Heterozygote advantage can explain the extraordinary diversity of immune genes

Mattias Siljestam[1*] **and Claus Rueffler**[1]

***For correspondence:**
mattias.siljestam@ebs.uu.se (MS)

[1] Department of Ecology and Genetics, Animal Ecology, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden

Abstract The majority of highly polymorphic genes are related to immune functions and with over 100 alleles within a population genes of the major histocompatibility complex (MHC) are the most polymorphic loci in vertebrates. How such extraordinary polymorphism arose and is maintained is controversial. One possibility is heterozygote advantage which can in principle maintain any number of alleles but biologically explicit models based on this mechanism have so far failed to reliably predict the coexistence of significantly more than ten alleles. We present an eco-evolutionary model showing that evolution can in a self-organising process result in the emergence of more than 100 alleles maintained by heterozygote advantage. Thus, our model shows that heterozygote advantage is a much more potent force in explaining the extraordinary polymorphism found at MHC loci than currently recognised. Our model applies more generally and can explain polymorphism at any gene involved in multiple functions important for survival.

Introduction

Heterozygote advantage is a well established explanation for single locus polymorphism with the sickle cell locus as a classical text book example (***Allison, 1954***). However, whether heterozygote advantage is generally important for the maintenance of genetic polymorphism is questioned (***Hedrick, 2012***; ***Sellis et al., 2016***). Genes of the major histocompatibility complex (MHC), which are responsible for inducing immune defence by recognising pathogenic antigens, are the most polymorphic loci among vertebrates (***Duncan et al., 1979***; ***Apanius et al., 1997***; ***Penn, 2002***; ***Sommer, 2005***; ***Eizaguirre and Lenz, 2010***). Heterozygote advantage as an explanation for this high level of polymorphism was introduced more than 40 years ago by ***Doherty and Zinkernagel*** (***1975***) with the idea that individuals with MHC molecules from two different alleles can recognize a broader spectrum of pathogens resulting in higher fitness. While early theoretical work suggested that heterozygote advantage can maintain the high number of alleles found at MHC loci (***Kimura and Crow, 1964***; ***Wright, 1966***; ***Maruyama and Nei, 1981***) later work discredited this idea as more mechanistic models failed to reliably predict very high allele numbers (***Lewontin et al., 1978***; ***Spencer and Marks, 1988***; ***Hedrick, 2002***; ***de Boer et al., 2004***; ***Borghans et al., 2004***; ***Stoffels and Spencer, 2008***; ***Trotter and Spencer, 2008***, ***2013***; ***Ejsmond and Radwan, 2015***; ***Lau et al., 2015***). Therefore, heterozygote advantage plays only a minor role in current explanations of polymorphism at MHC loci (***Hedrick, 1999***; ***Gould et al., 2004***; ***Wegner, 2008***; ***Kekäläinen et al., 2009***; ***Eizaguirre and Lenz, 2010***; ***Lenz, 2011***; ***Loiseau et al., 2011***) despite empirical evidence for it (***Doherty and Zinkernagel, 1975***; ***Hughes and Nei, 1989***; ***Jeffery and Bangham, 2000***; ***McClelland et al., 2003***; ***Froeschke and Sommer, 2005***; ***Kekäläinen et al., 2009***; ***Oliver et al., 2009***; ***Lenz, 2011***) and other mechanisms are

suggested to be important (***Apanius et al., 1997***; ***Hedrick, 1999***; ***Penn, 2002***; ***Borghans et al., 2004***; ***Wegner, 2008***; ***Spurgin and Richardson, 2010***; ***Loiseau et al., 2011***; ***Ejsmond and Radwan, 2015***).

In contrast, our work showing that heterozygote advantage is a potent force which can drive the evolution and subsequent maintenance of more than 100 alleles. Importantly, our results do not rely on carefully fine-tuned genotypic fitness values. Instead, allelic values that allow for extraordinary polymorphism are found by evolution in a self-organised process. We do not claim that heterozygote advantage is the sole mechanism responsible for the diversity of MHC alleles found in nature. However, our results show that heterozygote advantage can be more important than currently recognised.

Model

In our model, MHC molecules produced by the two alleles at a focal locus govern the performance of the immune response against m pathogens. This immune performance determines the condition c of an individual, for instance, body weight (***Eizaguirre et al., 2012***), which, in turn, determines survival s. A crucial assumption is the existence of a trade-off in the efficiencies of MHC molecules to induce a defence against different pathogens. Thus, no MHC molecule can perform optimally with respect to all pathogens and an improved efficiency against one set of pathogens can only be achieved at the expense of a decreased efficiency against another set of pathogens. Hence, generalist MHC molecules can detect many pathogens with low efficiency and specialist MHC molecules can detect few pathogens but with high efficiency. Two lines of empirical evidence support such trade-offs. First, many MHC molecules can detect only a certain set of antigens (***Wakeland et al.,***

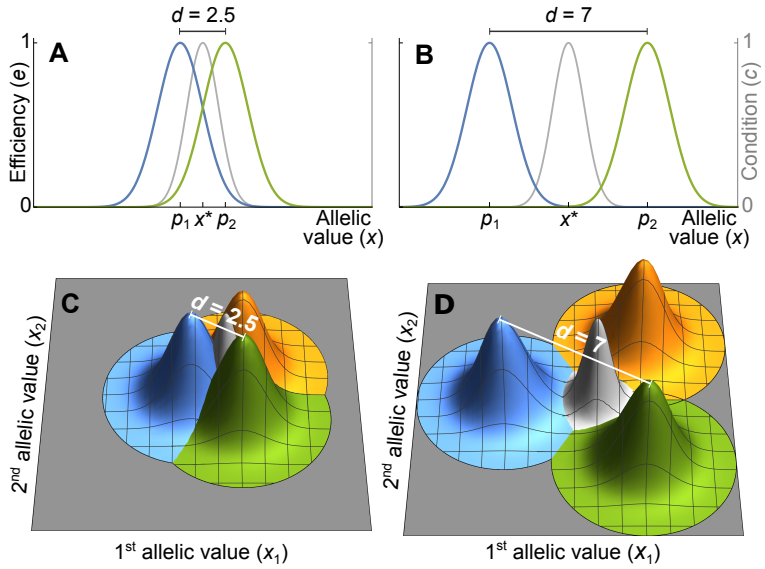

Figure 1. Gaussian trade-offs for an MHC molecule facing (***A-B***) two and (***C-D***) three pathogens. (***A-B***) The two pathogens are defined by the numbers p_1 and p_2, respectively, and separated by distance d. The blue and green curves show the Gaussian efficiency functions $e(p_1 - x)$ and $e(p_2 - x)$ for these pathogens as a function of the allelic value x. (***C-D***) Alleles and pathogens are defined by two-dimensional vectors $\boldsymbol{x} = (x_1, x_2)$ and $\boldsymbol{p} = (p_1, p_2)$, respectively, describing points on the plane. The three vectors $\boldsymbol{p}_1$, $\boldsymbol{p}_2$ and $\boldsymbol{p}_3$ describing the different pathogens are separated by distance d. Coloured cones give the two-dimensional Gaussian efficiency functions $e(\boldsymbol{p}_1 - \boldsymbol{x})$, $e(\boldsymbol{p}_2 - \boldsymbol{x})$ and $e(\boldsymbol{p}_3 - \boldsymbol{x})$ for these pathogens as a function of the allelic trait vector $\boldsymbol{x}(x_1, x_2)$. We refer to the allele located in the centre with equal distance to the pathogen optima as the generalist allele $\boldsymbol{x}^*$. The efficiency of this generalist allele for the different pathogens is determined by the distance d between the pathogen optima. If pathogens are similar (***A, C***) the generalist has a high efficiency against both pathogens while for more dissimilar pathogens (***B, D***) the generalist has a low efficiency. The grey lines (***A-B***) and grey cones (***C-D***) give the condition c of a homozygote genotype with the corresponding allelic trait values. The generalist allele (indicated by x^* in ***A, B***) gives the maximal homozygote condition c^*. Without loss of generality c^* is standardised to 1 by choosing $c_{\max}$ in ***Equation 1*** appropriately.

1990; *Froeschke and Sommer, 2012*; *Eizaguirre et al., 2012*; *Chappell et al., 2015*) and therefore provide different degrees of protection against different pathogens (*Wakeland et al., 1990*; *Apanius et al., 1997*; *Eizaguirre and Lenz, 2010*; *Froeschke and Sommer, 2012*; *Eizaguirre et al., 2012*; *Cortazar-Chinarro et al., 2019*). Second, it has been found that specialist MHC molecules are expressed at higher levels at the cell surface while generalist MHC molecules that bind less selectively are expressed at lower levels, potentially to reduce the harm of binding self-peptides (*Chappell et al., 2015*). This could explain the lower efficiency of generalist MHC molecules.

The ability of MHC molecules to recognise pathogens is modelled as follows. We assume that each MHC molecule contains ten functionally important parts or modules. Each of these modules is characterised by a real number x_i, collected in a vector $\boldsymbol{x} = (x_1, x_2, ..., x_{10})$, and coded by a different region of the MHC allele (note that throughout letters in bold italic font refer to multi-dimensional vectors while letters in normal italic font refer to scalars). Recombination between regions occurs with probability r. The ability of MHC molecules to recognise pathogens depends on the vector $\boldsymbol{x}$ and properties of the pathogen. Specifically, we assume that for each module of the MHC molecule pathogens have a corresponding property, characterised by a real number p_i and collected in a vector $\boldsymbol{p} = (p_1, p_2, ..., p_{10})$, that mediates the interaction between MHC molecules and the pathogen. The efficiency of an MHC molecule with vector $\boldsymbol{x}$ against a pathogen with vector $\boldsymbol{p}$ is then modelled as a Gaussian function $e(\boldsymbol{p} - \boldsymbol{x})$ of the Euclidean distance between these two vectors (*Figure 1* and *Equation S10*). Thus, the efficiency of an MHC molecule to recognise a pathogen decreases with increasing mismatch between the vectors describing the MHC molecule and pathogen. In our model we vary the number m of pathogens from two to eight.

We investigate three versions of this general model. The simplest version is based on no recombination ($r = 0$) and two symmetry assumptions. The m vectors describing the different pathogens are arranged symmetrically such that the distance between any two pathogens equals d (see *Figure 1*) and the Gaussian functions e are rotationally symmetric and of equal width (as in *Figure 1*c,d). In the second version, we allow for recombination in this symmetric model. The third version of the model is again based on $r = 0$ but relaxes the above symmetry assumptions. Since the results for the latter two version are qualitatively similar to the results from the first version we focus on the case without recombination and full symmetry and refer to *SI Appendix 4 and 7* and *SI Appendix 5* for results of the more complex models. Note, that in the absence of recombination it is without loss of generality possible to reduce the dimension of the trait space to $m - 1$, i.e., to one dimension less than the pathogen number, such that $\boldsymbol{x} = (x_1, x_2, ..., x_{m-1})$ and $\boldsymbol{p} = (p_1, p_2, ..., p_{m-1})$ (*Figure 1*). Mathematically speaking, this is because the m vectors describing the pathogens can be encapsulated in a convex set of dimension $m - 1$ (known as convex hull) and allelic vectors outside this set are maladapted for all pathogens along at least one dimension.

Since MHC alleles are codominantly expressed (*Eizaguirre and Lenz, 2010*; *Abbas et al., 2014*) an individual's efficiency to recognise pathogens of type k is given by the arithmetic mean of the efficiencies from its two alleles, $(e(\boldsymbol{p}_k - \boldsymbol{x}_i) + e(\boldsymbol{p}_k - \boldsymbol{x}_j))/2$. The phenotypic condition c of an individual is determined by its performance against all m pathogens. Specifically, we assume that during their life time individuals are exposed to all pathogens and therefore condition is given by the product of its defences against all m pathogens,

$$c(\boldsymbol{x}_i, \boldsymbol{x}_j) = c_{\max} \prod_{k=1}^{m} \frac{e(\boldsymbol{p}_k - \boldsymbol{x}_i) + e(\boldsymbol{p}_k - \boldsymbol{x}_j)}{2}, \tag{1}$$

where $c_{\max}$ is the condition of a hypothetical individual with perfect defence against all pathogens. The multiplicative nature of Eq. (1) implies that a poor defence against a single pathogen is sufficient to severely compromise condition c. Note that, since the Gaussian function $e(\boldsymbol{p}_k - \boldsymbol{x}_i) \leq 1$ (see *Equation S10*), the realised condition is always less than $c_{\max}$.

Finally, survival s is an increasing but saturating function of an individual condition c,

$$s(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{c(\boldsymbol{x}_i, \boldsymbol{x}_j)}{K + c(\boldsymbol{x}_i, \boldsymbol{x}_j)}. \tag{2}$$

Here, K is the survival half-saturation constant. This survival function increases approximately linearly for c smaller than K and saturates for large c where s approaches 1 (orange lines in bottom panel of ***Figure 2***).

Analysis

We study the evolutionary dynamics of allelic values $\boldsymbol{x}$ in a diploid Wright-Fisher model with selection (***Fisher, 1930***; ***Wright, 1931***) by following the fate of recurrent mutations. Thus, we consider a diploid population of fixed size N with non-overlapping generations and random mating resulting in Hardy-Weinberg proportions before viability selection. Initially the population is fixed for a single allele. We then ask whether a mutant allele can invade and if so, if it replaces its ancestor or coexists with it. In the latter case, we start following two allelic lineages that each can give rise to new variants through mutation. The long-term evolutionary dynamics is obtained by iterating this procedure (***Figure 2*** top panel). The evolution of the emerging allelic lineages is in this model driven by increasing the allele's expected survival (or marginal fitness). This follows from standard population genetics theory (see ***Equation S2-S3*** in ***SI Appendix 6***). Assuming rare mutations of small effect allows to analytically determine the evolutionary dynamics using the adaptive dynamics framework (***Metz et al., 1992***; ***Geritz et al., 1998***). For detailed derivations of all results see ***SI Appendix 6 & 7***. These assumptions are relaxed in individual-based simulations, where we also allow for the generation of new allelic variants through recombination (see SI Appendix 5).

Results

Our analytical derivations show that initially the allelic value evolves towards an allele whose MHC molecule offers protection against all pathogens. This specific allelic value, labelled $\boldsymbol{x}^*$, is given by the arithmetic mean of the vectors describing the m pathogens $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m$ (***Figure 1***; ***SI Appendix 7.1***). We refer to $\boldsymbol{x}^*$ as the generalist allele. This allele maximises the condition c for homozygote genotypes (grey lines and cones in ***Figure 1***).

We now present in two steps how evolution proceeds after the generalist allele $\boldsymbol{x}^*$ has been reached. First, we explain how an allelic dimorphism can emerge. Second, we proceed by describing how after this initial diversification allelic diversity can increase further and reach extraordinarily high levels that are maintained by heterozygote advantage.

When pathogens are sufficiently similar such that the distance d between the m pathogens is less than $2\sqrt{m}$ (as in ***Figure 1A, C***), then the generalist allele $\boldsymbol{x}^*$ cannot be invaded by nearby mutant alleles and is therefore an endpoint of evolution (***Figure S1***). For the derivation of this condition see ***SI Appendix 7.5***. The reason is in this case the MHC molecule coded by $\boldsymbol{x}^*$ has a relatively high efficiency against all pathogens. If, however, the pathogens are more different such that $d > 2\sqrt{m}$ (as in ***Figure 1B, D***), then the MHC molecule coded by the generalist allele $\boldsymbol{x}^*$ has a relatively low efficiency against all pathogens and any nearby mutant allele $\boldsymbol{x}'$ can invade and coexist with the generalist allele $\boldsymbol{x}^*$. In this case, the heterozygote genotype $(\boldsymbol{x}', \boldsymbol{x}^*)$ provides better protection against a broader spectrum of pathogens – and therefore a higher survival s – than either homozygote genotype. Hence, the allelic dimorphism is maintained by heterozygote advantage, $s(\boldsymbol{x}^*, \boldsymbol{x}') > s(\boldsymbol{x}^*, \boldsymbol{x}^*) > s(\boldsymbol{x}', \boldsymbol{x}')$, as suggested by ***Doherty and Zinkernagel*** (***1975***). From this dimorphism a new mutant allele can invade if it is more specialised than the allele from which it originated. General theory predicts that such a successful mutant allele also replaces its parent allele (***Dercole and Rinaldi, 2008***, Appendix B). In summary, two allelic lineages emerge from the generalist allele $\boldsymbol{x}^*$ and subsequently diverge (***Figure 2***). Thus, $\boldsymbol{x}^*$ is an evolutionary branching point in the sense of adaptive dynamics theory (***Metz et al., 1992***; ***Geritz et al., 1998***).

Increasing the difference between the two alleles present in such a dimorphism has two opposing effects. The condition and thereby the survival of the heterozygote genotype increases because the MHC molecules of the two more specialised alleles provide increasingly better protection against complementary sets of pathogens (divergent allele advantage in the sense of ***Wakeland***

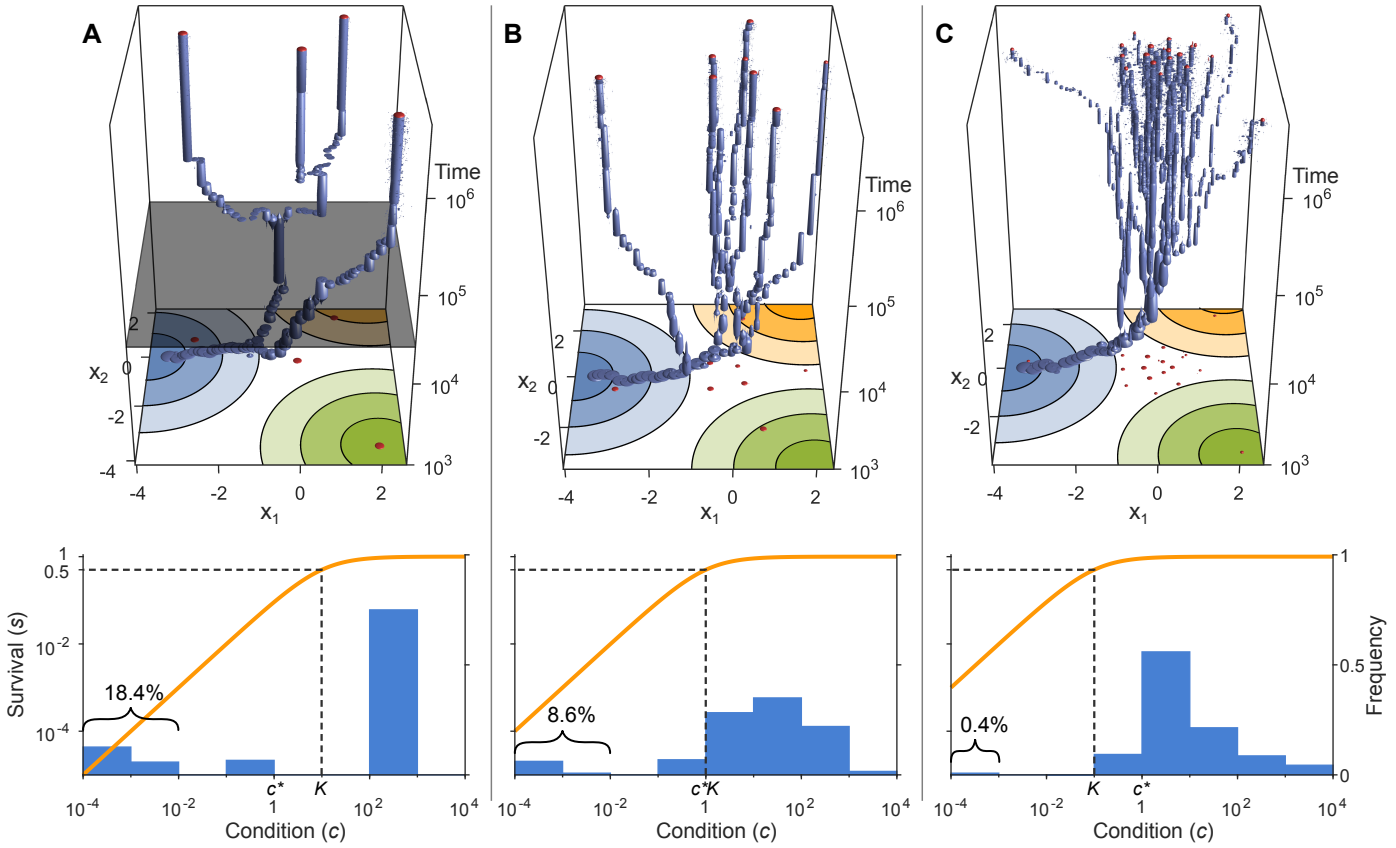

Figure 2. Evolution of allelic values in the presence of three pathogens (arranged as in ***Figure 1D***) for three different values of the survival half-saturation constant K (***A***: $K = 10$, ***B***: $K = 1$, ***C***: $K = 0.1$; dashed line in lower panel). The top panel shows individual-based simulations. The two horizontal axes give the two allelic values $\boldsymbol{x} = (x_1, x_2)$ that characterises an allele while the vertical axes shows evolutionary time. The thickness of the blue tubes is proportional to allele frequencies. Allelic values at the last generation are projected as red dots on the top as well as on the bottom plane. The coloured circles represent the contour lines of the Gaussian efficiency functions $e(\boldsymbol{p}_k - \boldsymbol{x})$ shown in ***Figure 1D***. The bottom panel gives survival s as a function of condition c as defined by ***Equation 2*** on a log-log scale (orange line, left vertical axis) and the frequencies of individual conditions at the final generation (blue bars, right vertical axis; conditions from 0 to 10^{-4} are incorporated into the first bar). In all simulations gradual evolution leads toward the generalist allele $\boldsymbol{x}^* = (0, 0)$ and branching occurs in its neighbourhood, as predicted by our analytical derivations (***SI Appendix 7***). In (***A***) there are three consecutive branching events (the second branching event is marked by the grey plane). This leads to the evolution of three specialist alleles and the generalist allele. Random mating results in ten genotypes (four homozygotes and six heterozygotes). The homozygote specialists have a condition and thereby a survival close to zero (two left bars in bottom panel ***A***), the homozygote for the generalist allele x^* has a condition $c^* = 1$ (middle bar) and thereby a survival equalling $1/(K + 1) \approx 0.091$ (***Equation 2***), and the heterozygote genotypes have a survival close to 1 (right bar). (***B***) and (***C***) show an increasing number of branching events resulting in increasing numbers of coexisting alleles. The lower panel shows that the increasing number of alleles results in a lower proportion of homozygotes with low survival (left most bars in lower panels). Mutations are drawn from a rotational symmetric normal distribution with expected step size $\delta = 0.11$. Other parameter values: $d = 7$, per-capita mutation probability $\mu = 5 \times 10^{-7}$, population size $N = 2 \times 10^5$.

et al. (***1990***)). On the other hand, survival of the two homozygote genotypes decreases because they become increasingly vulnerable for the set of pathogens for which their MHC molecules do not offer protection. Note that, due to random mating and assuming equal allele frequencies, half of the population are high survival heterozygotes and the remaining half homozygotes with low survival. Since survival is a saturating function of condition c (Eq. 2) the increase in survival of heterozygotes slows down with increasing condition (plateau of the orange curves in ***Figure 2***) so that the two opposing forces eventually balance each other and divergence comes to a halt.

At this point, one of the allelic lineages branches again resulting in three coexisting alleles. As a result, the proportion of low-survival homozygotes decreases, assuming equal allele frequencies, from one-half to one-third. Subsequently, the coexisting alleles diverge further from each other because the increase in heterozygote survival once again outweighs the decreased survival of

the (now less frequent) homozygotes (see ***Figure 2***A at time $t = 3 \times 10^4$, grey plane). This process of evolutionary branching and allelic divergence repeats itself so that allelic diversity continues to increase while simultaneously the proportion of vulnerable homozygote genotypes decreases (***Figure 2***). The number of consecutive branching events increases with the number of pathogens m and their dissimilarity d, but also depends on the survival half-saturation constant K (Eq. 2). ***Figure 3*** presents the final number of coexisting alleles for different model parameters derived from individual-based simulations. It shows that for a large part of the parameter space more than 100 (solid contour lines in ***Figure 3***) and up to over 200 alleles can emerge and coexist.

We perform several robustness checks for our results. First, ***Figure S2*** shows simulations in which we vary the expected mutational step size. These simulations show that the gradual build-up of diversity occurs most readily when the mutational step size is neither very small, since then the evolutionary dynamics becomes exceedingly slow, nor very large, since then the self-organised assembly of alleles becomes dominated by a stochastic assembly that is less efficient in finding the maximum set of alleles able to coexist. Second, the results presented in ***Figure 2*** and ***3*** are based on the assumptions of equally spaced pathogen optima and equal width and rotational symmetry of the Gaussian functions $e(\boldsymbol{p} - \boldsymbol{x})$ as shown in ***Figure 1***. In ***SI Appendix 4*** and ***SI Appendix 7*** we present simulation and analytical results, respectively, for the non-symmetric case. ***Figure S3*** shows

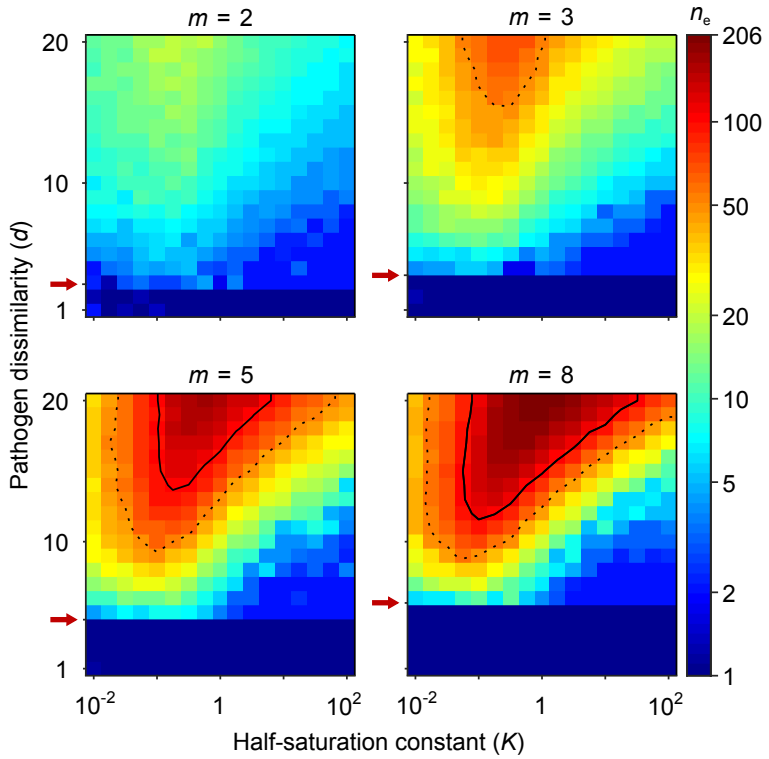

Figure 3. Number of coexisting alleles for m pathogens as a function of pathogen dissimilarity d and the survival half-saturation constant K. Note, that the x-axis can be translated into survival $s(\boldsymbol{x}^*, \boldsymbol{x}^*)$ of the generalist genotype using ***Equation 2***, which then varies between 0.01 for $K = 10^{-2}$ and 0.99 for $K = 10^2$. Results are based on a single individual-based simulation per pixel. These were run for 10^6 generations assuring that the equilibrium distribution of alleles is reached. The clear pattern in the figures indicates a high degree of determinism in the simulations. Results are reported in terms of the effective number of alleles n_e, which discounts for alleles arising from mutation-drift balance (see ***Appendix 1***). Note that for our demographic parameters the expected n_e under mutation-drift balance alone equals 1.2. Dashed and solid lines give the contours for $n_e = 50$ and $n_e = 100$, respectively. Red arrows indicate $d = 2\sqrt{m}$, the threshold for polymorphism to emerge from branching. Accordingly, simulations in the dark blue area result in a single abundant allele with n_e close to one. Other parameters: population size $N = 2 \times 10^5$, per-capita mutation probability $\mu = 5 \times 10^{-7}$, expected mutational step size $\delta = 0.024d$.

that the predictions presented here are qualitatively robust against deviations from symmetry. Finally, in ***SI Appendix 5*** we present the results for the model in which we allow for recombination between the MHC allele's ten different regions. We find that recombination such as gene conversion generally accelerates the build-up of genetic diversity but does not affect the final number of alleles maintained by heterozygote advantage. This can be seen from ***Figure S4***, which is qualitatively similar to ***Figure 3***.

Discussion

Heterozygote advantage as an explanation for the coexistence of a large number of alleles at a single locus has a long history in evolutionary genetics. ***Kimura and Crow*** (***1964***) and subsequently ***Wright*** (***1966***) showed that it is in principle possible that heterozygote advantage can result in the coexistence of an arbitrary number of alleles at a single locus if two conditions are met: (1) all heterozygotes have a similarly high fitness, and (2) all homozygotes have a similarly low fitness. One special class of genes fulfilling these assumptions are highly polymorphic self-incompatibility genes where mating partners need to carry different alleles for fertilisation to be successful (***Wright, 1939***; ***Castric and Vekemans, 2004***). However, more generally these conditions were deemed unrealistic by Kimura, Crow and Wright themselves. This assessment was subsequently confirmed by ***Lewontin et al.*** (***1978***) who investigated a model in which all heterozygotes are more fit than all homozygotes but where the exact fitnesses are determined by drawing random numbers. They found that the proportion of fitness arrays that leads to a stable equilibrium of more than six or seven alleles is vanishingly small. Similarly, the idea that the high allelic diversity found at MHC loci can be explained by heterozygote advantage was initially accepted by theoreticians (e.g. ***Maruyama and Nei, 1981***; ***Takahata and Nei, 1990***), but several later authors studying models based on more mechanistic assumptions were unable to reliably predict the coexistence of significantly more than ten alleles (***Spencer and Marks, 1988***; ***Hedrick, 2002***; ***de Boer et al., 2004***; ***Borghans et al., 2004***; ***Stoffels and Spencer, 2008***; ***Trotter and Spencer, 2008***, ***2013***; ***Ejsmond and Radwan, 2015***; ***Lau et al., 2015***). Thus, currently heterozygote advantage is largely dismissed as an explanation for highly polymorphic loci (***Gould et al., 2004***; ***Eizaguirre and Lenz, 2010***; ***Lenz, 2011***; ***Hedrick, 2012***).

Our study shows that a high number of alleles, matching those found at MHC loci in natural populations, can indeed arise in an evolutionary process driven by heterozygote advantage. Thus, our results revive the idea that heterozygote advantage has the potential to explain extraordinary allelic diversity. Importantly, and in contrast to several of the above mentioned studies, this is achieved without making specific assumptions about homozygote and heterozygote fitnesses. Instead, alleles that are able to coexist emerge through evolutionary branching and self-organise through adaptive evolution such that at all times during this gradual build-up of diversity the two conditions formulated by Crow and Kimura are fulfilled. Specifically, divergence of coexisting alleles is driven by increased heterozygote advantage. Since, as long as K is not too large the survival of heterozygotes quickly saturates with increasing condition (orange line in ***Figure 2*** for $c > K$), all heterozygote genotypes consisting of alleles that are not very similar to each other have survival close to 1. Thus, alleles evolve to fulfil condition (1) stated above. Furthermore, in the presence of several pathogens (m large) of sufficient difference (d large) survival of homozygote individuals that do not carry the generalist allele x^*, or alleles very similar to x^*, approaches zero (grey line and cone in ***Figure 1B, D***). Thus, divergence of allelic values also results in condition (2) of Crow and Kimura being fulfilled by the majority of homozygotes. Further build-up of allelic diversity is now favoured as increasing the number of alleles decreases the proportion of low-fitness homozygotes.

However, for K very small the number of coexisting alleles decreases again (left side in the panels in ***Figure 3***). This is because then survival decreases only marginally with decreasing condition and even individuals homozygote for alleles that are different form the generalist allele x^* can have high survival, violating condition (2) above.

Several other mechanisms are currently invoked to explain diversity at MHC loci. Negative frequency-dependent selection arising from Red-Queen dynamics, fluctuating selection and mating

patterns can all favour polymorphism at MHC loci (***Apanius et al., 1997***; ***Hedrick, 1999***; ***Penn, 2002***; ***Borghans et al., 2004***; ***Wegner, 2008***; ***Spurgin and Richardson, 2010***; ***Loiseau et al., 2011***; ***Ejsmond and Radwan, 2015***; ***Lighten et al., 2017***). In addition, neutral diversity can be enhanced by recombination (***Klitz et al., 2012***; ***Linnenbrink et al., 2018***; ***Robinson et al., 2017***). If many individuals are heterozygous the particularly high levels of gene conversion found at MHC genes can be effective in creating new allelic variants. For instance, for urban human populations with a large effective population size of $N_e = 10^6$ and a per-capita gene conversion probability of $r = 10^{-4}$ an effective number of alleles as high as $n_e = 1 + 4rN_e = 401$ can theoretically be maintained by gene conversion (***Klitz et al., 2012***). However, it is important to point out that recombination such as gene conversion relies on some genetic polymorphism due to balancing selection for it to be efficient at creating new allelic variants.

We do not claim that these mechanisms do not play an important role in maintaining allelic diversity at MHC loci. Rather, our results show that, contrary to the currently widespread view, heterozygote advantage should not be dismissed as a potent force. In any real system, these different mechanisms will jointly affect the allelic diversity at MHC loci. For instance, ***Lighten et al.*** (***2017***) present a model in which for Red-Queen co-evolution to maintain allelic polymorphism heterozygote advantage in the form of divergent allele advantage (***Wakeland et al., 1990***) seems to be a necessary ingredient. Similarly, ***Borghans et al.*** (***2004***) show that allowing for pathogen co-evolution can further increase the number of co-existing alleles compared to heterozygote advantage alone.

Here, we focus at evolution at a single MHC locus while many vertebrates have several MHC loci with similar functions (***Wegner, 2008***; ***Eizaguirre and Lenz, 2010***; ***Spurgin and Richardson, 2010***). The diversity generating mechanism described here still applies if the different loci are responsible for largely non-overlapping sets of pathogens.

Our model does not only apply to MHC genes but to any gene that affects several functions important for survival. Examples are genes that are expressed in different ontogenetic stages or different tissues with competing demands on the optimal gene product. In principle, gene duplication will reduce the potential number of coexisting alleles per locus and eventually leads to a situation where the number of duplicates equals the number of functions (***Proulx and Phillips, 2006***). Under this scenario, the high degree of polymorphism reported here would be transient. However, for MHC genes evidence exist that other forces limit the number of MHC loci (***Penn, 2002***; ***Wegner, 2008***; ***Eizaguirre and Lenz, 2010***; ***Spurgin and Richardson, 2010***) indicating that the mechanism presented here can in principle explain the high number of coexisting MHC alleles.

Appendix 1

Here we provide the calculations for the effective number of alleles n_e given in ***Figure 3***. The expected homozygosity G under mutation-drift balance in a population of size N and per-capita mutation probability μ at equilibrium is given by

$$G = \frac{1}{1 + 4N\mu} \tag{3}$$

(***Gillespie, 2004***). The effective number of alleles n_e under mutation drift balance is defined as $n_e = 1/G = 1 + 4N\mu$. With a population size of $N = 10^5$ and per capita per generation mutation probability $\mu = 5 \times 10^{-7}$ as used in all our simulations $n_e \approx 1.2$. Hence, values of n_e significantly higher than 1.2 indicate the presence of alleles maintained by balancing selection.

Author contributions

Conceptualisation: M.S. and C.R.; Methodology: M.S. and C.R.; Formal analysis: M.S.; Visualisation: M.S.; Writing: M.S. and C.R.; Supervision: C.R.; Funding Acquisition: C.R.

Acknowledgements

We thank Yvonne Meyer-Lucht and Tobias Lenz for helpful discussions and Göran Arnqvist, Helena Westerdahl, Joachim Hermisson and Sophie Karrenberg for comments on an earlier version of the manuscript.

References

Abbas AK, Lichtman AH, Pillai S. Basic immunology: functions and disorders of the immune system. Elsevier Health Sciences

Allison AC. Protection afforded by sickle-cell trait against subtertian malarial infection. British medical journal. 1954; doi: 10.1136/bmj.1.4857.290.

Apanius V, Penn D, Slev PR, Ruff LR, Potts WK. The nature of selection on the major histocompatibility complex. Critical Reviews™ in Immunology. 1997; doi: 10.1615/critrevimmunol.v17.i2.40.

de Boer RJ, Borghans JA, Van Boven M, Keşmir C, Weissing FJ. Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. Immunogenetics. 2004; doi: 10.1007/s00251-003-0629-y.

Borghans JA, Beltman JB, De Boer RJ. MHC polymorphism under host-pathogen coevolution. Immunogenetics. 2004; doi: 10.1007/s00251-003-0630-5.

Castric V, Vekemans X. Plant self-incompatibility in natural populations: a critical assessment of recent theoretical and empirical advances. Molecular Ecology. 2004; doi: 10.1111/j.1365-294x.2004.02267.x.

Chappell PE, Meziane EK, Harrison M, Magiera Ł, Hermann C, Mears L, Wrobel AG, Durant C, Nielsen LL, Buus S, et al. Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. Elife. 2015; doi: 10.7554/elife.05345.

Cortazar-Chinarro M, Meurling S, Schroyens L, Siljestam M, Ritcher-Boix A, Laurila A, Höglund J. Latitudinal MHC variation and haplotype associated differential survival in response to experimental infection of two strains of Batrachochytrium dendrobatitis (Bd-GPL) in common toads. bioRxiv. 2019; doi: 10.1101/597559.

Dercole F, Rinaldi S. Analysis of evolutionary processes: the adaptive dynamics approach and its applications: the adaptive dynamics approach and its applications. Princeton University Press

Doherty PC, Zinkernagel RM. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature. 1975; doi: 10.1038/256050a0.

Duncan WR, Wakeland EK, Klein J. Histocompatibility-2 system in wild mice. Immunogenetics. 1979; doi: 10.1007/bf01570420.

Eizaguirre C, Lenz T. Major histocompatibility complex polymorphism: dynamics and consequences of parasite-mediated local adaptation in fishes. Journal of Fish Biology. 2010; doi: 10.1111/j.1095-8649.2010.02819.x.

Eizaguirre C, Lenz TL, Kalbe M, Milinski M. Divergent selection on locally adapted major histocompatibility complex immune genes experimentally proven in the field. Ecology letters. 2012; doi: 10.1111/j.1461-0248.2012.01791.x.

Ejsmond MJ, Radwan J. Red Queen processes drive positive selection on major histocompatibility complex (MHC) genes. PLoS Comput Biol. 2015; doi: 10.1371/journal.pcbi.1004627.

Fisher RA. The genetical theory of natural selection. Oxford University Press

Froeschke G, Sommer S. MHC class II DRB variability and parasite load in the striped mouse (Rhabdomys pumilio) in the southern Kalahari. Molecular biology and evolution. 2005; doi: 10.1093/molbev/msi156.

Froeschke G, Sommer S. Insights into the complex associations between MHC class II DRB polymorphism and multiple gastrointestinal parasite infestations in the striped mouse. PLoS One. 2012; doi: 10.1371/journal.pone.0031820.

Geritz SAH, Kisdi É, Meszéna G, Metz JAJ. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evolutionary Ecology. 1998; doi: 10.1023/a:1006554906681.

Gillespie JH. Population Genetics: A Concise Guide. The Johns Hopkins University Press

Gould SJ, Hildreth JE, Booth AM. The evolution of alloimmunity and the genesis of adaptive immunity. The Quarterly review of biology. 2004; doi: 10.1086/426088.

Hedrick PW. Balancing selection and MHC. Genetica. 1999; doi: 10.1023/A:1026494212540.

Hedrick PW. Pathogen resistance and genetic variation at MHC loci. Evolution. 2002; doi: 10.1554/0014-3820(2002)056[1902:pragva]2.0.co;2.

Hedrick PW. What is the evidence for heterozygote advantage selection? Trends in ecology & evolution. 2012; doi: 10.1016/j.tree.2012.08.012.

Hughes AL, Nei M. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proceedings of the National Academy of Sciences. 1989;

Jeffery KJ, Bangham CR. Do infectious diseases drive MHC diversity? Microbes and Infection. 2000; doi: 10.1016/s1286-4579(00)01287-9.

Kekäläinen J, Vallunen JA, Primmer CR, Rättyä J, Taskinen J. Signals of major histocompatibility complex overdominance in a wild salmonid population. Proceedings of the Royal Society of London B: Biological Sciences. doi: 10.1098/rspb.2009.0727.

Kimura M, Crow JF. The number of alleles that can be maintained in a finite population. Genetics. 1964;

Klitz W, Hedrick P, Louis EJ. New reservoirs of HLA alleles: pools of rare variants enhance immune defense. Trends in Genetics. 2012; doi: 10.1016/j.tig.2012.06.007.

Lau Q, Yasukochi Y, Satta Y. A limit to the divergent allele advantage model supported by variable pathogen recognition across HLA-DRB1 allele lineages. Tissue Antigens. 2015; doi: 10.1111/tan.12667.

Lenz TL. Computational prediction of MHC II-antigen binding supports divergent allele advantage and explains trans-species polymorphism. Evolution. 2011; doi: 10.1111/j.1558-5646.2011.01288.x.

Lewontin R, Ginzburg L, Tuljapurkar S. Heterosis as an explanation for large amounts of genic polymorphism. Genetics. 1978;

Lighten J, Papadopulos AS, Mohammed RS, Ward BJ, Paterson IG, Baillie L, Bradbury IR, Hendry AP, Bentzen P, Van Oosterhout C. Evolutionary genetics of immunological supertypes reveals two faces of the Red Queen. Nature communications. 2017; doi: 10.1038/s41467-017-01183-2.

Linnenbrink M, Teschke M, Montero I, Vallier M, Tautz D. Meta-populational demes constitute a reservoir for large MHC allele diversity in wild house mice (Mus musculus). Frontiers in zoology. 2018; doi: 10.1186/s12983-018-0266-9.

Loiseau C, Zoorob R, Robert A, Chastel O, Julliard R, Sorci G. Plasmodium relictum infection and MHC diversity in the house sparrow (Passer domesticus). Proceedings of the Royal Society of London B: Biological Sciences. 2011;

Maruyama T, Nei M. Genetic variability maintained by mutation and overdominant selection in finite populations. Genetics. 1981;

McClelland EE, Penn DJ, Potts WK. Major histocompatibility complex heterozygote superiority during coinfection. Infection and immunity. 2003; doi: 10.1128/iai.71.4.2079-2086.2003.

Metz JAJ, Nisbet RM, Geritz SAH. How should we define 'fitness' for general ecological scenarios? Trends in Ecology & Evolution. 1992;

Oliver M, Telfer S, Piertney S. Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (Arvicola terrestris). Proceedings of the Royal Society of London B: Biological Sciences. 2009; doi: 10.1098/rspb.2008.1525.

Penn DJ. The scent of genetic compatibility: sexual selection and the major histocompatibility complex. Ethology. 2002; doi: 10.1046/j.1439-0310.2002.00768.x.

Proulx SR, Phillips PC. Allelic divergence precedes and promotes gene duplication. Evolution. 2006; doi: 10.1554/05-507.1.

Robinson J, Guethlein LA, Cereb N, Yang SY, Norman PJ, Marsh SG, Parham P. Distinguishing functional polymorphism from random variation in the sequences of> 10,000 HLA-A,-B and-C alleles. PLoS Genetics. 2017; doi: 10.1371/journal.pgen.1006862.

Sellis D, Kvitek DJ, Dunn B, Sherlock G, Petrov DA. Heterozygote advantage is a common outcome of adaptation in Saccharomyces cerevisiae. Genetics. 2016; doi: 10.1534/genetics.115.185165.

Sommer S. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Frontiers in Zoology. 2005; doi: 10.1186/1742-9994-2-16.

Spencer HG, Marks RW. The maintenance of single-locus polymorphism. I. Numerical studies of a viability selection model. Genetics. 1988;

Spurgin LG, Richardson DS. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proceedings of the Royal Society of London B: Biological Sciences. 2010; doi: 10.1098/rspb.2009.2084.

Stoffels RJ, Spencer HG. An asymmetric model of heterozygote advantage at major histocompatibility complex genes: degenerate pathogen recognition and intersection advantage. Genetics. 2008; doi: 10.1534/genetics.107.082131.

Takahata N, Nei M. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics. 1990;

Trotter MV, Spencer HG. The generation and maintenance of genetic variation by frequency-dependent selection: constructing polymorphisms under the pairwise interaction model. Genetics. 2008; doi: 10.1534/genetics.108.088880.

Trotter MV, Spencer HG. Models of frequency-dependent selection with mutation from parental alleles. Genetics. 2013; doi: 10.1534/genetics.113.152496.

Wakeland EK, Boehme S, She JX, Lu CC, McIndoe RA, Cheng I, Ye Y, Potts WK. Ancestral polymorphisms of MHC class II genes: divergent allele advantage. Immunologic Research. 1990; doi: 10.1007/bf02918202.

Wegner K. Historical and contemporary selection of teleost MHC genes: did we leave the past behind? Journal of Fish Biology. 2008; doi: 10.1111/j.1095-8649.2008.02051.x.

Wright S. Evolution in Mendelian populations. Genetics. 1931; doi: 10.1007/bf02459575.

Wright S. The distribution of self-sterility alleles in populations. Genetics. 1939;

Wright S. Polyallelic random drift in relation to evolution. Proceedings of the National Academy of Sciences. 1966; doi: 10.1073/pnas.55.5.1074.

Supplementary Information for

Heterozygote advantage can explain the extraordinary diversity of immune genes

Mattias Siljestam* and Claus Rueffler

***Corresponding author: mattias.siljestam@ebc.uu.se**

Contents

1 Supplementary figure S1: Evolutionary dynamics without branching **2**

2 Supplementary figure S2: Varying the mutational step size **3**

3 Table of mathematical notation **4**

4 Deviations from symmetry **5**
4.1 Random placement of pathogen vectors . . . 5
4.2 Random covariance matrices for the pathogen efficiencies . . . 5

5 Model with recombination **7**

6 Mathematical analysis: Preliminaries **9**
6.1 Adaptive dynamics and invasion fitness . . . 9
6.2 Model description . . . 10

7 Analytical results **11**
7.1 The evolutionarily singular point . . . 11
7.2 Derivation of the Hessian matrix . . . 12
7.3 Derivation of the Jacobian matrix . . . 13
7.4 Special case: Identically shaped Gaussian efficiency functions . . . 15
7.5 Special case: Maximal symmetry . . . 15

1. Supplementary figure S1: Evolutionary dynamics without branching

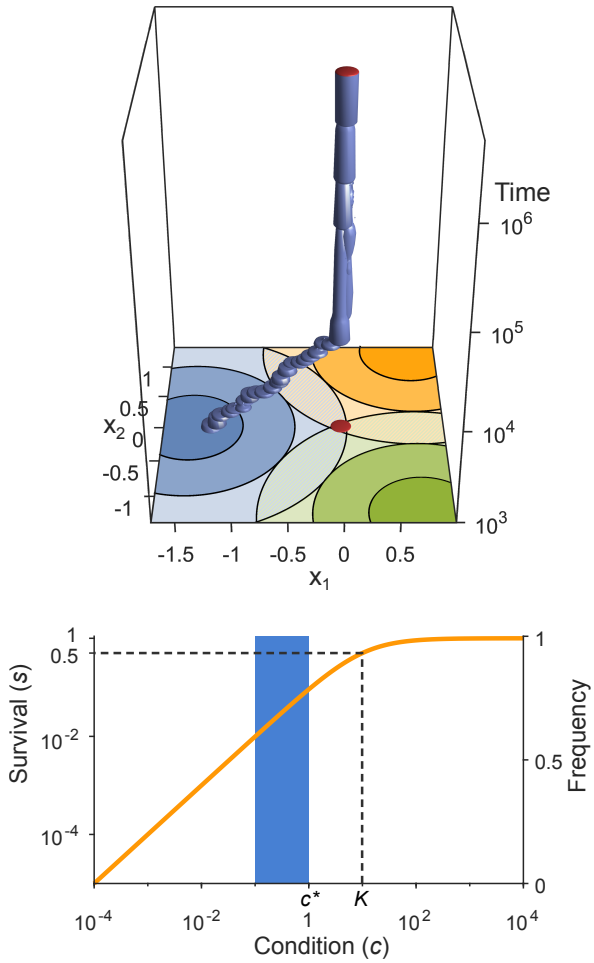

Fig. S1. Evolution of allelic values in the presence of three pathogens. This figure is analogous to ***Figure 2*** but the distance between pathogen vectors equals $d = 2.5$ (as in ***Figure 1C***) while in ***Figure 2*** this distance is much larger ($d = 7$, as in ***Figure 1D***). As a consequence, the condition for evolutionary branching ($d > 2\sqrt{m}$) is not fulfilled and the evolutionary dynamics result in a monomorphic population consisting essentially of only the generalist allele $\boldsymbol{x}^* = (0, 0)$. This result is independent of the half-saturation constant K, here chosen to be $K = 10$. The lower panel shows that the generalist allele x^*, which is the final outcome of the evolutionary dynamics, has condition $c^* \approx 1$ (blue bar) and thereby survival $s \approx 1/(K+1) = 0.\overline{09}$ (***Equation S8***). Note that while in ***Figure 2A*** the configuration $c^* < K$ resulted in three branching events leading to four alleles this does not occur here since the condition for the first branching is not fulfilled. Mutations are drawn from a rotational symmetric normal distribution with expected step size $\delta = 0.11$. Other parameter values: per-capita mutation probability $\mu = 5 \times 10^{-7}$, population size $N = 2 \times 10^5$.

2. Supplementary figure S2: Varying the mutational step size

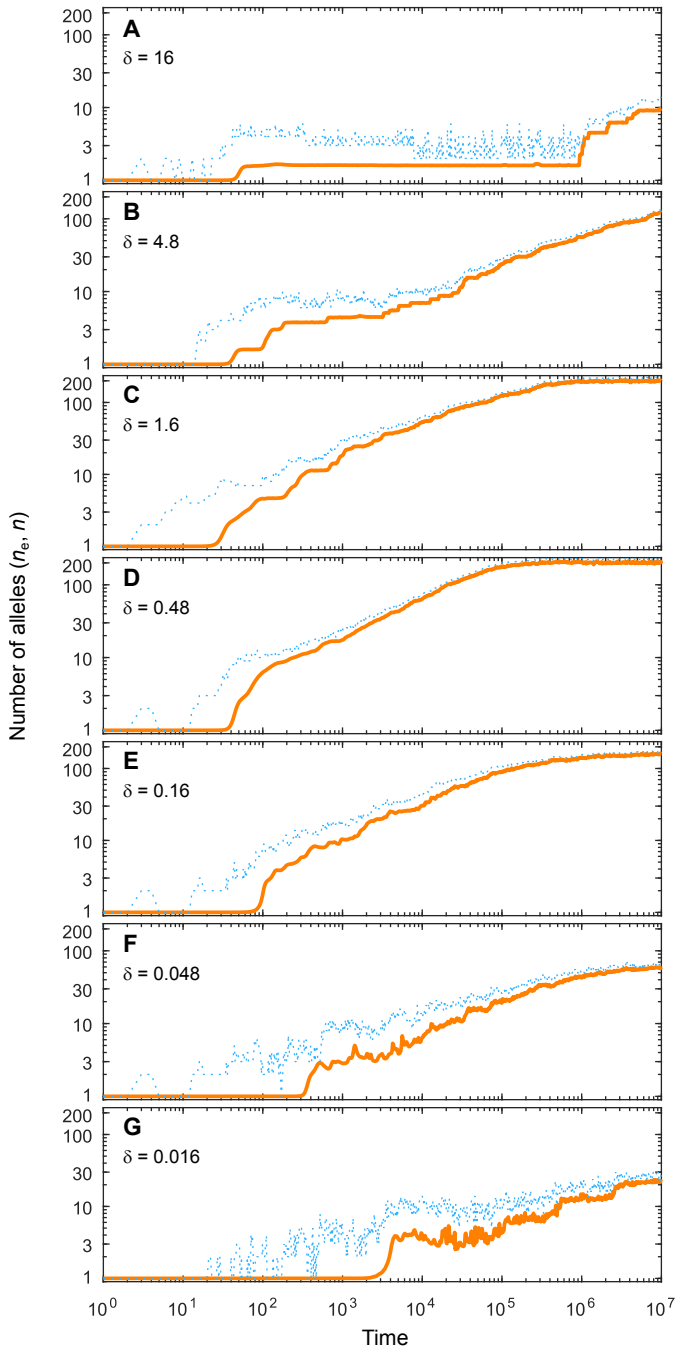

Fig. S2. Number of coexisting alleles as they emerge in individual-based simulations for different expected mutational step sizes δ and eight pathogens ($m = 8$). Parameters are chosen such that up to 200 alleles can evolve ($K = 0.5$, $d = 20$; see bottom right panel in ***Figure 3*** in the main text). Orange solid lines and blue dotted lines give the effective number n_e and the absolute number n of alleles, respectively. The number of alleles increases fastest and saturates earliest for an intermediate mutational step size of $\delta = 0.48$ (***D***; pathogen vectors are $d/\delta = 20/0.48 \approx 42$ average mutational steps apart) as used in ***Figure 3***. Decreasing the average mutational step size slows down the build-up of allelic diversity (***E***-***G***). In the extreme case shown in ***G*** (pathogen vectors are $20/0.016 \approx 1250$ average mutational steps apart) the evolutionary dynamics is strongly limited by the rate of phenotypic change due to the small step size and the number of alleles after 10^7 time steps has reached only 10% the number reached in ***D***. Increasing the average mutational step size also slows down the build-up of allelic diversity (***A***-***C***). In the extreme case shown in ***A*** (pathogen vectors are 1.25 average mutational steps apart) the evolutionary dynamics are strongly limited by the very large proportion of maladapted mutants and the number of alleles after 10^7 time steps has reached only 5% the number reached in ***D***. Other parameters as in ***Figure 3***: population size $N = 2 \times 10^5$, per-capita mutation probability $\mu = 5 \times 10^{-7}$.

3. Table of mathematical notation

Bold italic font indicates vectors (e.g. $\boldsymbol{x}$) while normal italic font indicates scalars (e.g. x). Capital letter in sans serif font indicate matrices (e.g. H). Below is a list of all mathematical notations used in this supplementary material.

Notation	Explanation
c	condition function
C	mutational variance-covariance matrix
d	distance between equally spaced pathogen vectors
δ	expected mutational step size
$e_k(\boldsymbol{x}) = e(\boldsymbol{p}_k - \boldsymbol{x})$	efficiency function for pathogen k
f_a	frequency of allele $\boldsymbol{x}_a$
h	dimensionality of the allelic trait space
H	Hessian matrix
I	identity matrix
J	Jacobian matrix
k	index for pathogens
K	half-saturation constant of survival function
m	number of pathogens
μ	per-capita per-generation mutation probability
n	number of alleles
n_{e}	effective number of alleles
N	population size
$\boldsymbol{p}_k$	vector describing the kth pathogen
s	survival function
$s_{\max}$	maximum survival
Σ_{k}	Gaussian covariance matrix of the efficiency function e_k (***Equation S10***)
V	variance-covariance matrix of the position of the pathogen vectors
v^2	variance of the position of the pathogen vectors under full symmetry
$\boldsymbol{x}_a$	trait vector of allele a
$\boldsymbol{X}$	vector of n allelic trait vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$
$\boldsymbol{y}$	allelic trait vector of rare mutant allele

4. Deviations from symmetry

Figure 3 in the main text and ***Figure S4*** show the final number of coexisting alleles for different parameter combinations. These results are based on two symmetry assumptions. First, the m point described by the pathogen vectors are placed equidistantly resulting in a regular $(m-1)$-simplex. Second, the multivariate Gaussian functions e_k describing the MHC-molecule's efficiencies against pathogen k are rotational symmetric as shown in ***Figure 1***. Thus, the variance-covariance matrices Σ_k in ***Equation S10*** are equal to the identity matrix I. Here, we test the robustness of the outcome shown ***Figure 3*** with respect to violations of these symmetry assumptions. We focus on the case with eight pathogens. The results are summarised in ***Figure S3***. Panel (***A***) is identical to the bottom right panel in ***Figure 3*** and shown here for comparison. Panels (***B-D***) show the final number of coexisting alleles for increasing deviations from symmetry as explained in the following. Note that each pixel is derived from a single simulation with a unique random perturbation from symmetry.

In panel (***B***) the assumption of symmetrically placed pathogen vectors is perturbed while the Gaussian functions e_k are kept rotationally symmetric. Section 4.1 describes the procedure how the positions of the pathogen vectors are randomised while controlling for their variance. The similarity between panel (***A***) and (***B***) indicates that deviations from a symmetric placement of pathogen vectors has a minor effect on the number of coexisting alleles. The slightly decreased smoothness of the contours corresponding to 50 and 100 coexisting alleles stems from the fact that each pixel is derived from a unique perturbation. Note that polymorphism can emerge for d-values such that the branching condition $d > 2\sqrt{m}$ derived for the symmetric case is not fulfilled (below the red arrow). This can be understood based on the expression for the Hessian matrix given in ***Equation S37***. This Hessian is more likely to be non-negative definite for asymmetrically placed pathogen vectors.

In ***Figure S3C, D*** we, additionally to the non-symmetric placement of pathogen vectors, allow for non-rotational symmetric Gaussian functions e_k. Thus, Σ_k differs between pathogens and neither needs to be a diagonal matrix nor a matrix with 1's on the diagonal. The variances of the perturbed variance-covariance matrices are drawn from the interval $(1-\varepsilon, 1+\varepsilon)$ and constrained such that the average variance is equals 1. Section 4.2 describes the procedure in detail. Panel (***C***) shows the result for modest ($\varepsilon = 0.2$) and panel (***D***) for strong ($\varepsilon = 0.5$) deviations from rotational symmetry. Comparing panel (***B***) and (***C***) indicates that modest deviations from rotational symmetry still have a relatively minor effect on the final number of alleles. In panel ***D***) configurations exist where significantly fewer alleles are able to coexist. Interestingly, configurations resulting in a high number of alleles are more likely to occur in combination with high K-values. The highly irregular pattern results from each pixel corresponding to single simulation with a unique random perturbation from symmetry. Furthermore, the threshold for polymorphism decreases even more because the Hessian matrix given in ***Equation S5*** is even more likely to be non-negative definite with perturbations in Σ_k.

4.1. Random placement of pathogen vectors.

The distribution of eight pathogen vectors can be described by their seven dimensional variance-covariance matrix V calculated from the coordinates $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_8$. Since each diagonal element of V describes the variance of the pathogen vectors along a different dimension of the trait space the average variance equals $\mathrm{tr}(\mathrm{V})/7$, where $\mathrm{tr}(\mathrm{V})$ denotes the trace. This measure is unaffected by rotation of the points $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_8$. For symmetrically placed pathogen vectors $\mathrm{V}_{\mathrm{sym}} = d^2/(2m)\mathrm{I}$, where I denotes the identity matrix, and therefore $\mathrm{tr}(\mathrm{V}_{\mathrm{sym}}) = d^2(m-1)/(2m)$. In panel (***B-D***) We obtain randomly placed pathogen vectors with $\mathrm{tr}(\mathrm{V}_{\mathrm{per}}) = d^2(m-1)/2m = \mathrm{tr}(\mathrm{V}_{\mathrm{sym}})$ in the following manner. First, eight preliminary points $\hat{\boldsymbol{p}}_1, \ldots, \hat{\boldsymbol{p}}_8$ are placed randomly within a unit 7-sphere. These points are subsequently multiplied with a scalar α given by

$$\alpha = \sqrt{\frac{\mathrm{tr}(\mathrm{V}_{\mathrm{sym}})}{\mathrm{tr}(\hat{\mathrm{V}})}} = \sqrt{\frac{d^2(m-1)}{2m\,\mathrm{tr}(\hat{\mathrm{V}})}}$$

to obtain the final set of pathogen vectors $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_8$. Here, $m = 8$ and $\hat{\mathrm{V}}$ is the variance-covariance matrix calculated based on the preliminary points $\hat{\boldsymbol{p}}_1, \ldots, \hat{\boldsymbol{p}}_8$.

4.2. Random covariance matrices for the pathogen efficiencies.

In ***Figure 1-3*** the variance-covariance matrices Σ_k determining the shape of the efficiency functions e_k are equal to the identity matrix I. In ***Figure S3C, D*** we allow for arbitrary variance-covariance matrices. In order to keep the results between the symmetric and asymmetric case comparable we fix the mean variance over all Σ_k equal to 1. We obtain the eight random variance-covariance matrices $\Sigma_1, \ldots, \Sigma_8$ in the following manner. First, eight random diagonal matrices $\mathrm{D}_1, \ldots, \mathrm{D}_8$

are determined (one per pathogen vectors) with entries drawn from a uniform distribution $U(1-\varepsilon, 1+\varepsilon)$. These matrices are then multiplied with a scalar β given by

$$\beta = \frac{m-1}{\frac{1}{m}\sum_{k=1}^{m} \mathrm{tr}(\mathsf{D}_k)},$$

with $m = 8$, to obtain the set of matrices $\mathsf{M}_1, \ldots, \mathsf{M}_8$ obeying $1 = 1/8 \sum_{i=1}^{8} (\mathrm{tr}(\mathsf{M}_i)/7)$. In a final step we draw eight random rotation matrices $\mathsf{R}_1, \ldots, \mathsf{R}_8$ and calculate our final variance-covariance matrices $\mathsf{P}_1, \ldots, \mathsf{P}_8$ as $\mathsf{P}_k = \mathsf{R}_k \mathsf{M}_k \mathsf{R}_k^{\mathrm{T}}$.

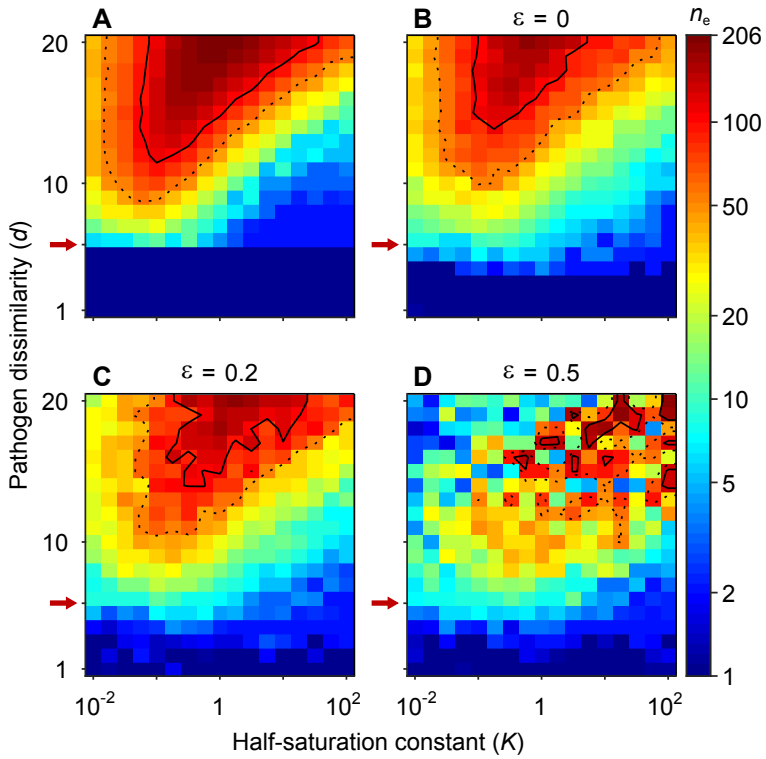

Fig. S3. Number of coexisting alleles for eight pathogens as a function of pathogen dissimilarity d and the half-saturation constant K for symmetrically (***A***) and non-symmetrically placed pathogen vectors (***B-D***). Results are based on a single individual-based simulation per pixel. These were run for 10^6 generations assuring that the equilibrium distribution of alleles is reached. Panel (***A***) shows results for equally spaced pathogen vectors and rotational symmetric functions e_k (***Equation S10***). It is identical to the bottom right panel in ***Figure 3*** and shown for comparison. Panel (***B-D***) show the result for increasing perturbations from symmetry. In panel (***B***) pathogen vectors are placed randomly (see text for details) while the functions e_k are kept rotationally symmetric. In panel (***C***) and (***D***), additionally to the non-symmetric placement of pathogen vectors, the functions e_k are independently perturbed from rotational symmetry (see text for details). In panel (***C***) the deviations from rotational symmetry are moderate while in panel (***D***) they are strong. Note that in panel (***B-D***) pathogen vectors are no longer at a constant distance d. In these panels the y-axis shows the value d that corresponds to symmetrically placed pathogen vectors with a variance equal to the mean variance of the randomly placed pathogen vectors. Results are reported in terms of the effective number of alleles n_e, which discounts for alleles arising from mutation-drift balance (see ***Appendix 1***). Dashed and solid lines give the contours for $n_\mathrm{e} = 50$ and $n_\mathrm{e} = 100$, respectively. Red arrows indicate $d = 2\sqrt{8}$, the minimal value for polymorphism to emerge from branching under full symmetry. Accordingly, simulations in the dark blue area result in a single abundant allele with n_e close to one. Other parameters: population size $N = 2 \times 10^5$, per-capita mutation probability $\mu = 5 \times 10^{-7}$, expected mutational step size $\delta = 0.024d$.

5. Model with recombination

Figure 3 in the main text shows the final number of coexisting alleles based on the symmetry assumptions described in Sections 7.4 and 7.5 and no recombination, $r = 0$. Here we relax the assumption of no recombination and present in ***Figure S4*** simulation results for the per-capita recombination probability $r = 10^5$.

MHC molecules and pathogen vectors are described by ten-dimensional vectors $\boldsymbol{x} = (x_1, x_2, \ldots, x_{10})$ and $\boldsymbol{p} = (p_1, p_2, \ldots, p_{10})$, respectively, as described in the main text. In the absence of recombination a parent with alleles $\boldsymbol{x}$ and $\boldsymbol{y}$ passes on only one of them. With recombination the same parent passes on a recombined allele that originates through crossing over at a randomly chosen break point. For example, a recombination event at the third position results in $(x_1, x_2, x_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10})$. For m pathogens ($m \leq 8$) the $(m-1)$-simplex is embedded in the ten-dimensional trait space according to a random rotation. For each simulation a new random rotation is chosen. This ensures that all ten MHC trait dimensions are of similar importance. Due to the higher dimensionality of the trait-space compared to the $m-1$-dimensional simplex a larger number of mutants lies outside the $m-1$-dimensional simplex and is therefore maladapted. This results in a considerable slow-down of the build-up of allelic diversity. In order to be able to compare the results from simulations with and without recombination we increase the per-capita mutation probability by a factor four from $\mu = 5 \times 10^{-7}$ in ***Figure 3*** to $\mu = 2 \times 10^{-6}$ such that in test runs the same diversity is reached for the ten-dimensional trait space without recombination as for the $(m-1)$-dimensional trait spaces as shown in ***Figure 3***.

Comparing ***Figure S4*** with ***Figure 3*** shows that recombination does not qualitatively change the final number of alleles. Nevertheless, recombination has some effects due to the increased influx of new alleles. First, once allelic polymorphism evolves the further build-up of genetic diversity is accelerated by recombination. Second, when polymorphism is not selected for, i.e., when the branching condition is not fulfilled ($d < 2\sqrt{m}$, below the red arrow in ***Figure S4***) a higher level of neutral polymorphism is reached.

Third, the highest number of coexisting alleles in ***Figure S4*** is $n_{\mathrm{e}} = 236$ compared to $n_{\mathrm{e}} = 206$ in ***Figure 3***. It is informative to discuss this difference in slightly more detail. In an infinite population in which all heterozygotes have equally high fitness and all homozygotes have equally low fitness an infinite number of alleles can coexist. In real populations the maximum number of coexisting alleles is limited by population size and by extinctions due to drift. The maximum number of alleles that can coexist in finite populations has been explored by Kimura and Crow (13) and Wright (21). From Kimura and Crow (13) we estimate that for the parameter values in ***Figure 3*** ($N = 4 \times 10^5$ and $\mu = 5 \times 10^-7$) this maximum number is ≈ 250. The maximum number in our simulations, 206 in the case without recombination, is lower because in our model not all homozygotes and not all heterozygotes have exactly the same fitness, (see ***Figure 3***, bottom panel), which violates the assumptions of the theory of Kimura and Crow (13). We therefore expect that the maximum number of alleles in ***Figure 3*** is at the theoretical maximum and therefore limited by demographic processes.

Once an appreciable number of alleles has evolved recombination creates new alleles just like mutation. Based on this twenty-fold increase in the influx of new allelic variants due to recombination we estimate based on Kimura and Crow (13) that the theoretically maximum number alleles is ≈ 340. This is considerably higher than 236, the maximum number in our simulations that include recombination (***Figure S4***). From this we draw two conclusions. First, the increase of alleles from 206 to 236 can be explained by the increased influx of new alleles. Second, 236 is the true maximum number of alleles that can be maintained by heterozygote advantage given the parameters in ***Figure 3*** and ***Figure S4*** as its is unlikely to be capped by demographic processes.

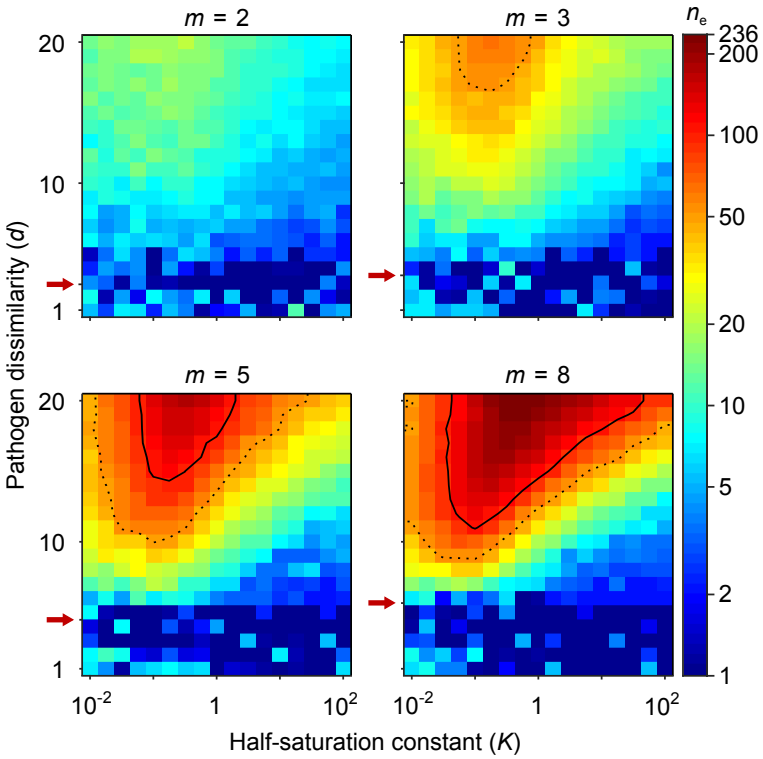

Fig. S4. Number of coexisting alleles for m pathogens as a function of pathogen dissimilarity d and the survival half-saturation constant K for symmetrically placed pathogens and recombination. Results are based on a single individual-based simulation per pixel. These were run for 10^6 generations assuring that the equilibrium distribution of alleles is reached. Results are reported in terms of the effective number of alleles n_e, which discounts for alleles arising from mutation-drift balance (see ***Appendix 1***). Dashed and solid lines give the contours for $n_\mathrm{e} = 50$ and $n_\mathrm{e} = 100$, respectively. Red arrows indicate $d = 2\sqrt{m}$, the threshold for polymorphism to emerge from branching. Accordingly, simulations in the dark blue area below the red arrow result in a single abundant allele with n_e close to one. Higher values of n_e below the red arrow are due the extra influx of new alleles due to recombination. Other parameters: population size $N = 2 \times 10^5$, recombination rate $r = 10^5$, per-capita mutation probability $\mu = 2 \times 10^{-6}$, expected mutational step size $\delta = 0.024d$.

6. Mathematical analysis: Preliminaries

6.1. Adaptive dynamics and invasion fitness.

We analyse the model with an evolutionary invasion analysis using the adaptive dynamics formalism (5, 9, 18). The main tool to predict the evolutionary dynamics is invasion fitness $w(\boldsymbol{y}, \boldsymbol{X})$, which is the expected long-term exponential growth rate of an infinitesimally rare mutant allele with allelic vector $\boldsymbol{y}$ in a resident population at equilibrium consisting of the alleles $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ (16, 18).

We consider a large population consisting of N individuals containing n alleles $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ under Wright-Fisher population dynamics (8, 20). The recurrence equation for allele $\boldsymbol{x}_a$ is given by

$$f_{a,t+1} = f_{a,t} \sum_{b=1}^{n} f_{b,t} \frac{s(\boldsymbol{x}_a, \boldsymbol{x}_b)}{\bar{s}_t}, \qquad \text{[S1]}$$

where $f_{a,t}$ is the frequency of allele $\boldsymbol{x}_a$ at time t, $s(\boldsymbol{x}_a, \boldsymbol{x}_b)$ is the survival of an individual carrying the alleles $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ and

$$\bar{s}_t = \sum_{a=1}^{n} \sum_{b=1}^{n} f_{a,t} f_{b,t} s(\boldsymbol{x}_a, \boldsymbol{x}_b)$$

is the population mean survival at time t. Note, that the sum on the right hand side of ***Equation S1*** describes the marginal fitness of allele $\boldsymbol{x}_a$.

Consider a rare mutant allele $\boldsymbol{y} = \boldsymbol{x}_a + \boldsymbol{\epsilon}$ for some $\boldsymbol{x}_a \in \boldsymbol{X}$ introduced into a resident population of n coexisting alleles $\boldsymbol{X}$ at equilibrium. As long as the mutant allele has a negligible frequency its marginal fitness simplifies to

$$w(\boldsymbol{y}, \boldsymbol{X}) \;=\; \frac{1}{\bar{s}} \sum_{b=1}^{n} f_j s(\boldsymbol{y}, \boldsymbol{x}_b), \qquad \text{[S2]}$$

which is its invasion fitness. Note that because of the equilibrium assumption the time-subscript could be dropped. Allele $\boldsymbol{y}$ has a positive probability to invade and increase in frequency if $w(\boldsymbol{y}, \boldsymbol{X}) > 1$ and disappears otherwise.

The gradient $\nabla w(\boldsymbol{x}_a, \boldsymbol{X})$ of invasion fitness with entries

$$\nabla w(\boldsymbol{x}_a, \boldsymbol{X})_i = \frac{\partial w(\boldsymbol{y}, \boldsymbol{X})}{\partial y_i}\bigg|_{\boldsymbol{y}=\boldsymbol{x}_a} \qquad \text{[S3]}$$

gives the direction in the h-dimensional allelic trait space in which invasion fitness of mutants of allele $\boldsymbol{x}_a$ increases fastest. If mutations occur rarely a mutant allele $\boldsymbol{y}$ will either go extinct or reach an equilibrium frequency before a new mutant appears. If, additionally, $\nabla w(\boldsymbol{x}_a, \boldsymbol{X}) \neq \boldsymbol{0}$ and mutational effects are sufficiently small, i.e., $\boldsymbol{y} = \boldsymbol{x}_a + \boldsymbol{\epsilon}$ for $\boldsymbol{x}_a \in \boldsymbol{X}$ and $\boldsymbol{\epsilon}$ small, then invasion of $\boldsymbol{y}$ implies extinction of $\boldsymbol{x}_a$ (4, Appendix B) resulting in a trait substitution sequence. In the limit of small mutational steps the evolutionary dynamics of an allelic lineage becomes gradual and is given by

$$\frac{\mathrm{d}\boldsymbol{x}_a}{\mathrm{d}t} = \mu f_a N \mathsf{C} \nabla w(\boldsymbol{x}_a, \boldsymbol{X}) \qquad \text{[S4]}$$

(2, 5, 7, 17). Here, μ is the per-capita mutation probability and C the mutational variance-covariance matrix. We note that ***Equation S4*** is structurally similar to the gradient equation of quantitative genetics, which is based on the assumption of weak selection or, equivalently, small genetic variances (1, 3, 12, 14). In this case, $\boldsymbol{x}_a$ characterises the mean of the phenotype distribution, the variance-covariance matrix describes the distribution of the standing genetic variation, and the factor $\mu f_a N$ is replaced with a constant.

Combinations of alleles $\boldsymbol{X}$ such that $0 = \nabla w(\boldsymbol{x}_a, \boldsymbol{X})_i$ for all alleles are of special interest. In the case of a single resident allele $\boldsymbol{X} = \boldsymbol{x}$, the trait vector $\boldsymbol{x}$ is then said to be an evolutionarily singular strategy. If n coexisting alleles all fulfil $0 = \nabla w(\boldsymbol{x}_a, \boldsymbol{X})_i$, then these form an evolutionarily singular coalition. Evolutionarily singular coalitions can be either attractors or repellors of the evolutionary dynamics described by ***Equation S4***. Furthermore, an evolutionarily singular coalition of alleles can be either invadable or uninvadable by nearby mutants. For a single resident allele with a one-dimensional trait $\boldsymbol{x} = x$ a classification of singular strategies is straight forward (9). Evolutionarily singular points that are not approached, irrespective of whether they are invadable or uninvadable, act as repellors and we do not expect to ever find alleles with such values. Evolutionarily singular strategies that are attractors and uninvadable are endpoints of the evolutionary dynamics. Finally, evolutionarily singular points that are attractors and invadable are know as evolutionary branching points. In this case, any nearby mutant can invade the singular strategy and coexist with it in a protected

dimorphism. Further evolution leads to divergence of the alleles present in the dimorphism. Thus, evolutionary branching points are points in trait space at which diversity emerges (19).

The classification of singular points becomes more complicated in multivariate trait spaces or when several strategies coexist in an evolutionarily singular coalition (6, 10, 15). First, in multivariate trait spaces or polymorphic populations, whether or not a singular coalition is an attractor does not only depend on the direction of the fitness gradient in the vicinity of the singular coalition but also on the mutational input (15). Second, in multivariate trait spaces or polymorphic populations attractivity of a singular coalition and invadability are necessary but generally not sufficient conditions for evolutionary branching (10).

In Section 7.1, we show that for our model, as long as the resident population consists of a single allele $\boldsymbol{X} = \boldsymbol{x}$, a unique singular point $\boldsymbol{x}^*$ exists. This allele is uninvadable if it is a minimum of $w(\boldsymbol{y}, \boldsymbol{x}^*)$ as a function of $\boldsymbol{y}$. This is the case if the h-dimensional Hessian matrix H with entries

$$h_{ij} \;=\; \frac{\partial^2 w(\boldsymbol{y}, \boldsymbol{x})}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}. \qquad \text{[S5]}$$

is negative definite (6, 15). In Section 7.2 we derive an explicit expression for H for the fully general case of our model that allows to determine invadability of $\boldsymbol{x}^*$ as a function of the positions of the pathogen vectors, the half-saturation constant K, and the variance-covariance matrices Σ_k that determine the shape of the efficiency functions e_k.

Whether or not the singular point $\boldsymbol{x}^*$ is an attractor of the evolutionary dynamics can be evaluated based on the Jacobian matrix J of the fitness gradient $\nabla w(\boldsymbol{x}^*, \boldsymbol{x}^*)$ (15), which is given by

$$\mathsf{J} \;=\; \mathsf{H} + \mathsf{Q} \qquad \text{[S6]}$$

and where Q is the h-dimensional matrix of mixed derivatives with entries

$$q_{ij} \;=\; \frac{\partial^2 w(\boldsymbol{y}, \boldsymbol{x})}{\partial y_i \partial x_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}. \qquad \text{[S7]}$$

Leimar (15) showed that if the symmetric part of J, i.e., $(\mathsf{J} + \mathsf{J}^{\mathrm{T}})/2$, is negative definite, then the singular point is an attractor of the evolutionary dynamics described by ***Equation S4*** in dependent of the mutational variance-covariance matrix C. Leimar refers to this case as strong convergence stability. In Section 7.3 we show that for our model the Jacobian matrix is in fact a negative definite symmetric matrix. For this case, Leimar proved a stronger result to which he refers to as absolute convergence stability. In this case, all conceivable gradualistic, adaptive paths starting near the point $\boldsymbol{x}^*$ converge to it.

To summarise, for the presented model we prove that a unique singular strategy $\boldsymbol{x}^*$ exists, to which we refer as the generalist allele. This allele is an attractor of any adaptive gradualistic trait substitution sequence. Furthermore, we compute an explicit expression for the Hessian matrix at this point. This expression allows to numerically calculate whether the singular point is invadable or not. We note that the definiteness of the Hessian matrix is independent of the half-saturation constant K. We also make some heuristic arguments about which factors make invadability more likely. For the case of two-dimensional trait spaces results in Geritz *et al.* (10) allow us to conclude that if $\boldsymbol{x}^*$ is invadable, then it is indeed an evolutionary branching point. For trait spaces of dimension three or higher, whether strong convergence and invadability imply evolutionary branching is an open problem (10). Individual-based simulations indicate, however, that for our model this indeed the case.

6.2. Model description.

In this section, we describe the model ingredients. Survival $s(\boldsymbol{x}_a, \boldsymbol{x}_b)$ of a genotype carrying alleles $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ is a saturating function of condition c and described by the well known Michaelis-Menten equation

$$s(\boldsymbol{x}_a, \boldsymbol{x}_b) = \frac{s_{\max} c(\boldsymbol{x}_a, \boldsymbol{x}_b)}{K + c(\boldsymbol{x}_a, \boldsymbol{x}_b)}. \qquad \text{[S8]}$$

Here, the half saturation constant K gives the condition c at which half of the maximum survival is reached and $s_{\max}$ is the maximum survival probability that is approached when c is large.

The condition of a genotype is given by

$$c(\boldsymbol{x}_a, \boldsymbol{x}_b) = c_{\max} \prod_{k=1}^{m} \frac{e_k(\boldsymbol{x}_a) + e_k(\boldsymbol{x}_b)}{2}, \qquad \text{[S9]}$$

where $e_k(\boldsymbol{x})$ is the efficiency of an allele's MHC molecule against pathogen k in an environment with m pathogens. We assume that these efficiencies are traded off and we model this trade-off with Gaussian functions

$$e_k(\boldsymbol{x}) = \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{p}_k)^{\mathrm{T}}\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}-\boldsymbol{p}_k)\right) \qquad \text{[S10]}$$

(see ***Figure 1***). Note, that, with some abuse of notation, we write $e_k(\boldsymbol{x})$ as $e(\boldsymbol{p}_k - \boldsymbol{x})$ in the main part. Each of these multivariate Gaussian functions describes how the efficiency of an allele characterised by the h-dimensional vector $\boldsymbol{x}$ decreases with increasing distance from the pathogen vector $\boldsymbol{p}_k$. The closer an allelic trait vector is to a pathogen vector, the higher is the efficiency of the MHC molecule against that pathogen. The magnitude of the decrease in efficiency with increasing distance is determined by the shape of the Gaussian function given by the h-dimensional variance-covariance matrix $\boldsymbol{\Sigma}_k$.

7. Analytical results

7.1. The evolutionarily singular point.

In this section, we analyse the evolutionary dynamics of a monomorphic resident population ($\boldsymbol{X} = \boldsymbol{x}$) in full generality. Note, that in this case recombination and gene conversion can be ignored. By subsequently applying several symmetry assumptions we then derive the analytical results presented in the main text (see Section 7.4 and 7.5). The invasion fitness of a rare mutant allele $\mathbf{y}$ is given by its marginal fitness and for a monomorphic population ***Equation S2*** simplifies to

$$w(\boldsymbol{y},\boldsymbol{x}) = \frac{s(\boldsymbol{y},\boldsymbol{x})}{s(\boldsymbol{x},\boldsymbol{x})}. \qquad \text{[S11]}$$

Note that $s_{\max}$ cancels out. Since this is also true for the polymorphic case it is omitted from all further calculations. The direction of the evolutionary dynamics is governed by the selection gradient. Its ith entry calculates to

$$\nabla w(\boldsymbol{x},\boldsymbol{x})_i = \frac{\partial w(\boldsymbol{y},\boldsymbol{x})}{\partial y_i} = \frac{1}{s(\boldsymbol{x},\boldsymbol{x})}\frac{\partial s(\boldsymbol{y},\boldsymbol{x})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}}. \qquad \text{[S12]}$$

Using the definitions for s (***Equation S8***) and c (***Equation S9***) and their derivatives

$$\frac{\partial s(\boldsymbol{y},\boldsymbol{x})}{\partial y_i} = \frac{K}{\big(K + c(\boldsymbol{y},\boldsymbol{x})\big)^2}\frac{\partial c(\boldsymbol{y},\boldsymbol{x})}{\partial y_i}. \qquad \text{[S13]}$$

and

$$\frac{\partial c(\boldsymbol{y},\boldsymbol{x})}{\partial y_i} = c(\boldsymbol{y},\boldsymbol{x})\sum_{k=1}^{m}\frac{1}{e_k(\boldsymbol{y})+e_k(\boldsymbol{x})}\frac{\partial e_k(\boldsymbol{y})}{\partial y_i}, \qquad \text{[S14]}$$

where ***Equation S14*** is obtained by applying the generalised product rule

$$\frac{\partial}{\partial x}\left(\prod_{i=1}^{n} f_i(x)\right) = \left(\prod_{i=1}^{n} f_i(x)\right)\sum_{i=1}^{n}\frac{f_i'(x)}{f_i(x)}, \qquad \text{[S15]}$$

we obtain

$$\nabla w(\boldsymbol{x},\boldsymbol{x})_i = \frac{K}{c(\boldsymbol{x},\boldsymbol{x})\big(K + c(\boldsymbol{x},\boldsymbol{x})\big)}\frac{\partial c(\boldsymbol{y},\boldsymbol{x})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}} = \frac{K}{2\big(K + c(\boldsymbol{x},\boldsymbol{x})\big)}\sum_{k=1}^{m}\frac{1}{e_k(\boldsymbol{x})}\frac{\partial e_k(\boldsymbol{y})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}}. \qquad \text{[S16]}$$

In the next step we calculate the derivative of the function $e_k(\boldsymbol{y})$ (***Equation S10***). Applying the chain rule and simplifying results in

$$\begin{aligned}
\frac{\partial e_k(\boldsymbol{y})}{\partial y_i} &= -\frac{1}{2}e_k(\boldsymbol{y})\frac{\partial}{\partial y_i}\left((\boldsymbol{y}-\boldsymbol{p}_k)^{\mathrm{T}}\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{y}-\boldsymbol{p}_k)\right)\\
&= -\frac{1}{2}e_k(\boldsymbol{y})\frac{\partial}{\partial y_i}\sum_{j=1}^{h}\sum_{l=1}^{h}\sigma_{kjl}^{-1}(y_j-p_{kj})(y_l-p_{kl})\\
&= -\frac{1}{2}e_k(\boldsymbol{y})\left(\sum_{j=1}^{h}\sum_{l=1}^{h}\sigma_{kjl}^{-1}(y_l-p_{kl})\frac{\mathrm{d}y_j}{\mathrm{d}y_i} + \sum_{j=1}^{h}\sum_{l=1}^{h}\sigma_{kjl}^{-1}(y_j-p_{kj})\frac{\mathrm{d}y_l}{\mathrm{d}y_i}\right),
\end{aligned} \qquad \text{[S17]}$$

where the entries of the matrix Σ_k^{-1} are denoted by σ_{kjl}^{-1}. Using that $\mathrm{d}y_j/\mathrm{d}y_i = 0$ for $i \neq j$ and $\mathrm{d}y_j/\mathrm{d}y_i = 1$ for $i = j$ this further simplifies to

$$\frac{\partial e_k(\boldsymbol{y})}{\partial y_i} = e_k(\boldsymbol{y}) \sum_{j=1}^{h} \sigma_{k\,il}^{-1}(p_{kj} - y_j). \qquad \text{[S18]}$$

Substituting ***Equation S18*** into ***Equation S16*** finally results in

$$\nabla w(\boldsymbol{x}, \boldsymbol{x})_i \;=\; \frac{K}{2\big(K + c(\boldsymbol{x}, \boldsymbol{x})\big)} \sum_{k=1}^{m} \sum_{j=1}^{h} \sigma_{k\,ij}^{-1}(p_{kj} - x_j) \qquad \text{[S19]}$$

and

$$\boxed{\nabla \boldsymbol{w}(\boldsymbol{x}) \;=\; \frac{K}{2\big(K + c(\boldsymbol{x}, \boldsymbol{x})\big)} \sum_{k=1}^{m} \Sigma_k^{-1}(\boldsymbol{p}_k - \boldsymbol{x})}. \qquad \text{[S20]}$$

As mentioned in 6.1, singular points $\boldsymbol{x}^*$ are allelic trait vectors for which ***Equation S20*** equals zero. From ***Equation S20*** follows that in our model singular points have to fulfil

$$\Big(\sum_{k=1}^{m} \Sigma_k^{-1}\Big)\boldsymbol{x}^* \;=\; \sum_{k=1}^{m} \Sigma_k^{-1} \boldsymbol{p}_k. \qquad \text{[S21]}$$

Solving for $\boldsymbol{x}^*$ yields

$$\boxed{\boldsymbol{x}^* \;=\; \Big(\sum_{k=1}^{m} \Sigma_k^{-1}\Big)^{-1} \sum_{k=1}^{m} \Sigma_k^{-1} \boldsymbol{p}_k := \bar{\boldsymbol{p}}_{\mathrm{w}}}. \qquad \text{[S22]}$$

Thus, the unique singular point $\boldsymbol{x}^*$ equals the arithmetic mean of the pathogen vectors $\boldsymbol{p}_1, \ldots \boldsymbol{p}_m$, each weighted by the inverse of their Gaussian covariance matrices $\Sigma_1, \ldots, \Sigma_m$. For a one dimensional trait space ($h = 1$) this simplifies to

$$x^* \;=\; \frac{\sum_{k=1}^{m} \sigma_k^{-2} p_k}{\sum_{k=1}^{m} \sigma_k^{-2}} =: \bar{p}_{\mathrm{w}}, \qquad \text{[S23]}$$

which is the well-known weighted average for scalars. If $\Sigma_1 = \ldots = \Sigma_m$, then ***Equation S22*** simplifies to the arithmetic mean pathogen vector

$$\boldsymbol{x}^* \;=\; \frac{1}{m} \sum_{k=1}^{m} \boldsymbol{p}_k := \bar{\boldsymbol{p}} \qquad \text{[S24]}$$

as stated in the main text.

7.2. Derivation of the Hessian matrix.

As stated in ***Equation S5***, the entries of the Hessian matrix are given by

$$h_{ij} \;=\; \frac{\partial^2 w(\boldsymbol{y}, \boldsymbol{x}^*)}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} \;=\; \frac{1}{s(\boldsymbol{x}^*, \boldsymbol{x}^*)} \frac{\partial^2 s(\boldsymbol{y}, \boldsymbol{x}^*)}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*}. \qquad \text{[S25]}$$

The second derivative of the function s is obtained by differentiating ***Equation S13*** with respect to y_j resulting in

$$\begin{aligned}
\frac{\partial^2 s(\boldsymbol{y}, \boldsymbol{x}^*)}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} \;&=\; K \frac{\partial}{\partial y_j}\left(\frac{1}{\big(K + c(\boldsymbol{y}, \boldsymbol{x})\big)^2} \frac{\partial c(\boldsymbol{y}, \boldsymbol{x})}{\partial y_i}\right)\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} \\
&=\; \frac{K}{\big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)^4}\left(\frac{\partial^2 c(\boldsymbol{y}, \boldsymbol{x}^*)}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} \big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)^2 - \frac{\partial c(\boldsymbol{y}, \boldsymbol{x}^*)}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} \frac{\partial c(\boldsymbol{y}, \boldsymbol{x}^*)^2}{\partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*}\right) \\
&=\; \frac{K}{\big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)^2} \frac{\partial^2 c(\boldsymbol{y}, \boldsymbol{x}^*)}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*},
\end{aligned} \qquad \text{[S26]}$$

where in the final simplification step we use the conclusion drawn from ***Equation S16*** that $0 = \partial c(\boldsymbol{y}, \boldsymbol{x}^*)/\partial y_i|_{\boldsymbol{y}=\boldsymbol{x}^*}$ and therefore the term after the minus sign disappears.

The second derivative for the function c is obtained by differentiating ***Equation S14*** with respect to y_j resulting in

$$\begin{aligned}
\frac{\partial^2 c(\boldsymbol{y},\boldsymbol{x}^*)}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} &= \frac{\partial}{\partial y_j}\left(c(\boldsymbol{y},\boldsymbol{x}^*)\sum_{k=1}^{m}\left(\frac{1}{e_k(\boldsymbol{y})+e_k(\boldsymbol{x}^*)}\frac{\partial e_k(\boldsymbol{y})}{\partial y_i}\right)\right)\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} \\
&= c(\boldsymbol{x}^*,\boldsymbol{x}^*)\sum_{k=1}^{m}\frac{\partial}{\partial y_j}\left(\frac{1}{e_k(\boldsymbol{y})+e_k(\boldsymbol{x}^*)}\frac{\partial e_k(\boldsymbol{y})}{\partial y_i}\right)\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} \\
&= c(\boldsymbol{x}^*,\boldsymbol{x}^*)\sum_{k=1}^{m}\frac{1}{4e_k(\boldsymbol{x}^*)^2}\left(2\frac{\partial^2 e_k(\boldsymbol{y})}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*}e_k(\boldsymbol{x}^*)-\frac{\partial e_k(\boldsymbol{y})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*}\frac{\partial e_k(\boldsymbol{y})}{\partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*}\right). \qquad \text{[S27]}
\end{aligned}$$

Here, the one but last simplification step again follows from the fact that $0 = \partial c(\boldsymbol{y},\boldsymbol{x}^*)/\partial y_i|_{\boldsymbol{y}=\boldsymbol{x}^*}$.

The second derivative for the function e_k is obtained by differentiating ***Equation S18*** with respect to y_j resulting in

$$\begin{aligned}
\frac{\partial^2 e_k(\boldsymbol{y})}{\partial y_i \partial y_j} &= \frac{\partial}{\partial y_j}e_k(\boldsymbol{y})\sum_{l=1}^{h}\sigma^{-1}_{k\,il}(p_{kl}-y_l) \\
&= \frac{\partial e_k(\boldsymbol{y})}{\partial y_j}\sum_{l=1}^{h}\sigma^{-1}_{k\,il}(p_{kl}-y_l) \;-\; e_k(\boldsymbol{y})\left(\sum_{l=1}^{h}\sigma^{-1}_{k\,il}\frac{\partial y_l}{\partial y_j}\right) \\
&= \frac{1}{e_k(\boldsymbol{y})}\frac{\partial e_k(\boldsymbol{y})}{\partial y_i}\frac{\partial e_k(\boldsymbol{y})}{\partial y_j} - e_k(\boldsymbol{y})\sigma^{-1}_{k\,ij}, \qquad \text{[S28]}
\end{aligned}$$

where the last simplification uses that $\mathrm{d}y_j/\mathrm{d}y_i = 0$ for $i \neq j$ and $\mathrm{d}y_j/\mathrm{d}y_i = 1$ for $i = j$.

By recursively substituting ***Equations S26-S28*** into ***Equation S5*** we obtain

$$\begin{aligned}
h_{ij} &= \frac{K}{K+c(\boldsymbol{x}^*,\boldsymbol{x}^*)}\sum_{k=1}^{m}\frac{1}{4e_k(\boldsymbol{x}^*)^2}\left(\frac{\partial e_k(\boldsymbol{y})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*}\frac{\partial e_k(\boldsymbol{y})}{\partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} - 2e_k(\boldsymbol{x}^*)^2\sigma^{-1}_{k\,ij}\right) \\
&= \frac{K}{4\big(K+c(\boldsymbol{x}^*,\boldsymbol{x}^*)\big)}\sum_{k=1}^{m}\left(\left(\sum_{l=1}^{h}\sigma^{-1}_{k\,il}(p_{kl}-x^*_l)\right)\left(\sum_{l=1}^{h}\sigma^{-1}_{k\,jl}(p_{kl}-x^*_l)\right) - 2\sigma^{-1}_{k\,ij}\right),
\end{aligned}$$

where in the last step we substituted ***Equation S18***. This result can be rewritten as a matrix

$$\mathsf{H} = \frac{K}{4\big(K+c(\boldsymbol{x}^*,\boldsymbol{x}^*)\big)}\left(\sum_{k=1}^{m}\left(\boldsymbol{\Sigma}_k^{-1}\big(\boldsymbol{p}_k-\boldsymbol{x}^*\big)\big(\boldsymbol{p}_k-\boldsymbol{x}^*\big)^{\mathrm{T}}\boldsymbol{\Sigma}_k^{-1}\right) - 2\sum_{k=1}^{m}\boldsymbol{\Sigma}_k^{-1}\right).$$

Finally, substituting $\boldsymbol{x}^*$ with ***Equation S22*** we obtain

$$\boxed{\mathsf{H} = \frac{K}{4\big(K+c(\boldsymbol{x}^*,\boldsymbol{x}^*)\big)}\left(\sum_{k=1}^{m}\left(\boldsymbol{\Sigma}_k^{-1}\big(\boldsymbol{p}_k-\bar{\boldsymbol{p}}_w\big)\big(\boldsymbol{p}_k-\bar{\boldsymbol{p}}_w\big)^{\mathrm{T}}\boldsymbol{\Sigma}_k^{-1}\right) - 2\sum_{k=1}^{m}\boldsymbol{\Sigma}_k^{-1}\right)}. \qquad \text{[S29]}$$

7.3. Derivation of the Jacobian matrix.

From ***Equations S5-S7*** it follows that the entries of the Jacobian matrix J are given by

$$j_{ij} = h_{ij} + q_{ij} = \frac{\partial^2 w(\boldsymbol{y},\boldsymbol{x})}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*} + \frac{\partial^2 w(\boldsymbol{y},\boldsymbol{x})}{\partial y_i \partial x_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}.$$

For q_{ij} we obtain

$$\begin{aligned}
q_{ij} &= \frac{\partial}{\partial x_j}\left(\frac{1}{s(\boldsymbol{x},\boldsymbol{x})}\frac{\partial s(\boldsymbol{y},\boldsymbol{x})}{\partial y_i}\right)\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*} \\
&= \frac{1}{s(\boldsymbol{x}^*,\boldsymbol{x}^*)^2}\left(\frac{\partial^2 s(\boldsymbol{y},\boldsymbol{x})}{\partial y_i \partial x_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}s(\boldsymbol{x}^*,\boldsymbol{x}^*) - \frac{\partial s(\boldsymbol{x},\boldsymbol{x})}{\partial x_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}\frac{\partial s(\boldsymbol{y},\boldsymbol{x})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}\right) \\
&= \frac{1}{s(\boldsymbol{x}^*,\boldsymbol{x}^*)}\frac{\partial^2 s(\boldsymbol{y},\boldsymbol{x})}{\partial y_i \partial x_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}, \qquad \text{[S30]}
\end{aligned}$$

where in the final simplification step we use the conclusion drawn from ***Equation S12*** that $0 = \partial s(\boldsymbol{y}, \boldsymbol{x}^*)/\partial y_i|_{\boldsymbol{y}=\boldsymbol{x}^*}$ and therefore the term after the minus sign disappears.

The mixed derivative of the function s is obtained by differentiating ***Equation S13*** with respect to x_j resulting in

$$\begin{aligned}
\frac{\partial^2 s(\boldsymbol{y}, \boldsymbol{x})}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*} &= K \frac{\partial}{\partial x_j}\left(\frac{1}{\big(K + c(\boldsymbol{y}, \boldsymbol{x})\big)^2}\frac{\partial c(\boldsymbol{y}, \boldsymbol{x})}{\partial y_i}\right)\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*} \\
&= \frac{K}{\big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)^4}\left(\frac{\partial^2 c(\boldsymbol{y}, \boldsymbol{x})}{\partial y_i \partial x_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}\big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)^2 - \frac{\partial c(\boldsymbol{y}, \boldsymbol{x})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}\frac{\partial c(\boldsymbol{y}, \boldsymbol{x})^2}{\partial x_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}\right) \\
&= \frac{K}{\big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)^2}\frac{\partial^2 c(\boldsymbol{y}, \boldsymbol{x})}{\partial y_i \partial x_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*},
\end{aligned} \tag{S31}$$

where the final simplification follows from the fact that $0 = \partial c(\boldsymbol{y}, \boldsymbol{x}^*)/\partial y_i|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}$.

The mixed derivative for the function c is obtained by differentiating ***Equation S14*** with respect to x_j resulting in

$$\begin{aligned}
\frac{\partial^2 c(\boldsymbol{y}, \boldsymbol{x}^*)}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*} &= \frac{\partial}{\partial x_j}\left(c(\boldsymbol{y}, \boldsymbol{x})\sum_{k=1}^{m}\left(\frac{1}{e_k(\boldsymbol{y}) + e_k(\boldsymbol{x})}\frac{\partial e_k(\boldsymbol{y})}{\partial y_i}\right)\right)\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*} \\
&= c(\boldsymbol{x}^*, \boldsymbol{x}^*)\sum_{k=1}^{m}\frac{\partial}{\partial x_j}\left(\frac{1}{e_k(\boldsymbol{y}) + e_k(\boldsymbol{x})}\frac{\partial e_k(\boldsymbol{y})}{\partial y_i}\right)\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*} \\
&= -c(\boldsymbol{x}^*, \boldsymbol{x}^*)\sum_{k=1}^{m}\frac{1}{4e_k(\boldsymbol{x}^*)^2}\frac{\partial e_k(\boldsymbol{y})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}\frac{\partial e_k(\boldsymbol{x})}{\partial x_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}.
\end{aligned} \tag{S32}$$

Here, the one but last simplification step again follows from the fact that $0 = \partial c(\boldsymbol{y}, \boldsymbol{x}^*)/\partial y_i|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}$.

By recursively substituting ***Equations S31*** and ***S32*** into ***Equation S30*** we obtain

$$\begin{aligned}
q_{ij} &= -\frac{K}{K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)}\sum_{k=1}^{m}\frac{1}{4e_k(\boldsymbol{x}^*)^2}\frac{\partial e_k(\boldsymbol{y})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}\frac{\partial e_k(\boldsymbol{x}^*)}{\partial x_j^*}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}. \\
&= -\frac{K}{4\big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)}\sum_{k=1}^{m}\left(\sum_{l=1}^{h}\sigma_{k\,il}^{-1}(p_{kl} - x_l^*)\right)\left(\sum_{l=1}^{h}\sigma_{k\,jl}^{-1}(p_{kl} - x_l^*)\right),
\end{aligned} \tag{S33}$$

where in the last step we substituted ***Equation S18***. This result can be rewritten in matrix form as

$$\mathsf{Q} = -\frac{K}{4\big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)}\sum_{k=1}^{m}\left(\Sigma_k^{-1}\big(\boldsymbol{p}_k - \boldsymbol{x}^*\big)\big(\boldsymbol{p}_k - \boldsymbol{x}^*\big)^{\mathrm{T}}\Sigma_k^{-1}\right). \tag{S34}$$

Finally, substituting $\boldsymbol{x}^*$ with ***Equation S22*** we obtain

$$\boxed{\mathsf{Q} = -\frac{1}{4}\frac{K}{K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)}\sum_{k=1}^{m}\left(\Sigma_k^{-1}\big(\boldsymbol{p}_k - \bar{\boldsymbol{p}}_w\big)\big(\boldsymbol{p}_k - \bar{\boldsymbol{p}}_w\big)^{\mathrm{T}}\Sigma_k^{-1}\right)}. \tag{S35}$$

Combining ***Equations S29*** and ***S35*** and simplifying results in the Jacobian matrix

$$\mathsf{J} = \mathsf{H} + \mathsf{Q} = -\frac{1}{2}\frac{K}{K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)}\sum_{k=1}^{m}\Sigma_k^{-1}. \tag{S36}$$

Note that all matrices Σ_k are symmetric and therefore J is symmetric. Furthermore, J is the product of a negative scalar and sum of m positive-definite matrices. Taken together, we can conclude that J is negative definite and the singular point $\boldsymbol{x}^*$ is absolutely convergence stable in the sense of Leimar (15).

7.4. Special case: Identically shaped Gaussian efficiency functions.

For the special case that the covariance matrices Σ_i are equal ($\Sigma_1 = \Sigma_2 = \ldots = \Sigma_m = \Sigma$) the Hessian matrix simplifies to

$$\begin{aligned} \mathsf{H} &= \frac{Km}{4\big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)} \Sigma^{-1} \left(\frac{1}{m} \sum_{k=1}^{m} \Big(\big(\boldsymbol{p}_k - \bar{\boldsymbol{p}}\big) \big(\boldsymbol{p}_k - \bar{\boldsymbol{p}}\big)^{\mathrm{T}} \Big) - 2\Sigma \right) \Sigma^{-1} \\ &= \frac{Km}{4\big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)} \Sigma^{-1} \Big(\mathsf{V} - 2\Sigma \Big) \Sigma^{-1}, \end{aligned} \qquad \text{[S37]}$$

where we used that $\mathsf{V} = 1/m(\sum_{k=1}^{m} (\boldsymbol{p}_k - \bar{\boldsymbol{p}})(\boldsymbol{p}_k - \bar{\boldsymbol{p}})^{\mathrm{T}})$ is the variance-covariance matrix V of the positions $\boldsymbol{p}_k$ of the pathogen vectors. Since the ratio in ***Equation S37*** is positive and Σ^{-1} is positive definite it follows that the definiteness of H is given by the definiteness of the matrix $\mathsf{V} - 2\Sigma$. Hence, the singular point $\boldsymbol{x}^*$ is uninvadable whenever $\mathsf{V} - 2\Sigma$ is negative definite. Panel (***B***) in ***Figure S3*** is based on this case.

How do the entries of V and Σ affect the definiteness of H? Generally, increasing a diagonal element of a symmetric matrix either increases all its eigenvalues or increases some eigenvalues while leaving the remaining ones constant (e.g. 11, corollary 4.3.12). Thus, increasing diagonal elements of V or decreasing diagonal elements of Σ can increase the eigenvalues of $\mathsf{V} - 2\Sigma$ and never decreases them. Thus, increasing the variance in the pathogen vectors and decreasing the variance in the efficiency functions e_i makes evolutionary branching more likely. No simple statement is possible with respect to the off-diagonal elements.

7.5. Special case: Maximal symmetry.

For the even more special case that $\Sigma = \sigma^2 \mathsf{I}$ and that the pathogen vectors are arranged in a symmetric manner resulting in $\mathsf{V} = v^2 \mathsf{I}$ we obtain

$$\mathsf{H} = \frac{Km}{4\sigma^{-4}\big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)} \big(v^2 - 2\sigma^2\big) \mathsf{I}. \qquad \text{[S38]}$$

In this case, the singular point $\boldsymbol{x}^*$ is a branching point whenever $v^2 > 2\sigma^2$. Assuming that all m pathogen vectors have an equal distance d to each other implies that they are arranged in an $m-1$-dimensional regular simplex (see ***Figure 1***). For this case $v^2 = d^2/(2m)$ and we obtain the branching condition $d > 2\sigma\sqrt{m}$. Panel (***A***) and the results in the main text are based on this case.

References

1. **Abrams PA**, Matsuda H, Harada Y. Evolutionarily unstable fitness maxima and stable fitness minima of continuous traits. Evolutionary Ecology. 1993; doi: 10.1007/bf01237642.
2. **Champagnat N**, Ferrière R, Méléard S. Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. Theoretical Population Biology. 2006; doi: 10.1016/j.tpb.2005.10.004.
3. **Débarre F**, Nuismer SL, Doebeli M. Multidimensional (co) evolutionary stability. The American Naturalist. 2014; doi: 10.1086/677137.
4. **Dercole F**, Rinaldi S. Analysis of evolutionary processes: the adaptive dynamics approach and its applications: the adaptive dynamics approach and its applications. Princeton University Press
5. **Dieckmann U**, Law R. The dynamical theory of coevolution: a derivation from stochastic ecological processes. Journal of Mathematical Biology. 1996; doi: 10.1007/bf02409751.
6. **Doebeli M**. Adaptive Diversification (MPB-48). Princeton University Press doi: 10.2307/j.ctt7rgw4.
7. **Durinx M**, Metz JAJH, Meszéna G. Adaptive dynamics for physiologically structured population models. Journal of Mathematical Biology. 2008;
8. **Fisher RA**. The genetical theory of natural selection. Oxford University Press
9. **Geritz SAH**, Kisdi É, Meszéna G, Metz JAJ. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evolutionary Ecology. 1998; doi: 10.1023/a:1006554906681.
10. **Geritz SAH**, Metz JAJ, Rueffler C. Mutual invadability near evolutionarily singular strategies for multivariate traits, with special reference to the strongly convergence stable case. Journal of Mathematical Biology. 2016; doi: 10.1007/s00285-015-0944-6.
11. **Horn RA**, Johnson CR. Matrix analysis, 2nd. New York: Cambridge University Press
12. **Iwasa Y**, Pomiankowski A, Nee S. The evolution of costly mate preferences II. The "handicap" principle. Evolution. 1991; doi: 10.1111/j.1558-5646.1991.tb02646.x.
13. **Kimura M**, Crow JF. The number of alleles that can be maintained in a finite population. Genetics. 1964;
14. **Lande R**. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. Evolution. 1979; doi: 10.2307/2407630.

15. **Leimar O**. Multidimensional convergence stability. Evolutionary Ecology Research. 2009;
16. **Metz JAJ**. Fitness. In: Jorgensen SE, Fath B, editors. *Encyclopedia of ecology* Oxford: Elsevier, available online as IIASA Interim Report IR-06-061; 2008. doi: 10.1016/b978-008045405-4.00792-8.
17. **Metz JAJ**, de Kovel CGF. The canonical equation of adaptive dynamics for Mendelian diploids and haplo-diploids. Interface Focus. 2013; doi: 10.1098/rsfs.2013.0025.
18. **Metz JAJ**, Nisbet RM, Geritz SAH. How should we define 'fitness' for general ecological scenarios? Trends in Ecology & Evolution. 1992;
19. **Rueffler C**, Van Dooren TJM, Leimar O, Abrams PA. Disruptive selection and then what? Trends in Ecology & Evolution. 2006; doi: 10.1016/j.tree.2006.03.003.
20. **Wright S**. Evolution in Mendelian populations. Genetics. 1931; doi: 10.1007/bf02459575.
21. **Wright S**. Polyallelic random drift in relation to evolution. Proceedings of the National Academy of Sciences. 1966; doi: 10.1073/pnas.55.5.1074.