

Online Appendices to “The Evolution of Genetic Architecture under Frequency-dependent Disruptive Selection”

Michael Kopp

Joachim Hermisson

Appendix 2: Simulations

In this Appendix, we describe the numerical simulations.

General procedure.— We used deterministic simulations of haplotype frequencies, coupled with stochastic creation and extinction of modifier alleles. A haplotype is defined by the allelic values at both the primary and modifier loci, with the alleles at the modifier loci determining the mutational effects of the primary loci. We assume one modifier locus per primary locus. In principal, there are infinitely many modifier alleles and, therefore, infinitely many haplotypes. In order to keep the number of haplotypes finite we only allow a fixed number k of alleles per modifier locus to be simultaneously present in the population (between 2 and 20, depending on the number of loci).

Our simulated haplotypes undergo a repeated sequence of selection, segregation (in the diploid case), recombination, and mutation. More precisely, for each genotype, we first calculate the phenotype g , and then apply equation (5) or (6) to obtain the corresponding fitness values. To get haplotype frequencies after selection and recombination, we use the standard recursion relation

$$p'_r = \bar{W}^{-1} \sum_{s,t} W_s W_t p_s p_t R(st \rightarrow r) \quad (\text{A21a})$$

in the haploid case and

$$p'_r = \bar{W}^{-1} \sum_{s,t} W_{st} p_s p_t R(st \rightarrow r) \quad (\text{A21b})$$

in the diploid case. Here, haplotypes are labeled r , s , and t , and p_r and p'_r are the haplotype frequencies before and after selection and recombination, respectively. W_x is the fitness of a haploid individual with genotype x , and W_{xy} is the fitness of a diploid individual

containing haplotypes x and y . $R(st \rightarrow r)$ is the rate at which recombination events between haplotypes s and t produce haplotype r . The $R(st \rightarrow r)$'s are calculated from the recombination rates between pairs of adjacent loci, which are model parameters, and from the number of recombination events necessary to create the target genotype. Between each pair of adjacent loci, at most one recombination event can occur, and recombination events between different pairs of adjacent loci are assumed to be independent. Modifier loci are adjacent to their respective primary loci. Population sizes follow the recursion relation

$$N' = N\bar{W}. \quad (\text{A22})$$

Mutation is handled differently for primary loci and modifier loci. Primary locus alleles mutate from A_i to a_i and vice versa at rate 10^{-5} per individual and generation. Their allele frequencies change deterministically according to this rate. In contrast, mutations at modifier loci are modeled stochastically. If, at a modifier locus, less than k alleles are present in the population, one of them can give rise to a new, mutant allele with probability 0.01 per generation (note that this probability refers to the whole population, not to each single individual). The allelic value of the mutant allele is drawn from a normal distribution with mean equal to the value of the source allele and variance V_m . Mutations leading to a negative value of γ_i are rejected. The initial frequency of the new allele is set to $10/\tau N$, corresponding to 10 initial copies in the population. An allele of this frequency has a negligible chance of being lost due to genetic drift. In other words, we only consider alleles which have survived the initial threat of stochastic extinction. Once a modifier allele has entered the population, its subsequent dynamics are determined by the deterministic equations (A21). The only exception is extinction. A modifier allele with frequency μ can go extinct with probability $(1 - \mu)^{\tau N}$, which is the probability that it is not present in any of the τN haplotypes of the next generation.

Technical assumptions.— As described above, the simulations are very time consuming. In order to increase computation speed, we made several technical assumptions:

First, the time required for the recombination step in equations (A21) is of order $O(z^3)$, where z is the number of possible haplotypes. In simulations with free recombination ($r = 0.5$), this can be reduced to $O(z)$ by making the simplifying assumption that recombination always restores complete linkage equilibrium. Therefore, in many simulations, instead of using equations (A21), we calculated haplotype frequencies after selection and mutation, used these to extract allele frequencies, and then calculated haplotype frequencies at linkage equilibrium as the products of the respective allele frequencies. We will refer to this simplification as the linkage equilibrium assumption.

Second, in the full model (eq. 5), the time required for the computation of $\bar{\alpha}_\pi(g)$ is of order $O(z^2)$ in the haploid case and $O(z^4)$ in the diploid case. These relationships can be reduced to $O(z)$ and $O(z^2)$, respectively, if on the right-hand side of equation (2b), the average is taken not over all possible phenotypes, but only over the means of 20 equally

spaced phenotypic classes (like those in a histogram). The effect of this simplification on the resulting fitness values is negligible.

Third, we did not allow mutations in modifier loci that resulted in γ_i values too similar to those of existing alleles. More precisely, mutant alleles with a distance less than $\sqrt{V_m}/4$ from a previously existing allele were assigned a value at exactly this distance from the existing allele. In cases where the new allele fell in between two previously existing alleles with distance less than $\sqrt{V_m}/2$, its distance from the nearer of the two existing alleles was set to $1/4$ of the distance between them.

Fourth, instead of explicitly modeling mutations at primary loci, we only prevented allele frequencies from becoming less than 10^{-5} . This assumption has negligible effects provided selection on a locus is not extremely weak (i.e., unless the locus effect is very close to 0.)

The weak-selection approximation with constrained phenotypic range.—

The assumptions about the mutation process laid down in equations (16) were implemented as follows: We introduce “internal” variables $\hat{\gamma}_i$ which are assumed to be directly coded for by the modifier loci (one per primary locus). The $\hat{\gamma}_i$ can evolve independently, but they interact epistatically to produce the locus effects γ_i , according to

$$\gamma_i = \phi(\hat{\Gamma}) \frac{\hat{\gamma}_i}{\hat{\Gamma}} \quad (\text{A23a})$$

with

$$\hat{\Gamma} = \sum_i \hat{\gamma}_i \quad (\text{A23b})$$

and

$$\phi(\hat{\Gamma}) = \frac{\Gamma_{\max} \hat{\Gamma}}{\beta + \hat{\Gamma}}. \quad (\text{A23c})$$

where $\beta > 0$. We always used $\Gamma_{\max} = 1$ and $\beta = 1$. Other fixed parameter values were: $k = 2$, $a = 0.2$, $V_m = 0.04$, $\rho = 2$, $\kappa = 10^4$.

For the case with free recombination, we investigated the following parameter combinations: weak versus strong frequency-dependence ($f = 2$ versus $f = 10$), symmetric versus asymmetric stabilizing selection ($\theta = 0$ versus $\theta = 0.5$), and haploid versus diploid genetics with 2, 3, or 4 primary loci. For the four-locus cases, we used the linkage equilibrium assumption. All simulations were started with equal locus effects (i.e., a symmetric genetic architecture): either $\hat{\gamma}_i = 0.01$ (leading to a population with low initial phenotypic variance) or $\hat{\gamma}_i = 4/(\tau n)$ (leading to a population with high initial phenotypic variance). In the latter case, the sum of initial (mean) locus effects, $\sum \bar{\gamma}_j = 0.8\Gamma_{\max}$. The simulations were stopped when the mean effect of the strongest locus alone exceeded $0.8\Gamma_{\max}$. (Note that, due to eq. (A23c), an effect of $1.0\Gamma_{\max}$ can be reached only asymptotically.) This

condition was always reached, and when it was reached, the mean effects of all other loci were always less than $0.11\Gamma_{\max}$ and in many cases, less than $0.01\Gamma_{\max}$. Simulations with linkage were done only for the haploid case. We tested a range of values of r (0.01, 0.02, 0.03, 0.04, 0.05, 0.075, and 0.1) for $n = 2$ with $\theta = 0$ or $\theta = 0.5$ and $f = 2$ or $f = 10$, and for $n = 2$ with $\theta = 0$ and $f = 10$. Figure 2 shows the results for $n = 3$, $\theta = 0$, and $f = 10$.

The full model.— In simulations of the full model (eq. 5), we assumed that the effect of each locus was determined exclusively by the allelic values at the corresponding modifier locus. We usually used the linkage equilibrium assumption. As described in the main text, linkage disequilibrium is negligible for $r > 0.05$. Most simulations (including the ones presented in the main text) were started with equal locus effects of $\gamma_i = 0.01$ (i.e., assuming a low initial phenotypic variance). In some limited simulations, the initial locus effects were higher, but generically, we found the same results as those presented in the main text.

An important conclusion from our simulations is that an asymmetric genetic architecture can evolve in a reasonably short time-span (less than 5000 generations). Of course, the rate of evolution depends strongly on the mutation rate and on the variance of mutational effects at the modifier loci. Regarding mutation rates, we assumed that new alleles can invade the population (without being immediately lost by drift) at a rate of 0.01 per generation. For comparison, according to standard theory, the rate of fixation of beneficial alleles under stabilizing selection is approximately $\tau NusV_m$, with u being the mutation rate. With N between 10^4 and 10^5 (Fig. 8), u between 10^{-4} and 10^{-5} , $s = 0.1$ and $V_m = 0.04$, this is quite exactly in the order of 10^{-2} . As we only allow new mutations when there are less than k resident alleles, the effective mutation rate in our simulations is clearly conservative. Regarding the variance of mutational effects, we usually used $V_m = 0.04$. This means that the width of the corresponding normal distribution is about a fifth of the width of the stabilizing selection function (for $s = 0.1$), which is well in line with empirical estimates reviewed by Bürger (2005, pp. 263-267). In all simulations, we assumed $\rho = 2$, $\kappa = 10^4$, and $s = 0.1$.

References

- Bürger, R., 2005. A multilocus analysis of intraspecific competition and stabilizing selection on a quantitative trait. *J. Math. Biol.* 50:355–396.

Appendix 3: Analysis of the full model with asexual reproduction

The equilibrium phenotypic distribution in the asexual version of the full model (eq. 6) can be determined without considering explicit genetics. The task is to find a vector $\mathbf{g} = (g_1, \dots, g_k)$ of k coexisting phenotypes that satisfy the following three conditions: First, each phenotype is at population-dynamic equilibrium; that is, $W(g_i) = 1$ for $i = 1 \dots k$. This condition yields the equilibrium numbers $N_i = N\pi(g_i)$ of individuals with genotype g_i as solutions of a linear system of equations. Second, each phenotype corresponds to a fitness optimum, that is the selection gradient $\partial W(g_i)/\partial g_i = 0$ for $i = 1 \dots k$. This condition can be solved numerically. Third, no other phenotype can invade the population; that is, $W(h) < 1$ for all h that are not elements of \mathbf{g} . This condition defines the number of phenotypes, k . In practical terms, the bifurcation points shown in Figure 4 can be computed exactly by making the conjecture (which can be supported numerically) that, with increasing f , additional phenotypes always appear at $g_{\text{new}} = \theta$: When there was previously no $g_i = \theta$, $g_{\text{new}} = \theta$ invades if $W(g = \theta) = 1$. If a $g_i = \theta$ exists, it bifurcates into two new phenotypes if the second derivative $\partial^2 W(g = \theta)/\partial g^2 = 0$, that is if the point $g = \theta$ turns from a fitness maximum to a fitness minimum.

Appendix 4: Analysis of the full model with asymmetric stabilizing selection and linkage

In the main text, we have only analyzed cases with symmetric stabilizing selection and free recombination. However, our main results – in particular, that of an asymmetric genetic architecture with a ratio of locus effects of about 1 : 2 : 4... – also hold true for more general conditions. With asymmetric stabilizing selection, typically one or more primary loci become fixed and the combined effect of these loci evolves to a value equal to the optimal phenotype θ . In consequence, the population mean phenotype equals θ , and selection on the effects of the remaining polymorphic loci is the same as in the symmetric case. The only exceptions occur if the number of primary loci, n , is small and f is large. In this case, there is a trade-off between fixing loci to shift to the right the mean phenotype towards the optimum and keeping all loci polymorphic to increase the number of phenotypes. Selection may then favor the latter option, leading to an asymmetric phenotypic distribution with allele frequencies at the primary loci being greater than 0.5 but the mean phenotype $\bar{g} < \theta$ (assuming $\theta > 0$; results not shown).

We also tested the effects of linkage on the outcome of evolution. Significant linkage disequilibrium builds up only if the recombination rate between adjacent loci, r , is less than about 0.05. In these cases, the genetic architecture is less asymmetric than in the cases with free recombination, but the overall pattern still holds true (results not shown).