lemon, S. M., Maniloff, J., Mayo, M. A., McGeoch, D. J., Pringle, C. R., et al. (2000). *Virus taxonomy: Classification and nomenclature of viruses. Seventh report of the International Committee on Taxonomy of Viruses*. Academic Press.

Zhang, F., Dieckmann, U., Metz, J. A. J., & Hui, C. (2023). Condition for evolutionary branching and trapping in multidimensional adaptive dynamics. Manuscript in preparation.