The evolution and maintenance of extraordinary allelic diversity

Mattias Siljestam[1,*] & Claus Rueffler[1]

[1] Department of Ecology and Genetics, Animal Ecology, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden
[*] Corresponding author: mattias.siljestam@ebc.uu.se

Many loci harbour allelic polymorphism, but the majority of highly polymorphic genes are related to the immune function.[1] Genes of the major histocompatibility complex (MHC) are the most polymorphic loci in vertebrates[2,3] with over 100 alleles reported within a population.[4] How such extraordinary polymorphism arose and is maintained is controversial but balancing selection rather than purely neutral processes seems responsible.[2,5] Heterozygote advantage can in principle maintain an arbitrary number of alleles at a single locus[6–8] but biologically explicit models have failed to reliably predict the coexistence of significantly more than ten alleles.[9–18] Consequently, current explanations of polymorphism at MHC loci ascribe only a minor role to heterozygote advantage[19–25] despite the fact that this mechanism has been documented in many empirical studies.[22,24,26–31] Here, we present an eco-evolutionary model showing that evolution can through a self-organising process give rise to the emergence of more than 100 alleles at a single locus that are maintained by heterozygote advantage. In our model, MHC molecules trigger immune responses that are traded off: no single MHC molecule can efficiently induce a response against all pathogens simultaneously. The emerging heterozygote advantage allows for the build-up of up to 20 coexisting alleles in the case of two pathogens while with eight pathogens this number can reach 200. We conclude that evolution driven by heterozygote advantage alone has the potential to explain the high degree of polymorphism found at MHC loci. Our model does not only explain diversity at MHC loci but applies more generally whenever gene products are involved in multiple functions important for survival and can therefore also explain the evolution of polymorphism at other genes.

Heterozygote advantage is a well established explanation for single locus polymorphism with the sickle cell locus as a classical text book example.[32] However, whether heterozygote advantage is generally important for the maintenance of genetic polymorphism is questioned.[33,34] 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.[2,23,35] Heterozygote advantage as an explanation for this high level of polymorphism was introduced more than 40 years ago by Doherty and Zinkernagel with the idea that having MHC molecules of two different alleles allows for the recognition of a broader spectrum of pathogens resulting in higher fitness.[26] While early theoretical work suggested that heterozygote advantage can maintain the high number of alleles found at MHC loci[6–8] later work discredited this idea[9–18] and other mechanisms were suggested to be important.[2,5,13,17,19,21,25,35]

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,[36] 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[36–39] and therefore provide different degrees of protection against different pathogens.[2,23,36–38] 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.[39] This could explain the lower efficiency among generalist MHC molecules.

We model the trade-off using Gaussian functions $e_k(\boldsymbol{x}_i)$ as shown in Fig. 1. Here, each allele is characterised by an allelic value $\boldsymbol{x}_i$ determining its MHC molecule's efficiency $e_k(\boldsymbol{x}_i)$ against pathogen k. Since MHC alleles are codominantly expressed[23,40] the protection for an individual from both alleles against a pathogens k is modelled as the arithmetic mean $(e_k(\boldsymbol{x}_i) + e_k(\boldsymbol{x}_j))/2$. The phenotypic condition c is given by the product of an individual's defences against all m pathogens,

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

Here, $c_{\max}$ is the condition of a hypothetical individual with perfect defence against all pathogens.

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

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

where K is the survival half-saturation constant. This survival function increases approximately linearly for c smaller than K and saturates for large c where $s \approx 1$ (orange lines in bottom panel of Fig. 2).

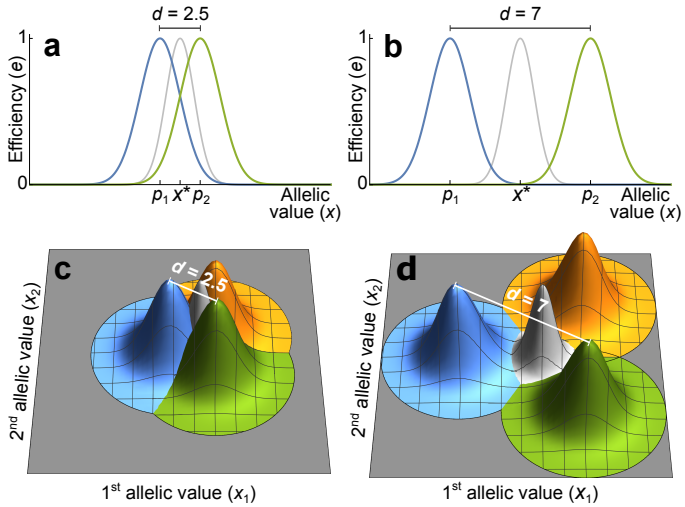

Figure 1: Gaussian trade-offs between an MHC molecule's efficiencies e_k for two (**a-b**) and three (**c-d**) pathogens. (**a-b**) The blue and green lines give the efficiencies e_1 and e_2 for pathogen 1 and 2, respectively, as a function of the allelic value x. These efficiency functions have maxima at p_1 and p_2 separated by distance d. (**c-d**) Coloured cones give the trade-off between three pathogens, which is modeled with two-dimensional Gaussian functions. Thus, allelic values $\boldsymbol{x} = (x_1, x_2)$ and pathogen optima $\boldsymbol{p}_k$ become two-dimensional vectors, i.e., points on the plane. 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 Eq. 1 appropriately). It is straightforward to extend this model to m pathogens and (m-1)-dimensional trait vectors (Supplementary Information 7.5). Here, we focus on the symmetric case with equally spaced pathogen optima and all covariance matrices of the Gaussians equalling the identity matrix. The general case with non-equally spaced pathogen optima and arbitrary covariance matrices is described in the Supplementary Information 5.

We study the evolutionary dynamics of allelic values $\boldsymbol{x}_i$ in a diploid Wright-Fisher model with selection[41,42] by following the fate of recurrent mutations of small effect.[43,44] The assumption of small mutational effects is made to allow for analytical derivations but simulations show that none of our results critically depend on it (Fig. S2). Thus, we consider a diploid population of fixed size N with non-overlapping generations, 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. The long-term evolutionary dynamics is obtained by iterating this procedure (Fig. 2 top panel). The evolution of the emerging allelic lineages in this model is driven by increasing the allele's expected survival (or marginal fitness). This follows from standard population genetics theory. Detailed derivations of all results are in Supplementary Information 6 & 7.

Initially, allelic values evolve towards an allele $\boldsymbol{x}^*$ given by the arithmetic mean of the optimal allelic values $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m$ against the m pathogens (Fig. 1). We refer to $\boldsymbol{x}^*$ as the generalist allele. From here on we present our results in two steps. First we explain how an allelic dimorphism can emerge after $\boldsymbol{x}^*$ has been reached. We then proceed by describing how after this initial diversification allelic diversity can increase further and reach extraordinarily high levels.

Our analytical derivations (Supplementary Information 7.5) for m pathogens with dissimilarity d (see Fig. 1) show that for similar pathogens such that $d < 2\sqrt{m}$ (Fig. 1a,c) the generalist allele $\boldsymbol{x}^*$ is an evolutionary stable strategy (ESS), i.e., it cannot be invaded by nearby mutants and therefore is the endpoint of evolution (Fig. S1), as it has a relatively high efficiency for all pathogens. If, however, the pathogens are more different such that $d > 2\sqrt{m}$ and the generalist's efficiency low (Fig. 1b,d), then any nearby mutant allele can invade and coexist with the generalist allele $\boldsymbol{x}^*$. Under this condition a heterozygote genotype carrying the generalist and a mutant allele has higher survival s than either homozygote and the allelic dimorphism is maintained by heterozygote advantage. A new mutant allele can invade if its more specialised than the allele from which it originated. If successful such a mutant also replaces its parent allele. In summary, two allelic lineages can emerge from the generalist allelic value $\boldsymbol{x}^*$ and subsequently diverge (Fig. 2). Thus, $\boldsymbol{x}^*$ is an evolutionary branching point in the sense of adaptive dynamics theory.[43,44]

Increasing the difference between the two alleles has two opposing effects. The condition and thereby the survival of the heterozygote genotype increases because the two alleles present in the dimorphism provide increasingly better protection against complementary sets of pathogens (divergent allele advantage in sense of Wakeland *et al.*[37]). On the other hand, survival of the two homozygote genotypes decreases because they become increasingly vulnerable for the set of pathogens for which their alleles do not offer protection. Allelic divergence proceeds as long as the increase in heterozygote survival outweighs the decreases in homozygote survival. Since the increase in survival of heterozygotes slows down (plateau of the orange curves in Fig. 2) the two opposing forces eventually balance each other and divergence comes halts. Due to random mating half of the population are high survival heterozygotes and the remaining half are homozygotes with low survival. At this point, one of the lineages branches again and thereby decreases the proportion of low-survival homozygotes. The three alleles coexisting in this trimorphism diverge from each other as the increase in heterozygote survival outweighs the decreased survival in the (now also less frequent) homozygotes once again (see Fig. 2a at time $t = 3 \times 10^4$, grey plane).

This process of evolutionary branching and allelic divergence repeats itself and the allelic diversity continues

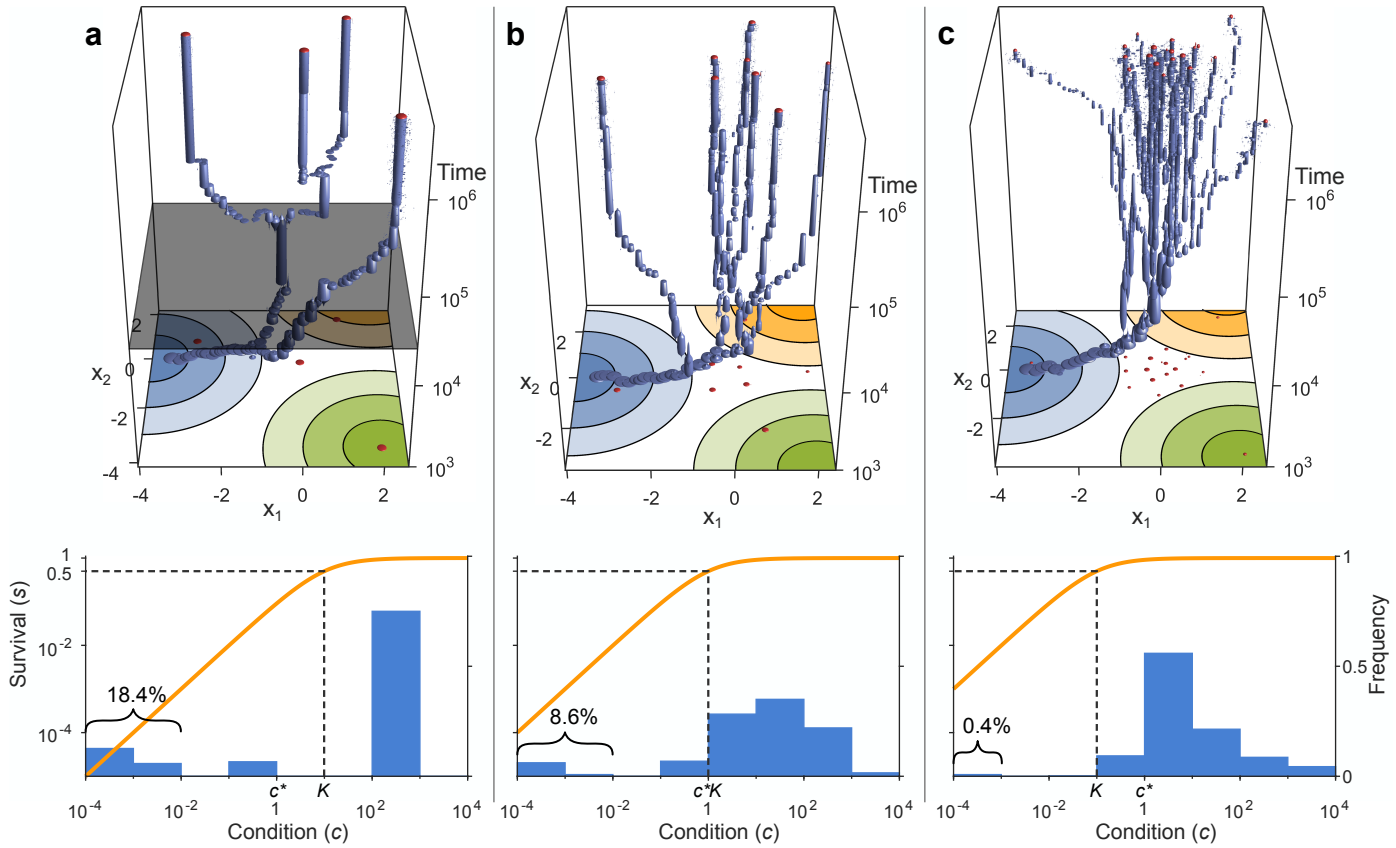

Figure 2: Evolution of allelic values in the presence of three pathogens (arranged as in Fig. 1d) for three different values of the survival half-saturation constant K (**a**: $K = 10$, **b**: $K = 1$, **c**: $K = 0.1$; dashed line in lower panel). The top panel shows the result of individual-based simulations where the two horizontal axes give the two values $\boldsymbol{x} = (x_1, x_2)$ that characterise 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 shown in Fig. 1d. The bottom panel gives survival s as a function of condition c as defined by Eq. 2 on a log-log scale (orange line, left vertical axis) and the frequencies of individual condition 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 (Supplementary Information 7). In (**a**) there are three consecutive branching events (the second branching event is marked out 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 $1/(K + 1) \approx 0.091$ (Eq. 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 $\delta = 0.14$. Other parameter values: $d = 7$, per-capita mutation probability $\mu = 5 \times 10^{-7}$, population size $N = 2 \times 10^5$.

to increase (Fig. 2). The number of consecutive branching events increases with the number of pathogens m and their dissimilarity d, but also depends on K, determining how quickly the survival function saturates. Fig. 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 this parameter space more than 100 (solid contour lines in Fig. 3) and up to 200 alleles can emerge and coexist. In Fig. S3 we show that qualitatively this prediction is robust against violations of the symmetry assumptions made in Fig. 1.

Kimura and Crow[6] and subsequently Wright[7] 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 (1) all heterozygotes have a similarly high fitness and (2) all homozygotes a similarly low fitness. These conditions were deemed unrealistic by the authors themselves, an assessment confirmed in later studies[9–18] In contrast, our study shows that heterozygote advantage alone can explain the high number of coexisting alleles found at MHC loci 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 are fulfilled. Specifically, as long as K is not too large the survival of heterozygotes quickly saturates with increasing condition (orange line in

Fig. 2 for $c > K$). As a result, 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) formulated by Kimura and Crow. If, additionally, pathogens are sufficiently different (d large), then 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 Fig. 1b,d). Thus, evolution of allelic values also results in condition (2) of Crow and Kimura being fulfilled by the majority of homozygotes. However, for K very small the number of coexisting alleles decreases again (left side in the panels in Fig. 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.

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 set of alleles able to coexist (Fig. S2).

Many vertebrates have several MHC loci with similar functions.[5,21,23] 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 fulfils the assumptions that it 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 lead to a situation where the number of duplicates equals the number of functions.[45] 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[5,21,23,35] indicating that the mechanism presented here can in principle explain the high number of coexisting MHC alleles.

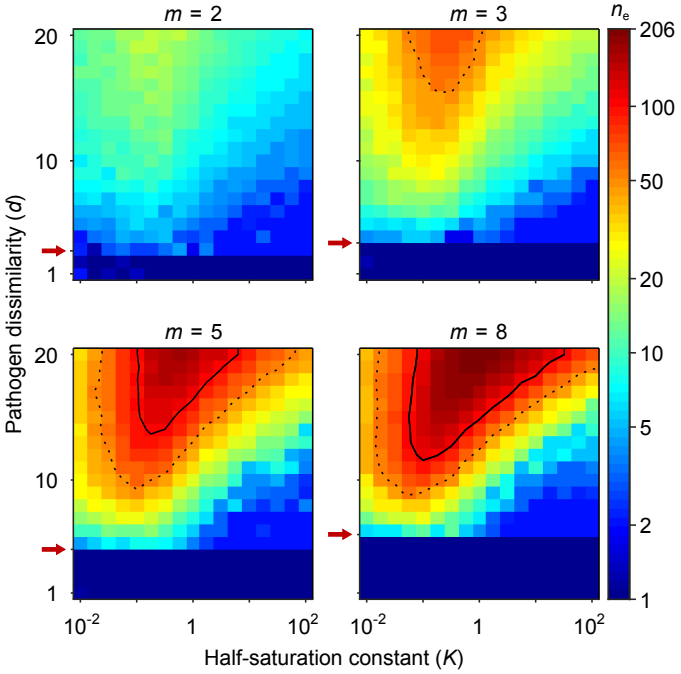

Figure 3: Number of coexisting alleles for m pathogens as a function of pathogen dissimilarity d and the survival half-saturation constant K. Results are based on a single individual-based simulation per pixel run for 10^6 generations. This assures that the equilibrium distribution of alleles is reached. 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. It is defined as the inverse of homozygosity and is lower or equal to the absolute number of allele. Note that for our demographic parameters the expected n_e under mutation-drift balance alone equals 1.2 (Supplementary Information 4). 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 $\delta = 0.03d$.

References

[1] Jin, P. & Wang, E. Polymorphism in clinical immunology–from hla typing to immunogenetic profiling. *Journal of Translational Medicine* **1**, 8 (2003).

[2] Apanius, V., Penn, D., Slev, P. R., Ruff, L. R. & Potts, W. K. The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology* **17** (1997).

[3] Sommer, S. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology* **2**, 16 (2005).

[4] Duncan, W., Wakeland, E. & Klein, J. Histocompatibility-2 system in wild mice. *Immunogenetics* **9**, 261–272 (1979).

[5] Spurgin, L. G. & Richardson, D. S. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society of London B: Biological Sciences* **277**, 979–988 (2010).

[6] Kimura, M. & Crow, J. F. The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738 (1964).

[7] Wright, S. Polyallelic random drift in relation to evolution. *Proceedings of the National Academy of Sciences* **55**, 1074–1081 (1966).

[8] Maruyama, T. & Nei, M. Genetic variability maintained by mutation and overdominant selection in finite populations. *Genetics* **98**, 441–459 (1981).

[9] Lewontin, R., Ginzburg, L. & Tuljapurkar, S. Heterosis as an explanation for large amounts of genic polymorphism. *Genetics* **88**, 149–169 (1978).

[10] Spencer, H. G. & Marks, R. W. The maintenance of single-locus polymorphism. I. Numerical studies of a viability selection model. *Genetics* **120**, 605–613 (1988).

[11] Hedrick, P. W. Pathogen resistance and genetic variation at MHC loci. *Evolution* **56**, 1902–1908 (2002).

[12] de Boer, R. J., Borghans, J. A., Van Boven, M., Keşmir, C. & Weissing, F. J. Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. *Immunogenetics* **55**, 725–731 (2004).

[13] Borghans, J. A., Beltman, J. B. & De Boer, R. J. MHC polymorphism under host-pathogen coevolution. *Immunogenetics* **55**, 732–739 (2004).

[14] Stoffels, R. J. & Spencer, H. G. An asymmetric model of heterozygote advantage at major histocompatibility complex genes: degenerate pathogen recognition and intersection advantage. *Genetics* **178**, 1473–1489 (2008).

[15] Trotter, M. V. & Spencer, H. G. The generation and maintenance of genetic variation by frequency-dependent selection: constructing polymorphisms under the pairwise interaction model. *Genetics* **180**, 1547–1557 (2008).

[16] Trotter, M. V. & Spencer, H. G. Models of frequency-dependent selection with mutation from parental alleles. *Genetics* **195**, 231–242 (2013).

[17] Ejsmond, M. J. & Radwan, J. Red queen processes drive positive selection on major histocompatibility complex (MHC) genes. *PLoS Comput Biol* **11** (2015).

[18] 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* **86**, 343–352 (2015).

[19] Hedrick, P. W. Balancing selection and MHC. *Genetica* **104**, 207–214 (1999).

[20] Gould, S. J., Hildreth, J. E. & Booth, A. M. The evolution of alloimmunity and the genesis of adaptive immunity. *The Quarterly review of biology* **79**, 359–382 (2004).

[21] Wegner, K. Historical and contemporary selection of teleost MHC genes: did we leave the past behind? *Journal of Fish Biology* **73**, 2110–2132 (2008).

[22] Kekäläinen, J., Vallunen, J. A., Primmer, C. R., 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* (2009).

[23] Eizaguirre, C. & Lenz, T. Major histocompatibility complex polymorphism: dynamics and consequences of parasite-mediated local adaptation in fishes. *Journal of Fish Biology* **77**, 2023–2047 (2010).

[24] Lenz, T. L. Computational prediction of MHC II-antigen binding supports divergent allele advantage and explains trans-species polymorphism. *Evolution* **65**, 2380–2390 (2011).

[25] Loiseau, C. *et al.* Plasmodium relictum infection and MHC diversity in the house sparrow (Passer domesticus). *Proceedings of the Royal Society of London B: Biological Sciences* **278**, 1264–1272 (2011).

[26] Doherty, P. C. & Zinkernagel, R. M. Enhanced immunological surveillance in mice heterozygous at the h-2 gene complex. *Nature* **256**, 50–52 (1975).

[27] Hughes, A. L. & Nei, M. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proceedings of the National Academy of Sciences* **86**, 958–962 (1989).

[28] Jeffery, K. J. & Bangham, C. R. Do infectious diseases drive MHC diversity? *Microbes and Infection* **2**, 1335–1341 (2000).

[29] McClelland, E. E., Penn, D. J. & Potts, W. K. Major histocompatibility complex heterozygote superiority during coinfection. *Infection and immunity* **71**, 2079–2086 (2003).

[30] 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* **22**, 1254–1259 (2005).

[31] Oliver, M., Telfer, S. & Piertney, S. Major histocompatibility complex (MHC) heterozygote superiority to natural multiparasite infections in the water vole (arvicola terrestris). *Proceedings of the Royal Society of London B: Biological Sciences* **276**, 1119–1128 (2009).

[32] Allison, A. C. Protection afforded by sickle-cell trait against subtertian malarial infection. *British medical journal* **1**, 290 (1954).

[33] Hedrick, P. W. What is the evidence for heterozygote advantage selection? *Trends in ecology & evolution* **27**, 698–704 (2012).

[34] Sellis, D., Kvitek, D. J., Dunn, B., Sherlock, G. & Petrov, D. A. Heterozygote advantage is a common outcome of adaptation in saccharomyces cerevisiae. *Genetics* **203**, 1401–1413 (2016).

[35] Penn, D. J. The scent of genetic compatibility: sexual selection and the major histocompatibility complex. *Ethology* **108**, 1–21 (2002).

[36] Eizaguirre, C., Lenz, T. L., Kalbe, M. & Milinski, M. Divergent selection on locally adapted major histocompatibility complex immune genes experimentally proven in the field. *Ecology letters* **15**, 723–731 (2012).

[37] Wakeland, E. K. *et al.* Ancestral polymorphisms of MHC class II genes: divergent allele advantage. *Immunologic Research* **9**, 115–122 (1990).

[38] 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* **7** (2012).

[39] Chappell, P. E. *et al.* Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. *Elife* **4** (2015).

[40] Abbas, A. K., Lichtman, A. H. & Pillai, S. *Basic immunology: functions and disorders of the immune system* (Elsevier Health Sciences, 2014).

[41] Fisher, R. A. *The genetical theory of natural selection* (Oxford University Press, 1930).

[42] Wright, S. Evolution in mendelian populations. *Genetics* **16**, 97–159 (1931).

[43] Metz, J. A. J., Nisbet, R. M. & Geritz, S. A. H. How should we define fitness for general ecological scenarios? *Trends in Ecology & Evolution* **7**, 198–202 (1992).

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

[45] Proulx, S. R. & Phillips, P. C. Allelic divergence precedes and promotes gene duplication. *Evolution* **60**, 881–892 (2006).

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.

Supplementary Information
The evolution and maintenance of extraordinary allelic diversity

Mattias Siljestam* and Claus Rueffler
* Corresponding author: mattias.siljestam@ebc.uu.se

Table of 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 Effective number of alleles due to demographic processes **4**

5 Robustness of the model: Deviations from symmetry **5**
5.1 Random placement of pathogen optima . . . 5
5.2 Random covariance matrices for the pathogen efficiencies . . . 6

6 Preliminaries **7**
6.1 Adaptive dynamics and invasion fitness . . . 7
6.2 Model description . . . 9

7 Mathematical derivations **9**
7.1 The evolutionarily singular point . . . 9
7.2 Derivation of the Hessian matrix . . . 11
7.3 Derivation of the Jacobian matrix . . . 12
7.4 Special case: Identically shaped pathogen optima . . . 14
7.5 Special case: Maximal symmetry . . . 14

1 Supplementary figure S1: Evolutionary dynamics without branching

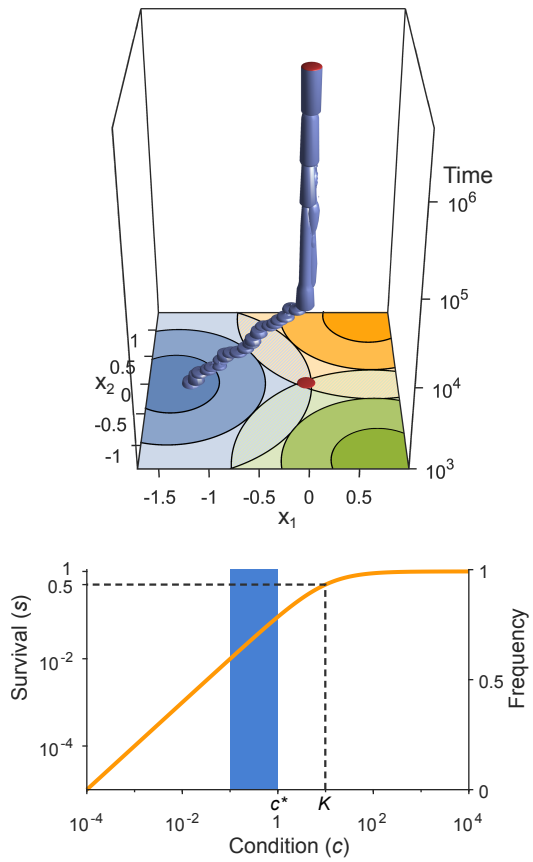

Figure S1: Evolution of allelic values in the presence of three pathogens. This figure is analogous to Fig. 2 but the distance between pathogen optima equals $d = 2.5$ (as in Fig. 1c) while in Fig. 2 this distance was much larger ($d = 7$, as in Fig. 1d). As a consequence, the condition for evolutionary 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 $s \approx 1/(K + 1) = 0.\overline{09}$ (Eq. 9). Note that while in Fig. 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 $\delta = 0.14$. 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

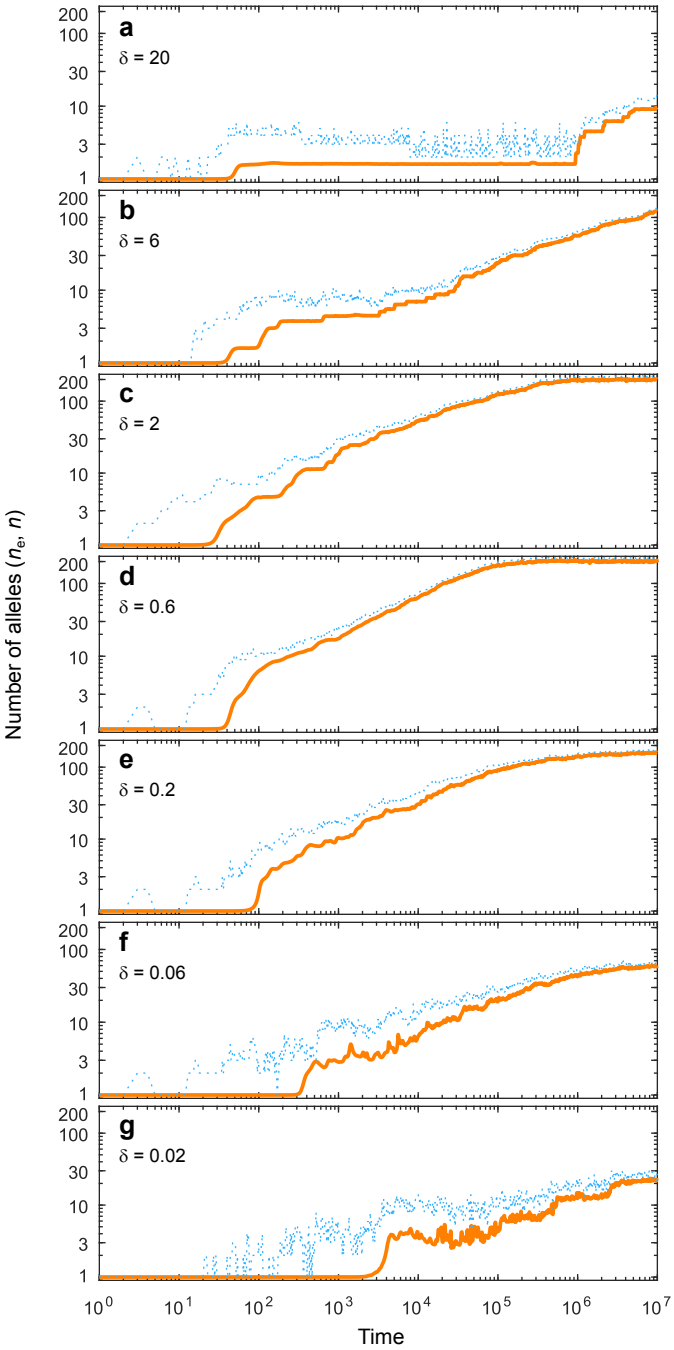

Figure S2: Number of coexisting alleles as they emerge in individual-based simulations for different expected mutational step sizes δ and eight pathogens ($m = 8$). Parameters are chosen such that up to 200 alleles can evolve ($K = 0.5$, $d = 20$; see bottom right panel in Fig. 3 in the main text). Orange 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 $\delta = 0.6$ (**d**; pathogen optima are $d/\delta = 20/0.6 = 33.3$ average mutational steps apart) as used in Fig. 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 optima are $20/0.02 = 1000$ 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 optima are one 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 Fig. 3: population size $N = 2 \times 10^5$, per-capita mutation probability $\mu = 5 \times 10^{-7}$.

3 Table of mathematical notation

Bold font indicates vector (e.g. $\boldsymbol{x}$), otherwise scalar (e.g. x). Capital letter in sans serif font indicate matrices (e.g. $\boldsymbol{X}$). 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 optima
δ	expected mutational step size
e_k	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
$\boldsymbol{p}_k$	kth pathogen optimum
s	survival function
s_max	maximum survival
$\mathsf{\Sigma}$	Gaussian covariance matrix of the efficiency function e_k (Eq. 11)
V	variance-covariance matrix of the position of the pathogen optima
v^2	variance of the position of the pathogen optima 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 mutant allele

4 Effective number of alleles due to demographic processes

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

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

(see e.g. Gillespie 2004). The effective number of alleles n_e under mutation drift balance is defined as $n_\mathrm{e} = 1/\mathrm{G} = 1 + 4N\mu$. 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 $n_\mathrm{e} \approx 1.2$. Hence, values of n_e significant higher than 1.2 indicate the presence of alleles maintained by balancing selection.

5 Robustness of the model: Deviations from symmetry

Fig. 3 in the main text shows the final number of coexisting alleles for different parameter combinations. This result is based on two symmetry assumptions, First, the m pathogen optima are placed fully symmetric, i.e., equidistantly resulting in a regular $(m-1)$-simplex. Second, the multivariate Gaussian functions describing the MHC-molecule's efficiencies e_k against pathogen k are assumed to be rotational symmetric as shown in Fig. 1. Thus, the variance-covariance matrix Σ_k in Eq. 11 is equal to the identity matrix I. Here, we test the robustness of the outcome shown Fig. 3 with respect to violations of these symmetry assumptions. We focus on the case with eight pathogens. The results are summarised in Fig. S3. Panel (a) is identical to the bottom right panel in Fig. 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 optima is perturbed while the Gaussian functions e_k are still rotational symmetric. Section 5.1 describes the procedure how the positions of the pathogen optima are randomised while controlling for their variance. The similarity between panel (a) and (b) indicates that deviations from a symmetric placement of pathogen optima have 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 Eq. 32. This Hessian is more likely to be non-negative definite for asymmetrically placed pathogen optima.

In Fig. S3(c,d) we, additionally to the non-symmetric placement of pathogen optima, allow for non-rotational symmetric Gaussian functions e_k. Thus, Σ_k can be different for each pathogen 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 constrained to the interval $(1-\varepsilon, 1+\varepsilon)$ such that the average variance is kept equal to 1. Section 5.2 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 Eq. 6 is even more likely to be non-negative definite with perturbations in Σ_k.

5.1 Random placement of pathogen optima

The distribution of eight pathogen optima 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 optima along a different dimension of the trait space the average variance equals $\mathrm{tr}(\mathsf{V})/7$, where $\mathrm{tr}(\mathsf{V})$ denotes the trace. This measure is unaffected by rotation of the points $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_8$. For symmetrically placed pathogen optima $\mathsf{V}_{\mathrm{sym}} = d^2/(2m)\mathrm{I}$, were I denotes the identity matrix, and therefore $\mathrm{tr}(\mathsf{V}_{\mathrm{sym}}) = d^2(m-1)/(2m)$. In panel (b-d) We obtain randomly placed pathogen optima with $\mathrm{tr}(\mathsf{V}_{\mathrm{per}}) = d^2(m-1)/2m = \mathrm{tr}(\mathsf{V}_{\mathrm{sym}})$ 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

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

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

5.2 Random covariance matrices for the pathogen efficiencies

In Fig. 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 Fig. S3(c,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 $\mathsf{D}_1, \ldots, \mathsf{D}_8$ are determined (one per pathogen optima) 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 $\mathsf{P}_k = \mathsf{R}_k \mathsf{M}_k \mathsf{R}_k^{\mathrm{T}}$.

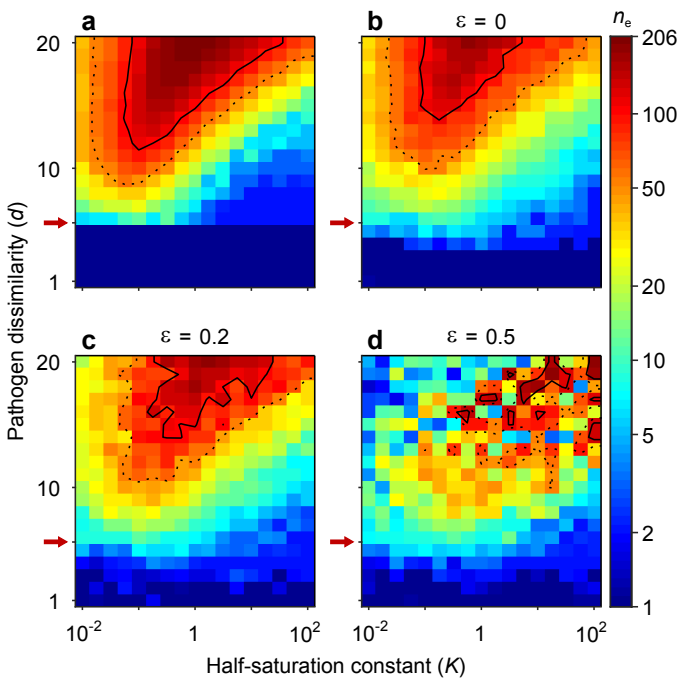

Figure S3: Number of coexisting alleles for eight pathogens as a function of pathogen dissimilarity d and the half-saturation constant K for symmetric (**a**) and non-symmetric pathogen optima (**b-d**). Results are based on a single individual-based simulation per pixel. Panel (**a**) shows results for equally spaced pathogen optima and rotational symmetric functions e_k (Eq. 11). It is identical to the bottom right panel in Fig. 3 and shown for comparison. Panel (**b-d**) show the result for increasing perturbations from symmetry. In panel (**b**) pathogen optima 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 optima, 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 optima are no longer at a constant distance d. In these panels, the y-axis shows the value d that corresponds to symmetrically placed pathogen optima with a variance equal to the mean variance of the randomly placed pathogen optima. Simulations were run for 10^6 generations to assure that the equilibrium distribution of alleles is reached. Results are reported in terms of the effective number of alleles n_e (see Fig. 3). Dashed and solid lines are contours corresponding to n_e equalling 50 and 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 $\delta = 0.03d$.

6 Preliminaries

6.1 Adaptive dynamics and invasion fitness

We analyse the model with an evolutionary invasion analysis using the adaptive dynamics formalism (Metz et al. 1992; Dieckmann and Law 1996; Geritz et al. 1998). 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 value $\boldsymbol{y}$ in a resident population at equilibrium consisting of the alleles $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ (Metz et al. 1992; Metz 2008).

Here, 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 (Fisher 1930; Wright 1931). The recurrence equation for allele $\boldsymbol{x}_a$ equals

$$f_{a,t+1} = f_{a,t} \sum_{b=1}^{n} f_{b,t} \frac{s(\boldsymbol{x}_a, \boldsymbol{x}_b)}{\bar{s}_t}, \tag{2}$$

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. The sum on the right hand side of Eq. 2 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

$$w(\boldsymbol{y}, \boldsymbol{X}) \;=\; \frac{1}{\bar{s}} \sum_{b=1}^{n} f_j s(\boldsymbol{y}, \boldsymbol{x}_b), \tag{3}$$

which is its invasion fitness. Note that because of the equilibrium assumption the time-subscript could be dropped. Allele $\boldsymbol{y}$ is able 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

$$\nabla w(\boldsymbol{x}_a, \boldsymbol{X})_i = \frac{\partial w(\boldsymbol{y}, \boldsymbol{X})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}_a} \tag{4}$$

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 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 affects are sufficiently small such that $\boldsymbol{y} = \boldsymbol{x}_a + \boldsymbol{\epsilon}$ for $\boldsymbol{x}_a \in \boldsymbol{X}$ and $\boldsymbol{\epsilon}$ small, then invasion of $\boldsymbol{y}$ implies extinction of $\boldsymbol{x}_a$ (Dercole and Rinaldi 2008, 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

$$\frac{\mathrm{d}\boldsymbol{x}_a}{\mathrm{d}t} = \mu f_a N \mathsf{C} \nabla w(\boldsymbol{x}_a, \boldsymbol{X}) \tag{5}$$

(Dieckmann and Law 1996; Champagnat et al. 2006; Durinx et al. 2008; Metz and de Kovel 2013). Here, μ is the per-capita mutation probability and C the mutational variance-covariance matrix. We note that Eq. 5 is structurally similar to the gradient equation of quantitative genetics, which is based on the assumption of weak selection or, equivalently, small genetic variances (Lande 1979; Iwasa et al. 1991; Abrams et al. 1993; Débarre et al. 2014). 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 μN is 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 $\boldsymbol{X} = \boldsymbol{x}$, this trait value of 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 Eq. 5. 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 (Geritz et al. 1998). 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 (Rueffler et al. 2006).

The classification of singular points becomes more complicated in multivariate trait spaces or when several strategies coexist in an evolutionarily singular coalition (Leimar 2009; Doebeli 2011; Geritz et al. 2016). 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 (Leimar 2009). 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 (Geritz et al. 2016).

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

$$h_{ij} \;=\; \frac{\partial^2 w(\boldsymbol{y}, \boldsymbol{x})}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}. \tag{6}$$

is negative definite (Leimar 2009; Doebeli 2011). 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 x^* as a function of the positions of the pathogen optima, 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 $\nabla w(\boldsymbol{x}^*, \boldsymbol{x}^*)$ (Leimar 2009), which is given by

$$\mathsf{J} \;=\; \mathsf{H} + \mathsf{Q} \tag{7}$$

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

$$q_{ij} \;=\; \frac{\partial^2 w(\boldsymbol{y}, \boldsymbol{x})}{\partial y_i \partial x_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}. \tag{8}$$

Leimar (2009) 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 Eq. 5. 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 and where 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, that 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. (2016) 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 (Geritz et al. 2016). 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

$$s(\boldsymbol{x}_a, \boldsymbol{x}_b) = \frac{s_{\max} c(\boldsymbol{x}_a, \boldsymbol{x}_b)}{K + c(\boldsymbol{x}_a, \boldsymbol{x}_b)}. \tag{9}$$

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

$$c(\boldsymbol{x}_a, \boldsymbol{x}_b) = c_{\max} \prod_{k=1}^{m} \frac{e_k(\boldsymbol{x}_a) + e_k(\boldsymbol{x}_b)}{2}, \tag{10}$$

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

$$e_k(\boldsymbol{x}) = \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{p}_k)^{\mathrm{T}} \Sigma_k^{-1} (\boldsymbol{x} - \boldsymbol{p}_k)\right) \tag{11}$$

(see Fig. 1). 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 optimum $\boldsymbol{p}_k$. Hence, pathogen optimum $\boldsymbol{p}_k$ and allelic trait values $\boldsymbol{x}$ can be interpreted as points in a h-dimensional space. The closer an allelic trait is to a pathogen optimum 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 Σ_k.

7 Mathematical derivations

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. Panel (c&d) in Fig. S3 are based on this general case. By 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 y is given by its marginal fitness and for a monomorphic population Eq. 3 simplifies to

$$w(\boldsymbol{y}, \boldsymbol{x}) = \frac{s(\boldsymbol{y}, \boldsymbol{x})}{s(\boldsymbol{x}, \boldsymbol{x})}. \tag{12}$$

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

$$\begin{aligned} \nabla w(\boldsymbol{x}, \boldsymbol{x})_i &= \frac{\partial w(\boldsymbol{y}, \boldsymbol{x})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}} \\ &= \frac{1}{s(\boldsymbol{x}, \boldsymbol{x})} \frac{\partial s(\boldsymbol{y}, \boldsymbol{x})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}}. \end{aligned} \tag{13}$$

Using the definitions for s (Eq. 9) and c (Eq. 10) and their derivatives

$$\frac{\partial s(\boldsymbol{y},\boldsymbol{x})}{\partial y_i} = \frac{K}{\big(K + c(\boldsymbol{y},\boldsymbol{x})\big)^2}\frac{\partial c(\boldsymbol{y},\boldsymbol{x})}{\partial y_i}. \tag{14}$$

and

$$\frac{\partial c(\boldsymbol{y},\boldsymbol{x})}{\partial y_i} = c(\boldsymbol{y},\boldsymbol{x})\sum_{k=1}^{m}\frac{1}{e_k(\boldsymbol{y}) + e_k(\boldsymbol{x})}\frac{\partial e_k(\boldsymbol{y})}{\partial y_i}, \tag{15}$$

where Eq. 15 is obtained by applying the generalised product rule $\frac{\partial}{\partial x}\Big(\prod_{i=1}^{n} f_i(x)\Big) = \Big(\prod_{i=1}^{n} f_i(x)\Big)\sum_{i=1}^{n}\frac{f_i'(x)}{f_i(x)}$, we obtain

$$\begin{aligned}\nabla w(\boldsymbol{x},\boldsymbol{x})_i &= \frac{K}{c(\boldsymbol{x},\boldsymbol{x})\big(K + c(\boldsymbol{x},\boldsymbol{x})\big)}\frac{\partial c(\boldsymbol{y},\boldsymbol{x})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}} \\ &= \frac{K}{2\big(K + c(\boldsymbol{x},\boldsymbol{x})\big)}\sum_{k=1}^{m}\frac{1}{e_k(\boldsymbol{x})}\frac{\partial e_k(\boldsymbol{y})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}}.\end{aligned} \tag{16}$$

In the next step we calculate the derivative of the function $e_k(\boldsymbol{y})$ (Eq. 11). Applying the chain rule and simplifying results in

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

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

$$\frac{\partial e_k(\boldsymbol{y})}{\partial y_i} = e_k(\boldsymbol{y})\sum_{j=1}^{h}\sigma_{kil}^{-1}(p_{kj} - y_j). \tag{17}$$

Substituting Eq. 17 into Eq. 16 finally results in

$$\nabla w(\boldsymbol{x},\boldsymbol{x})_i = \frac{K}{2\big(K + c(\boldsymbol{x},\boldsymbol{x})\big)}\sum_{k=1}^{m}\sum_{j=1}^{h}\sigma_{kij}^{-1}(p_{kj} - x_j)$$

and

$$\boxed{\nabla \boldsymbol{w}(\boldsymbol{x}) = \frac{K}{2\big(K + c(\boldsymbol{x},\boldsymbol{x})\big)}\sum_{k=1}^{m}\Sigma_k^{-1}(\boldsymbol{p}_k - \boldsymbol{x})}. \tag{18}$$

As mentioned in 6.1 singular points $\boldsymbol{x}^*$ are allelic trait vectors for which Eq. 18 equals zero. From Eq. 18 follows that in our model singular points have to fulfil

$$\Big(\sum_{k=1}^{m}\Sigma_k^{-1}\Big)\boldsymbol{x}^* = \sum_{k=1}^{m}\Sigma_k^{-1}\boldsymbol{p}_k.$$

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

$$\boxed{\boldsymbol{x}^* \;=\; \Big(\sum_{k=1}^{m} \Sigma_k^{-1}\Big)^{-1} \sum_{k=1}^{m} \Sigma_k^{-1} \boldsymbol{p}_k := \bar{\boldsymbol{p}}_{\mathrm{w}}}. \tag{19}$$

Thus, the unique singular point $\boldsymbol{x}^*$ equals the arithmetic mean of the pathogen optima $\boldsymbol{p}_1, \ldots \boldsymbol{p}_m$, 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

$$x^* \;=\; \frac{\sum_{k=1}^{m} \sigma_k^{-2} p_k}{\sum_{k=1}^{m} \sigma_k^{-2}} := \bar{p}_{\mathrm{w}}, \tag{20}$$

which is the well-known weighted average for scalars. If $\Sigma_1 = \ldots = \Sigma_m$, then Eq. 19 simplifies to the arithmetic mean pathogen optimum

$$\boldsymbol{x}^* \;=\; \frac{1}{m} \sum_{k=1}^{m} \boldsymbol{p}_k := \bar{\boldsymbol{p}} \tag{21}$$

as stated in the main text.

7.2 Derivation of the Hessian matrix

As stated in Eq. 6, the entries of the Hessian matrix are given by

$$h_{ij} \;=\; \frac{\partial^2 w(\boldsymbol{y}, \boldsymbol{x}^*)}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} \;=\; \frac{1}{s(\boldsymbol{x}^*, \boldsymbol{x}^*)} \frac{\partial^2 s(\boldsymbol{y}, \boldsymbol{x}^*)}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*}. \tag{22}$$

The second derivative of the function s is obtained by differentiating Eq. 14 with respect to y_j resulting in

$$\begin{aligned}
\frac{\partial^2 s(\boldsymbol{y}, \boldsymbol{x}^*)}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} &= K \frac{\partial}{\partial y_j}\Bigg(\frac{1}{\big(K + c(\boldsymbol{y}, \boldsymbol{x})\big)^2} \frac{\partial c(\boldsymbol{y}, \boldsymbol{x})}{\partial y_i}\Bigg)\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} \\
&= \frac{K}{\big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)^4}\Bigg(\frac{\partial^2 c(\boldsymbol{y}, \boldsymbol{x}^*)}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} \big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)^2 - \frac{\partial c(\boldsymbol{y}, \boldsymbol{x}^*)}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} \frac{\partial c(\boldsymbol{y}, \boldsymbol{x}^*)^2}{\partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*}\Bigg) \\
&= \frac{K}{\big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)^2} \frac{\partial^2 c(\boldsymbol{y}, \boldsymbol{x}^*)}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*},
\end{aligned} \tag{23}$$

where in the final simplification step we use the conclusion drawn from Eq. 16 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 Eq. 15 with respect to y_j resulting in

$$\begin{aligned}
\frac{\partial^2 c(\boldsymbol{y}, \boldsymbol{x}^*)}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} &= \frac{\partial}{\partial y_j}\Bigg(c(\boldsymbol{y}, \boldsymbol{x}^*) \sum_{k=1}^{m} \Big(\frac{1}{e_k(\boldsymbol{y}) + e_k(\boldsymbol{x}^*)} \frac{\partial e_k(\boldsymbol{y})}{\partial y_i}\Big)\Bigg)\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} \\
&= c(\boldsymbol{x}^*, \boldsymbol{x}^*) \sum_{k=1}^{m} \frac{\partial}{\partial y_j}\Bigg(\frac{1}{e_k(\boldsymbol{y}) + e_k(\boldsymbol{x}^*)} \frac{\partial e_k(\boldsymbol{y})}{\partial y_i}\Bigg)\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} \\
&= c(\boldsymbol{x}^*, \boldsymbol{x}^*) \sum_{k=1}^{m} \frac{1}{4 e_k(\boldsymbol{x}^*)^2}\Bigg(2 \frac{\partial^2 e_k(\boldsymbol{y})}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} e_k(\boldsymbol{x}^*) - \frac{\partial e_k(\boldsymbol{y})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*} \frac{\partial e_k(\boldsymbol{y})}{\partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}^*}\Bigg).
\end{aligned} \tag{24}$$

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 Eq. 17 with respect to y_j resulting in

$$\begin{aligned}
\frac{\partial^2 e_k(\boldsymbol{y})}{\partial y_i \partial y_j} &= \frac{\partial}{\partial y_j} e_k(\boldsymbol{y}) \sum_{l=1}^{h} \sigma_{k\,il}^{-1}(p_{kl} - y_l) \\
&= \frac{\partial e_k(\boldsymbol{y})}{\partial y_j} \sum_{l=1}^{h} \sigma_{k\,il}^{-1}(p_{kl} - y_l) \;-\; e_k(\boldsymbol{y}) \Big(\sum_{l=1}^{h} \sigma_{k\,il}^{-1} \frac{\partial y_l}{\partial y_j} \Big) \\
&= \frac{1}{e_k(\boldsymbol{y})} \frac{\partial e_k(\boldsymbol{y})}{\partial y_i} \frac{\partial e_k(\boldsymbol{y})}{\partial y_j} - e_k(\boldsymbol{y}) \sigma_{k\,ij}^{-1},
\end{aligned} \tag{25}$$

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 Eqs. 23-25 into Eq. 6 we obtain

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

where in the last step we substituted Eq. 17. This result can be rewritten as a matrix

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

Finally, substituting $\boldsymbol{x}^*$ with Eq. 19 we obtain

$$\boxed{\mathsf{H} = \frac{K}{4\big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)} \Bigg(\sum_{k=1}^{m} \Big(\Sigma_k^{-1} \big(\boldsymbol{p}_k - \bar{\boldsymbol{p}}_w\big) \big(\boldsymbol{p}_k - \bar{\boldsymbol{p}}_w\big)^{\mathrm{T}} \Sigma_k^{-1} \Big) \;-2 \sum_{k=1}^{m} \Sigma_k^{-1} \Bigg)}. \tag{26}$$

7.3 Derivation of the Jacobian matrix

From Eqs. 6 and 7 we can conclude 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

$$\begin{aligned}
q_{ij} &= \frac{\partial}{\partial x_j} \Big(\frac{1}{s(\boldsymbol{x}, \boldsymbol{x})} \frac{\partial s(\boldsymbol{y}, \boldsymbol{x})}{\partial y_i} \Big)\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*} \\
&= \frac{1}{s(\boldsymbol{x}^*, \boldsymbol{x}^*)^2} \Bigg(\frac{\partial^2 s(\boldsymbol{y}, \boldsymbol{x})}{\partial y_i \partial x_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*} s(\boldsymbol{x}^*, \boldsymbol{x}^*) - \frac{\partial s(\boldsymbol{x}, \boldsymbol{x})}{\partial x_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*} \frac{\partial s(\boldsymbol{y}, \boldsymbol{x})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*} \Bigg) \\
&= \frac{1}{s(\boldsymbol{x}^*, \boldsymbol{x}^*)} \frac{\partial^2 s(\boldsymbol{y}, \boldsymbol{x})}{\partial y_i \partial x_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*},
\end{aligned} \tag{27}$$

where in the final simplification step we use the conclusion drawn from Eq. 13 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 Eq. 14 with respect to x_j resulting in

$$\begin{aligned}
\frac{\partial^2 s(\boldsymbol{y},\boldsymbol{x})}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*} &= K\frac{\partial}{\partial x_j}\Big(\frac{1}{\big(K+c(\boldsymbol{y},\boldsymbol{x})\big)^2}\frac{\partial c(\boldsymbol{y},\boldsymbol{x})}{\partial y_i}\Big)\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*} \\
&= \frac{K}{\big(K+c(\boldsymbol{x}^*,\boldsymbol{x}^*)\big)^4}\Big(\frac{\partial^2 c(\boldsymbol{y},\boldsymbol{x})}{\partial y_i \partial x_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}\big(K+c(\boldsymbol{x}^*,\boldsymbol{x}^*)\big)^2 - \frac{\partial c(\boldsymbol{y},\boldsymbol{x})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}\frac{\partial c(\boldsymbol{y},\boldsymbol{x})^2}{\partial x_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}\Big) \\
&= \frac{K}{\big(K+c(\boldsymbol{x}^*,\boldsymbol{x}^*)\big)^2}\frac{\partial^2 c(\boldsymbol{y},\boldsymbol{x})}{\partial y_i \partial x_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}, \qquad (28)
\end{aligned}$$

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 Eq. 15 with respect to x_j resulting in

$$\begin{aligned}
\frac{\partial^2 c(\boldsymbol{y},\boldsymbol{x}^*)}{\partial y_i \partial y_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*} &= \frac{\partial}{\partial x_j}\Big(c(\boldsymbol{y},\boldsymbol{x})\sum_{k=1}^{m}\Big(\frac{1}{e_k(\boldsymbol{y})+e_k(\boldsymbol{x})}\frac{\partial e_k(\boldsymbol{y})}{\partial y_i}\Big)\Big)\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*} \\
&= c(\boldsymbol{x}^*,\boldsymbol{x}^*)\sum_{k=1}^{m}\frac{\partial}{\partial x_j}\Big(\frac{1}{e_k(\boldsymbol{y})+e_k(\boldsymbol{x})}\frac{\partial e_k(\boldsymbol{y})}{\partial y_i}\Big)\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*} \\
&= -c(\boldsymbol{x}^*,\boldsymbol{x}^*)\sum_{k=1}^{m}\frac{1}{4e_k(\boldsymbol{x}^*)^2}\frac{\partial e_k(\boldsymbol{y})}{\partial y_i}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}\frac{\partial e_k(\boldsymbol{x})}{\partial x_j}\Big|_{\boldsymbol{y}=\boldsymbol{x}=\boldsymbol{x}^*}. \qquad (29)
\end{aligned}$$

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

By recursively substituting Eqs. 28 and 29 into Eq. 27 we obtain

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

where in the last step we substituted Eq. 17. This result can be rewritten in matrix form as

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

Finally, substituting $\boldsymbol{x}^*$ with Eq. 19 we obtain

$$\boxed{\mathsf{Q} = -\frac{1}{4}\frac{K}{K+c(\boldsymbol{x}^*,\boldsymbol{x}^*)}\sum_{k=1}^{m}\Big(\Sigma_k^{-1}\big(\boldsymbol{p}_k-\bar{\boldsymbol{p}}_w\big)\big(\boldsymbol{p}_k-\bar{\boldsymbol{p}}_w\big)^{\mathrm{T}}\Sigma_k^{-1}\Big)}. \qquad (30)$$

Combining Eqs. 26 and 30) and simplifying results in the Jacobian matrix we obtain

$$\mathsf{J} = \mathsf{H}+\mathsf{Q} = -\frac{1}{2}\frac{K}{K+c(\boldsymbol{x}^*,\boldsymbol{x}^*)}\sum_{k=1}^{m}\Sigma_k^{-1}. \qquad (31)$$

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 (2009).

7.4 Special case: Identically shaped pathogen optima

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

$$\begin{aligned} \mathsf{H} &= \frac{Km}{4\big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)} \Sigma^{-1} \left(\frac{1}{m} \sum_{k=1}^{m} \Big(\big(\boldsymbol{p}_k - \bar{\boldsymbol{p}}\big) \big(\boldsymbol{p}_k - \bar{\boldsymbol{p}}\big)^{\mathrm{T}} \Big) - 2\Sigma \right) \Sigma^{-1} \\ &= \frac{Km}{4\big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)} \Sigma^{-1} \Big(\mathsf{V} - 2\Sigma \Big) \Sigma^{-1}, \end{aligned} \tag{32}$$

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 task optima. Since the ratio in Eq. 32 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 Fig. 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. Horn and Johnson 2013, 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 optima 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 $\Sigma = \sigma^2 \mathrm{I}$ and that the pathogen optima are arranged in a symmetric manner resulting in $\mathsf{V} = v^2 \mathrm{I}$ we obtain

$$\mathsf{H} = \frac{Km}{4\sigma^{-4}\big(K + c(\boldsymbol{x}^*, \boldsymbol{x}^*)\big)} \big(v^2 - 2\sigma^2\big)\mathrm{I}. \tag{33}$$

In this case, the singular point $\boldsymbol{x}^*$ is a branching point whenever $v^2 > 2\sigma^2$. Assuming that all m pathogen optima have an equal distance d to each other implies that they are arranged in a $m-1$-dimensional regular simplex (see Fig. 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.

Efficient binding of an MHC molecule to a pathogenic antigen is a high dimensional optimisation problem with many degrees of freedom. Here, we assume that in the presence of two pathogens, this problem reduces to an optimisation problem along a linear manifold that connects the two optima corresponding to the two pathogens. Similarly, in the presence of three pathogens the optimisation problem reduces to a two-dimensional manifold, and, more generally, in the presence of m pathogens to a (m-1)-dimensional manifold. This assumption seems reasonable as long as the number of pathogens does not exceed the number of dimensions that is needed to describe match between a MHC molecule and a pathogenic antigen. The only consequence of restricting the analysis in this manner is that directions in which mutations are deleterious, because they cause maladaptations for all pathogens, are omitted from the analysis. This effects the speed of the evolutionary dynamics but not the final number of coexisting alleles. The assumption that m pathogens are arranged in a (m-1)-dimensional regular simplex where any two pathogen optima are separated by the same distance d imposes maximal symmetry and facilitates analytical calculations.

References

Abrams, P. A., H. Matsuda, and Y. Harada. 1993. Evolutionarily unstable fitness maxima and stable fitness minima of continuous traits. Evolutionary Ecology 7:465–487.

Champagnat, N., R. Ferrière, and S. Méléard. 2006. Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. Theoretical Population Biology 69:297–321.

Débarre, F., S. L. Nuismer, and M. Doebeli. 2014. Multidimensional (co) evolutionary stability. The American Naturalist 184:158–171.

Dercole, F., and S. Rinaldi. 2008. Analysis of evolutionary processes: the adaptive dynamics approach and its applications: the adaptive dynamics approach and its applications. Princeton University Press.

Dieckmann, U., and R. Law. 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. Journal of Mathematical Biology 34:579–612.

Doebeli, M. 2011. Adaptive Diversification (MPB-48). Princeton University Press.

Durinx, M., J. A. J. H. Metz, and G. Meszéna. 2008. Adaptive dynamics for physiologically structured population models. Journal of Mathematical Biology 56:673–742.

Fisher, R. A. 1930. The genetical theory of natural selection. Oxford University Press.

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

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

Gillespie, J. H. 2004. Population Genetics: A Concise Guide. The Johns Hopkins University Press.

Horn, R. A., and C. R. Johnson. 2013. Matrix analysis, 2nd. New York: Cambridge University Press.

Iwasa, Y., A. Pomiankowski, and S. Nee. 1991. The evolution of costly mate preferences ii. the handicap principle. Evolution 45:1431–1442.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. Evolution 33:402–416.

Leimar, O. 2009. Multidimensional convergence stability. Evolutionary Ecology Research 11:191–208.

Metz, J. A. J. 2008. Fitness. Pages 1599–1612 *in* S. E. Jorgensen and B. Fath, eds. Encyclopedia of ecology. Oxford: Elsevier, available online as IIASA Interim Report IR06061.

Metz, J. A. J., and C. G. F. de Kovel. 2013. The canonical equation of adaptive dynamics for Mendelian diploids and haplo-diploids. Interface Focus 3.

Metz, J. A. J., R. M. Nisbet, and S. A. H. Geritz. 1992. How should we define fitness for general ecological scenarios? Trends in Ecology & Evolution 7:198–202.

Rueffler, C., T. J. M. Van Dooren, O. Leimar, and P. A. Abrams. 2006. Disruptive selection and then what? Trends in Ecology & Evolution 21:238–245.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97–159.