The diverse effects of phenotypic dominance on hybrid fitness

Hilde Schneemann[1,†,*], Aslı D. Munzur[2,*], Ken A. Thompson[2,3,⋆], and John J. Welch[1,⋆]

February 7, 2022

1. Department of Genetics, University of Cambridge, Downing Street, Cambridge, UK, CB23EH

2. Department of Zoology & Biodiversity Research Centre, University of British Columbia, Vancouver, Canada

3. Present address: Department of Biology, Stanford University & Department of Plant Biology, Carnegie Institution for Science, Stanford, CA, USA, 94305

† To whom correspondence should be addressed (hilde.schneemann@evobio.eu).

∗, ⋆ Equal contributions.

Abstract

When divergent populations interbreed, their alleles are brought together in hybrids. These hybrids may express novel phenotypes, not previously exposed to selection. In the initial F1 cross, most divergent alleles are present as heterozygotes. Therefore, F1 fitness can be influenced by dominance effects that first appear together in the hybrids, and so could not have been selected to function well together. We present a systematic study of these F1 dominance effects by introducing variable phenotypic dominance into Fisher's geometric model. We show that dominance often reduces hybrid fitness, which can lead to patterns of optimal outbreeding and a steady decline in F1 fitness at high levels of divergence. We also show that "lucky" beneficial effects sometimes arise by chance, which might be especially important when hybrids can access novel environments. We then explore the interaction of phenotypic dominance with uniparental inheritance, showing that dominance can lead to violations of Haldane's Rule (reduced fitness of the heterogametic sex) while strengthening Darwin's Corollary (fitness differences between cross direc-

tions). Taken together, our results show that dominance could play an important role in the outcomes of hybridisation after secondary contact, and thus to the maintenance or collapse of isolating barriers. Nevertheless, the telltale signs of dominance are relatively few and subtle. Results also suggest that dominance effects are smaller than the cost of segregation variance, implying that simple additive models may still give good predictions for later-generation recombinant hybrids, even when dominance qualitatively alters outcomes for the F1.

Keywords:

Speciation, Fisher's geometric model, heterosis, optimal outbreeding, Haldane's Rule, Darwin's corollary

Introduction

During hybridisation alleles from diverged genomes can be expressed together for the first time. The interactions between these alleles will help to determine the outcome of the hybridisation. If hybrids are sufficiently fit, for example, then the hybridisation might lead to ongoing gene flow (e.g. Rieseberg et al., 1999; Lee et al., 2013), or to a hybrid swarm, potentially adapted to a new niche (e.g. Taylor et al., 2006). If hybrids are are unfit, by contrast, the lineages will remain reproductively isolated, and may be subject to reinforcement selection. All hybrids trace their ancestry to a first-generation, or 'F1' hybrid, and so the fitness of the F1 has a special importance. Ongoing gene flow, for example, is only possible if the F1 are both viable and fertile (Wallace, 1889; Fisher, 1930; Dobzhansky, 1937; Blair, 1955; Butlin, 1987). Moreover, the F1 are often easier to form than later generation crosses (Muller, 1940; Vetukhiv, 1954), and so have been more extensively studied.

Data from F1 hybrids show some recurring patterns, that have been observed consistently across taxa (Arnold and Hodges, 1995; Coyne and Orr, 2004; Fraïsse et al., 2016, Table S1). For example, crosses between closely-related parental lines sometimes exhibit "optimal outbreeding", where the F1 are fittest between parents of intermediate genetic distance (Bateson, 1978; Butlin, 1987; Waser, 1993; Wei and Zhang, 2018; Dagilis et al., 2019); while at higher divergences, data often show an "F1 clock", where F1 fitness declines steadily with genetic distance (Edmands, 2002; Coughlan and Matute, 2020). Perhaps best known is Haldane's Rule, which describes the fitness inferiority of heterogametic F1 (Haldane, 1922; Coyne and Orr, 2004, Ch. 8; Schilthuizen et al., 2011); but equally well supported is "Darwin's Corollary", which describes strong fitness differences between the reciprocal F1 (i.e. female-male vs. male-female cross directions of the same parental lines; Darwin, 1859, Ch. 8; Kölreuters, 1766; Turelli and Moyle, 2007).

A class of fitness landscapes based on Fisher's geometric model (Fisher, 1930; Orr, 1998) has been successful in generating many of these empirical patterns (Mani and Clarke, 1990; Barton, 2001; Chevin et al., 2014; Fraïsse et al., 2016; Simon et al., 2018; Schneemann et al., 2020). However, the model makes some unrealistic predictions for the F1; in particular the simplest form of the model predicts F1 heterosis (i.e., hybrid fitness advantage) at all levels of genetic divergence (Barton, 2001; Fraïsse et al., 2016). One explanation for this incorrect prediction is that most studies of Fisher's model assume phenotypic additivity—that is, they assume that the phenotypic effect of a homozygous allele is exactly double its heterozygous effect. By contrast, empirical studies indicate that variable levels of dominance are common for loci affecting quantitative traits, including fitness components (Lynch and Walsh, 1998, p.485). Given its high level of heterozygosity, the F1 will be especially affected by the effects of this variable dominance.

Here, building on previous work (Schneemann et al., 2020), we present a systematic exploration of the effects of variable phenotypic dominance on hybrid fitness, with a particular focus on the F1. We show that the effects of dominance are highly diverse, but can nonetheless be divided into two broad categories: predictable deleterious effects, and "lucky" beneficial effects, which occur only occasionally and unpredictably. We then explore the interactions of dominance with uniparental inheritance or expression (as is essential to Haldane's Rule and Darwin's Corollary). We show that the two classes of effect act in similar ways, but that their interactions are sometimes synergistic and sometimes antagonistic. We end by discussing the overall importance of dominance to the outcome of hybridization.

Model

Fisher's (1930) geometric model assigns fitnesses to genotypes using a simple model of n quantitative traits under optimizing selection. The parameter n, which is sometimes called "organismal complexity", is also the dimensionality of the fitness landscape. An individual's phenotype is represented by a point in this landscape, with its trait values collected in the n-dimensional vector, $\mathbf{z}$. The relative fitness of the individual is denoted as w, and depends on the Euclidean distance of the phenotype from some optimum $\mathbf{o}$, whose position is determined by the current environmental conditions. Here, we will use a simple quadratic model, where log fitness declines with the squared distance to the optimum:

$$\ln w = -\left\|\mathbf{z} - \mathbf{o}\right\|^2 \tag{1}$$

$$\equiv -\sum_{i=1}^{n}(z_i - o_i)^2 \tag{2}$$

Substitutions are modelled as n-dimensional vectors of change to the phenotype. Thus, any two phenotypes can be connected by a chain of vectors that represent the divergent alleles accrued since their most recent common ancestor. This is illustrated in Figure 1a, where we show a chain of $d = 6$ genomic differences connecting two homozygous parental lines, labelled P1 and P2. In the illustration, each population fixed equal numbers of substitutions, but this is not an assumption of the analyses. What is important, however, is that all substitutions are defined relative to the P1 genotype. This means that none of our results depend on knowing whether the P1 alleles are ancestral or derived.

Any given hybrid between P1 and P2 will contain some combination of the d vectors. Under phenotypic additivity, the effects of these vectors sum together both between and within loci. This means that an allele carried as a heterozygote will have the same orientation, but half the length, of the allele carried as a homozygote. This is shown by the yellow arrow in Figure 1a inset. To model phenotypic dominance, we relax the assumption of additivity within loci, using dominance coefficients δ_{ij}. This is shown by the dark red arrow in Fig. 1a inset. Non-zero values of the δ_{ij} can alter both the length of the vector, and its orientation if different traits show different levels of dominance.

We note that the very simple model described above can be derived, either exactly or approximately, from a large family of more complex fitness functions (Martin and Lenormand, 2006; Martin, 2014). In that case, few if any of the n traits correspond to real quantitative traits, so that n becomes a phenomenological parameter of the fitness landscape (Orr, 2000; Welch and Waxman, 2003; Martin and Lenormand, 2006).

Results

To explore the various effects of dominance on the fitness of F1 hybrids, we consider several different scenarios and patterns. As such, throughout the results section, we present the major results without derivation, and relegate the full derivations to Appendix 1. Simulations are used solely to illustrate the analytical results.

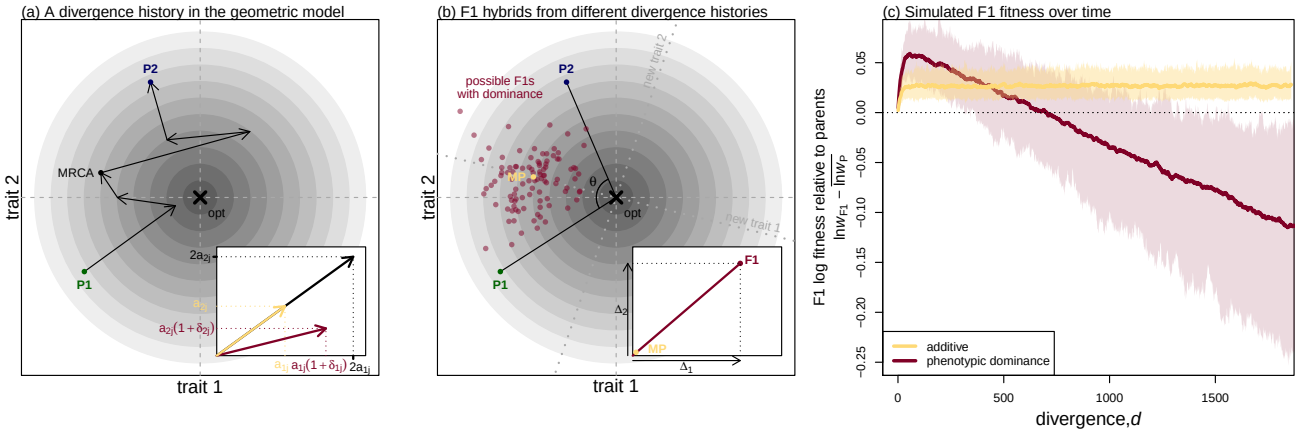

Figure 1: The effects of variable phenotypic dominance on F1 hybrid phenotypes and fitness in Fisher's geometric model. **(a)-(b)**: illustrations of the fitness landscape model with $n = 2$ traits for visualisation. Contour lines indicate fitness, with darker colours closer to the optimum ($\times$ symbol). **(a)**: a pair of parental lines, P1 and P2, connected by a chain of $d = 6$ fixed differences. These pass through their most recent common ancestor (MRCA), but are orientated from P1 to P2. The inset shows a single substitution in homozygous state (black arrow), and in heterozygous state either with or without dominance. With additivity (yellow arrow), the heterozygous substitution is simply half of the homozygous substitution. With dominance (dark red arrow), the heterozygous effect on trait i is $a_{ij}(1 + \delta_{ij})$, where a_{ij} is the additive effect, and δ_{ij} the dominance coefficient. **(b)**: parents and possible F1 hybrids. The yellow point (MP) shows the midparental phenotype located halfway between the parents, which is equivalent to the F1 under strictly biparental inheritance and additivity. Phenotypic dominance pushes the F1 phenotype away from the midparent. Dark red points show a cloud of possible F1 that could have resulted from different realisations of the divergence history. The inset panel shows the total dominance deviation on each trait (Δ_i, eq. 4), and dotted lines indicate the new axes after the rotation implied by eq. 5. **(c)**: simulation results illustrating the difference between F1 and parental log fitness over the course of divergence. Under additivity (yellow lines), the F1 maintains a roughly constant level of heterosis. With phenotypic dominance (dark red lines) the F1 shows a pattern of optimal outbreeding with an initial spike in F1 fitness followed by a steady decline (the F1 clock). Lines show the means across 100 replicate simulations and the shading shows 95% quantiles. The simulation procedure assumed that larger effect mutations were more likely to be recessive (Fig. S1), and is described in full in Appendix 2. Parameters were $n = 20$ and $U = \bar{s} = 1/N = 0.01$.

1. The F1 under additivity

With standard biparental inheritance and expression, F1 hybrids contain all d of the divergent alleles as heterozygotes. With complete phenotypic additivity, the F1 phenotype therefore matches the midparental phenotype, obtained by averaging the parental values for all n traits. The midparent is illustrated by the yellow point labelled MP in Figure 1b.

The fitness of the midparental phenotype, w_{MP}, therefore depends solely on the fitnesses of the parental lines, w_{P1} and w_{P2}, and their relative positions in the n-dimensional phenotypic space. This is characterized by θ, the angle in radians between the parental phenotypes (see Fig. 1b). Indeed, from the definition of cosine similarity, we find:

$$\ln w_{\mathrm{MP}} = -\sum_{i=1}^{n}\left(\frac{z_{\mathrm{P1,i}} + z_{\mathrm{P2,i}}}{2} - o_i\right)^2$$

$$= \frac{\overline{\ln w_{\mathrm{P}}} + \overline{\ln w_{\mathrm{P}}}^{(g)} \cos\theta}{2} \tag{3}$$

where $\overline{\ln w_{\mathrm{P}}} \equiv \frac{1}{2}\left(\ln w_{\mathrm{P1}} + \ln w_{\mathrm{P2}}\right)$ and $\overline{\ln w_{\mathrm{P}}}^{(g)} \equiv -\sqrt{(\ln w_{\mathrm{P1}} \times \ln w_{\mathrm{P2}})}$ are respectively the arithmetic and geometric mean log fitness of the parents. Since $\cos\theta$ varies between -1 and 1, it follows that $\ln w_{\mathrm{MP}} \geq \overline{\ln w_{\mathrm{P}}}$. Therefore, with phenotypic additivity, the F1 will always be at least as fit as the average of the fitnesses of the two parental lines (Barton, 2001; Fraïsse et al., 2016).

To understand the consequences of this result, let us consider a concrete situation, in which the two parental lines remain well adapted to a fixed phenotypic optimum but diverge from their common ancestor via "system drift" (Schneemann et al., 2020; Schiffman and Ralph, 2021). In the initial stages of divergence the parental lines fix different alleles, and so they wander slightly in phenotypic space while remaining near the optimum. This wandering tends to increase θ, bringing the midparent closer to the optimum and thereby increasing the potential for heterosis (eq. 3). After this initial divergence, heterosis remains at roughly constant levels because the phenotypic distance has no further tendency to increase. This is confirmed by individual-based simulations of Fisher's geometric model, which are described in detail in Appendix 2 and shown by the yellow lines in Figure 1c. With a fixed optimum and additive phenotypes, F1 heterosis shows a brief initial increase before settling at a constant positive value. The prediction of heterosis at all levels of genetic divergence is perhaps the least realistic prediction of the additive model.

2. The deleterious effects of phenotypic dominance

The principle effect of variable phenotypic dominance is to push the F1 phenotype away from the mid-parent, usually reducing its fitness (Schneemann et al., 2020). To see this, let us first define the total dominance deviation on trait i, including contributions from all d substitutions.

$$\Delta_i \equiv \sum_{j=1}^{d} a_{ij}\delta_{ij} \tag{4}$$

Here, a_{ij} is the additive effect of substitution j on trait i, and δ_{ij} is its dominance coefficient. These quantities are illustrated in the inset panels in Figure 1a and 1b. With this definition, the fitness of the F1 can be written as:

$$\ln w_{\mathrm{F1}} = -\left(\sqrt{|\ln w_{\mathrm{MP}}|} + \Delta_1\right)^2 - \sum_{i=2}^{n} \Delta_i^2 \tag{5}$$

Here, without loss of generality, we have used a simple rotation of the trait axes such that the midparent falls short of the optimum only for trait 1 (as newly defined), but is optimal for the remaining $n-1$ traits. This is shown in Figure 1b.

An important point about eq. 5 is that the Δ_i affect fitness only in hybrids, and so unlike the additive effects, which act together in the P2 genotype, the dominance effects are unlikely to be under selection together during divergence, especially allopatric divergence. As such, they show no tendency to be co-adapted, or remain close to any optimum. This means that similar scenarios of parental divergence, with similar trajectories for the parental phenotypes, can nonetheless yield very different Δ_i and thus very different F1 phenotypes. This is illustrated by the cloud of dark red points in Figure 1b, which show possible F1 that might have resulted from the same divergence scenario between P1 and P2.

To predict F1 fitness, we must average over these possible evolutionary histories. Let us first consider the scenario discussed above, where parental populations remain well adapted to the same constant environment. In this case, we can think of the Δ_i as undergoing a random walk away from the midparent, so that the set of possible F1 form a growing cloud (Schneemann et al., 2020). This yields:

$$E\left(\ln w_{\mathrm{F1}}\right) = E\left(\ln w_{\mathrm{MP}}\right) - nV_{\Delta} \tag{6}$$

$$= \frac{\overline{\ln w_{\mathrm{P}}}}{2} - dnV_{a\delta} \tag{7}$$

where V_Δ is the variance, across divergence histories, of the the total dominance deviations (see eqs. 4 and 25); while $V_{a\delta}$ is the equivalent variance in dominance effects of single substitutions (i.e. the d increments of the random walk; eq. 26). As with the additive model, eqs. 6-7 predict a transient increase in F1 fitness in the early stages of divergence, as long as the second terms remain small. But as the number of substitutions (d) increases, the second terms start to grow, causing a steady fitness decline. To illustrate this, we repeated our simulations after allowing for variable phenotypic dominance among the new mutations (see Appendix 2 for full details). The results, shown as dark red curves in Figure 1c, show a clear pattern of transient heterosis (i.e., optimal outbreeding) followed by a linear decline in log fitness (the F1 clock). The maximal heterosis is also higher with dominance, due to the parental lines fixing more deleterious recessives.

We can also think of the results above as exemplifying two regimes. In the early stages of divergence, the phenotypic distance between the parents increases faster than the cloud of possible F1. This means that most F1 trait values will be intermediate between the parental values. Later, the parents stop diverging phenotypically (because of effective stabilizing selection acting on their traits), but the cloud of possible F1 continues to expand. This leads to transgressive trait variation, with the F1 values lying outside of the range of the parental values. It is notable that both of these regimes have been observed in quantitative traits from F1 hybrids: transgressive variation sometimes decreases with d, and sometimes increases with d (Stelkens and Seehausen, 2009, see also Appendix 1 and Figure S2 for more details).

2.1. The tick rate of the F1 clock

Equation 7 gives the tick rate of the F1 clock as $nV_{a\delta}$, which is also the rate of expansion of the cloud of possible F1 phenotypes. This quantity summarizes the distribution of factors fixed, which, as previous work has shown, will depend on many different aspects of the system's biology (e.g. Charlesworth, 1992; Orr, 1998; Griswold, 2006; Yeaman and Whitlock, 2011; Matuszewski et al., 2014; Schneemann et al., 2020; Yamaguchi and Otto, 2020).

One important determinant of the tick rate is likely to be the distribution of dominance coefficients of new mutations. This is easiest to see with a simpler model of mutation than that used in Fig. 1c. If we have isotropic traits i.e. equal distributions of effects across traits, and no tendency for substitutions with larger additive effects to have more extreme dominance coefficients (e.g. no covariance between a_{ij}^2 and δ_{ij}^2), then we have

$$V_{a\delta} = V_a V_\delta \tag{8}$$

where V_a and V_δ are the variance of the additive effects and dominance coefficients respectively. Furthermore, with this model, V_δ for the fixed effects will often match the equivalent quantity for new mutations, denoted $V_{\delta_{\text{mut}}}$. To see this, we repeated our simulations with the assumptions just described. The dominance coefficient of each new mutation on each trait was drawn from a shifted beta distribution, with a vanishing mean (such that mutations were semi-dominant on average) and variance $V_{\delta_{\text{mut}}}$. We simulated under seven different variances, including the extremes of $V_{\delta_{\text{mut}}} = 0$ (phenotypic additivity) and $V_{\delta_{\text{mut}}} = 1$, such that new mutations were either fully dominant or fully recessive with a 50:50 probability. The seven distributions we used are illustrated in Figure 2a.

Figure 2b shows clouds of F1 phenotypes for sets of 100 replicate simulations, each with $n = 2$ for visualisation. In all cases, the cloud of possible F1 is centred on the optimum, but grows in width with $V_{\delta_{\text{mut}}}$. Figure 2c shows the consequence of this F1 cloud expansion for the F1 clock. Here, to match Fig. 1c, we simulated with $n = 20$ traits, but with a larger population size, so that parental populations remained well adapted and thus had little potential for F1 heterosis. Results show that the F1 tick rate also increases steadily with the variance in mutational dominance, $V_{\delta_{\text{mut}}}$. Figure 2d (circles) shows that this is because, as anticipated, the dominance coefficients of the fixed substitutions closely match the dominance coefficients of new mutations (i.e., $V_\delta \approx V_{\delta_{\text{mut}}}$; see also Schneemann et al. 2020). As shown in Figure S3, $V_\delta \approx V_{\delta_{\text{mut}}}$ does not hold in all parameter regimes. However, the increase in the slope with $V_{\delta_{\text{mut}}}$ was consistently observed, including when each value of δ_{mut} applied to all traits, so that heterozygous and homozygous mutations differed in size, but not direction.

2.2. Directional selection and directional dominance

Results above assume that dominance deviations have no tendency to point in any phenotypic direction, such that the cloud of possible F1s is centered on the midparent, and all $E(\Delta_i) = 0$ (see Fig 1b). However, it is well known that directional selection can lead to directional dominance, as with Haldane's Sieve (Haldane, 1924, 1927; Frankham, 1990; Crnokrak and Roff, 1995; Ala-Honkola et al., 2013; Sztepanacz and Blows, 2015; Kapun et al., 2016). This is because populations evolving in a particular direction preferentially fix alleles that are dominant in that direction.

To understand the effects of directional dominance for F1 fitness, let us consider the most extreme case, where both populations adapt independently to identical environmental change, so that all of the

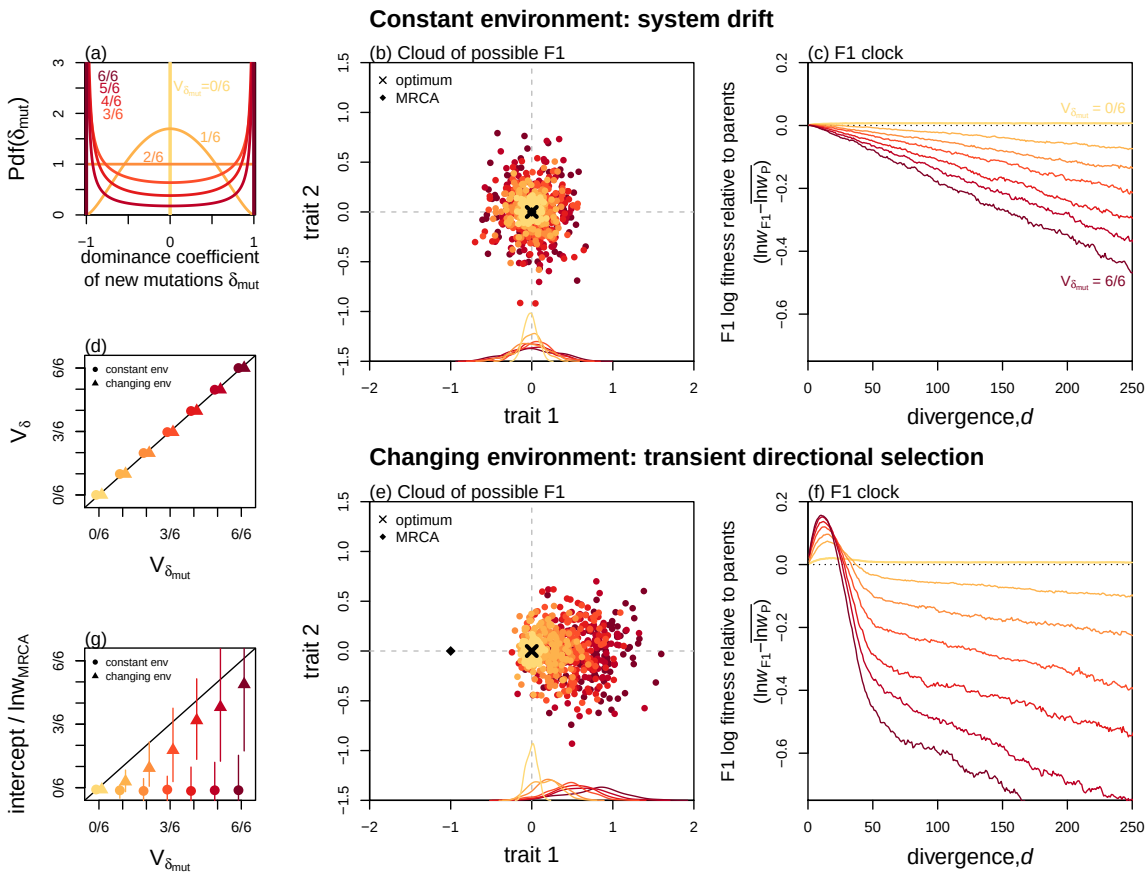

Figure 2: Deleterious effects of phenotypic dominance depend on the distribution of new mutations and the history of directional selection. (a): The distribution of dominance coefficients for new mutations used in each set of simulations. These were 7 different beta distributions, with vanishing means and 7 different variances, ranging from additivity (yellow: $V_{\delta_{\text{mut}}} = 0/6 = 0$) to complete dominance or recessivity (dark red: $V_{\delta_{\text{mut}}} = 6/6 = 1$). **(b)**–**(c)**: Simulations under constant environmental conditions show that **(b)** the variance in the cloud of possible F1s, and **(c)** the tick rate of the F1 clock both increase steadily with $V_{\delta_{\text{mut}}}$. **(d)**: This is because, for these simulations, the dominance coefficients of fixed mutations reflect those of the new mutations, such that $V_\delta \approx V_{\delta_{\text{mut}}}$. **(e)**: Directional selection can lead to directional dominance, causing the cloud of possible F1s to 'overshoot' the optimum in the direction of the past adaptation. **(f)**: Directional dominance creates transient heterosis as the parental lines evolve towards the new optimum, but a deleterious overshoot once the new optimum is reached. **(g)**: The result is that, after an initial period, the F1 clock ticks at the same rate, but with an intercept which also increases with $V_{\delta_{\text{mut}}}$. (b) and (e) show 100 simulations stopped after $d = 50$ substitutions, and with parameters chosen to best visualise the clouds ($N = 100$, $n = 2$). (c)–(d), and (f)–(g) show means across 100 replicate simulations with $n = 20$ and $N = 1000$ to generate a steady clock. All panels use $U = \bar{s} = 0.01$.

dominance deviations point in the same direction. In this case, the cloud of possible F1 becomes shifted in the direction of the past evolutionary change. This is illustrated in Figure 2e, where a bout of directional selection on trait 1 leads to $E(\Delta_1) > 0$. To understand why this happens, let us consider the evolution of trait 1 during the bout of adaptation. If we denote the F1 trait as $z_{\mathrm{F1},1}$, and the midparental value as $z_{\mathrm{MP},1}$, then eqs. 2 and 5 imply that:

$$E\left((z_{\mathrm{F1},1} - o_1)^2\right) = (z_{\mathrm{MP},1} - o_1)^2 + E(\Delta_1^2) - 2|(z_{\mathrm{MP},1} - o_1)|E(\Delta_1) \tag{9}$$

$$= V_{\Delta_1} + E^2(\Delta_1), \quad \text{if } z_{\mathrm{MP},1} = o_1 \tag{10}$$

The important point about eq. 9 is that it contains both positive and negative terms. This is because, during the adaptation phase, the directional dominance can increase F1 fitness, by taking its trait value closer to the still-distant optimum. But as shown by eq. 10, once the new optimum is approached, the directional dominance leads to a permanent and deleterious 'overshoot' of this optimum (see also Figure S4, and Ono et al. (2017) for a related phenomenon shown in yeast).

If the adaptation to the new optimum required d_{dir} substitutions, the expected log fitness of the F1 after this period (i.e. once system drift at the new optimum has begun) is:

$$E\left(\ln w_{\mathrm{F1}}\right) \approx \ln w_{\mathrm{MRCA}} V_{\delta_{\mathrm{dir}}} - (d - d_{\mathrm{dir}})\, n V_{a\delta}, \qquad \text{if } \mathbf{z}_{\mathrm{MP}} = \mathbf{o}, d > d_{\mathrm{dir}} \tag{11}$$

where $V_{\delta_{\mathrm{dir}}}$ is the variance in dominance coefficients of the adaptive substitutions, and $\ln w_{\mathrm{MRCA}}$ is the log fitness of the maladapted ancestor. Comparing eq. 11 to earlier results (eqs. 6 and 7), we see that the overshoot adds a new term. The history of directional selection leads to a non-zero intercept for the F1 clock.

Complete simulations of this scenario are reported in Figure 2f. Comparison of Figures 2f and 2c shows clearly the transient increase in F1 fitness (eq. 9), and the deleterious overshoot once the new optimum is approached (eq. 10). After the initial period, the F1 clock continues to tick at the same rate as before (see triangles in Figure 2d), but with a permanent intercept (eq. 11). Fitting linear regressions to the F1 clocks in Figure 2f after excluding the first 50 substitutions allowed us to calculate this intercept. Figure 2g (triangles) confirms that the intercept is also affected by the input of new mutations, such that $V_{\delta_{\mathrm{dir}}} \propto V_{\delta_{\mathrm{mut}}}$.

Results above concern directional selection in a common direction, but they generalize readily. For example, if each parental population underwent directional adaptive change on a different phenotypic trait, the resulting directional dominance would lead to an F1 with a mixture of the derived traits of the two parental lines. This is the case described as trait mismatch by Thompson et al. (2021). In all cases, with variable phenotypic dominance, a history of past directional selection can lead to an additional loss of fitness for the F1.

3. The lucky beneficial effects of dominance

So far, we have considered the typical effects of dominance in the F1, by averaging over the possible evolutionary histories. These effects are generally deleterious, and any heterosis tends to be transient. Nevertheless, even when effects are deleterious on average, by chance alone, some realisations of the evolutionary process will take the F1 closer to the optimum. In Fisher's geometric model, these "lucky" outcomes are far more likely when the number of phenotypic traits, n, is small. This is because random changes are more likely to go in the right direction when the dimensionality of the landscape is low (Fisher, 1930; Orr, 2000).

To see this, let us first consider again the scenario illustrated in Figures 1c and 2b-c, where parental populations remain well adapted to a single fixed optimum. We assume that the Δ_i have vanishing means (i.e., no directional dominance), and are approximately normally distributed. The normality is justified by the central-limit-like behaviour arising from the sum in eq. 4. With these assumptions, the coefficient of variation in log F1 fitness is approximately

$$\mathrm{CV}\left(\ln w_{\mathrm{F1}}\right) \equiv \frac{\sqrt{\mathrm{Var}\left(\ln w_{\mathrm{F1}}\right)}}{\left|\mathrm{E}\left(\ln w_{\mathrm{F1}}\right)\right|}$$

$$\approx \sqrt{\frac{2}{n}}, \qquad d \gg 1 \tag{12}$$

Equation 12 implies that, when n is large, the F1 clock will tick in a relatively deterministic way. When n is small, by contrast, increasing the divergence between the parental lines may lead to the chance re-appearance of heterosis, even after many generations of low fitness F1. This is confirmed by simulations reported in Figure S5a-b.

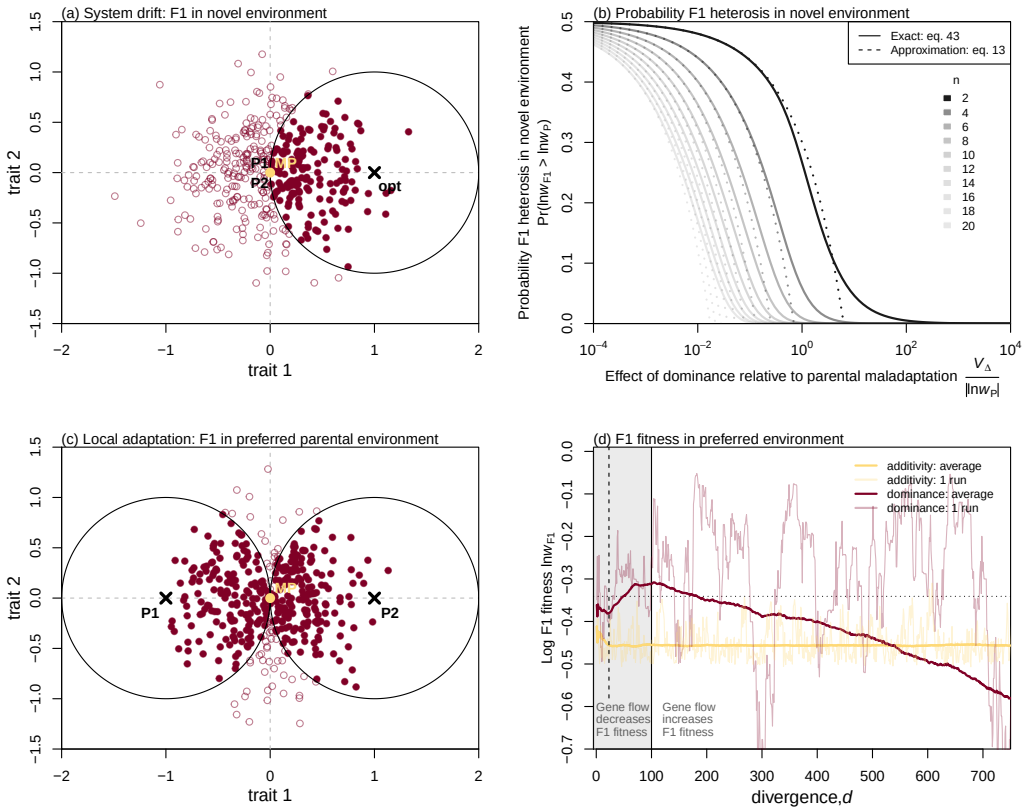

Figure 3: "Lucky" beneficial effects of dominance in heterogeneous environments. **(a)**: Even when two parental lines are well adapted to similar environments, their F1 may be well adapted to a novel environment, simply by chance. These fitter F1 (solid dark red points) lie within the circle shown, with the new optimum at its centre. **(b)**: The probability of such fitter F1 appearing decreases rapidly with n (the dimensionality) and with $V_\Delta/|\ln w_{\mathrm{P}}|$ (the size of the dominance deviations relative to the maladaptation of the parents). **(c)**: When the parents are locally adapted to two different environments, dominance might yield an F1 that is much fitter in one of the two environments, potentially leading to asymmetrical gene flow. **(d)**: Simulations of the F1 of locally adapted parents. F1 were scored in the "preferred" parental habitat, i.e. the habitat to which they were better adapted. Simulations began with the MRCA intermediate between the two optima, and the parents adapted to their optima rapidly, always before $d = 25$ (see vertical dotted line). With additivity, F1 fitness remained roughly constant thereafter (yellow curves), but with dominance, expected F1 fitness continued to increase with divergence (dark red curves), until a maximum roughly predicted by eq. 15 (horizontal dotted line). However, changes in F1 fitness were erratic for any single run (see fainter curves). The result is that gene flow, which reduces d, can either increase or decrease F1 fitness with dominance. Simulations in (d) were the same as in Fig. 2e, with the extremes of $V_{\delta_{\mathrm{mut}}} = 0$ and 1, and averages taken over all possible pairs of the 200 parental populations; see also Figure S7.

3.1. Lucky beneficial effects in a novel environment

The lucky beneficial effects of dominance may be particularly consequential when hybrids are formed in novel environments, to which one or both parental lines are severely maladapted. In the additive model, F1 heterosis can only appear when the parents are maladapted in different ways (Simon et al., 2018; Schneemann et al., 2020; Yamaguchi and Otto, 2020). For example, hybrids between parents adapted to low and high altitudes might thrive at intermediate altitudes (Wang et al., 1997). With phenoptypic dominance, however, hybrid advantage might appear under a broader range of conditions. Consider, for example, the situation shown in Figure 3a, where two genomically divergent, but phenotypically similar parental lines hybridize in a novel habitat. In this case, the midparent matches the parental phenotypes ($\ln w_{\mathrm{MP}} \approx \ln w_{\mathrm{P}}$), and so the effects of dominance will be deleterious on average (eq. 6). Nevertheless, for some divergence histories, the F1 will be fitter by chance. The probability of this lucky heterosis can be derived by noting that eq. 5 has a non-central chi-squared distribution if the Δ_i are normal. The probability therefore depends on n and the ratio $V_\Delta / \ln w_{\mathrm{P}}$, which compares the sizes of the dominance deviations to the maladaptation of the parents. Figure 3b plots this probability for a range of values, and we also have the approximation:

$$\Pr(\ln w_{\mathrm{F1}} > \ln w_{\mathrm{P}}) \approx \frac{1}{2}\left(1 - \frac{(n-1)}{\sqrt{2\pi}}\sqrt{\frac{V_\Delta}{|\ln w_{\mathrm{P}}|}}\right) \tag{13}$$

(see Appendix 1). As shown in Figure 3b, the maximum probability of heterosis is $\frac{1}{2}$, and it applies when dominance effects are small compared to the maladaptation of the parents such that $V_\Delta \ll |\ln w_{\mathrm{P}}|$ (toward the left of Fig. 3b). This implies that any fitness gain due to heterosis would also be very small, and simply reproduces Fisher's result that very small changes have a 50% chance of being beneficial (Fisher, 1930). Conversely when the dominance effects are very large, such that $V_\Delta \gg |\ln w_{\mathrm{P}}|$ (toward the right of Fig. 3b), the F1 are almost certain to overshoot the new optimum. Therefore, the area of interest concerns $V_\Delta \approx |\ln w_{\mathrm{P}}|$ (toward the centre of Fig. 3b), and in this regime, the probability of heterosis declines rapidly with n (see also Fig. S6). Only when n is small is there a non-negligible chance that dominance effects are both substantial and beneficial.

3.2. Hybrids between locally adapted parents

Now let us consider the situation shown in Figure 3c. This panel represents the outcome of divergent selection (Schluter, 2000), where each parental line is well adapted to a different habitat, characterized by different phenotypic optima. In this case, the midparent lies midway between the optima, such that,

in either habitat $\ln w_{\mathrm{MP}} \approx \ln w_{\mathrm{P}}/4$, with $\ln w_{\mathrm{P}}$ denoting the fitness of the maladapted parental line. In this scenario, results in Figure 3b now describe the probability of the F1 being fitter than the midparent in one of the two parental habitats. In most cases, the F1 will be fitter in one of the habitats than the other, and in this "preferred habitat", its expected log fitness is:

$$E\left(\ln w_{\mathrm{F1}_{\mathrm{pref}}}\right) = \frac{\ln w_{\mathrm{P}}}{4} - nV_{\Delta} + \sqrt{\frac{2V_{\Delta}|\ln w_{\mathrm{P}}|}{\pi}} \tag{14}$$

Equation 14 neatly separates the negative and positive effects of dominance in its second and third terms, and shows that the deleterious effects—but not the beneficial effects—grow with n. Equation 14 also shows that the fitness benefits of dominance will be greatest at intermediate values of V_{Δ}, and therefore, at intermediate levels of divergence (eqs. 6-7). This means that, in contrast to the scenario shown in Fig. 2e–f, F1 fitness in one of the two habitats, might continue to increase even after two parental populations have adapted to their new optima. At this intermediate level of divergence, when the fitness benefit is greatest, we find:

$$\max\left(E\left(\ln w_{\mathrm{F1}_{\mathrm{pref}}}\right)\right) = \ln w_{\mathrm{MP}}\left(1 - \frac{2}{\pi n}\right), \tag{15}$$

so, at suitable levels of divergence, phenotypic dominance can yield a substantial fitness increase for the F1 over the midparent. However, these lucky effects of dominance occur only in one of the two habitats and only when n is small. All of these results are confirmed in Figure 3d, which shows simulation results with $n = 2$ (see also Figure S7). Results confirm that on average (solid lines), F1 fitness with dominance initially increases with d, before reaching a maximum approximated by eq. 15, and then decreasing. However, eq. 12 still applies, and so for any single realisation of the divergence history (faint lines in Figure 3d) the fitness of F1 with phenotypic dominance is highly erratic.

Taken together, these results imply that phenotypic dominance may lead to asymmetrical gene flow between locally adapted populations after secondary contact. The asymmetry arises from fitness differences between the F1 in the two habitats and is therefore a form of "dominance drive" (Mallet and Barton, 1989; Barton, 1992). The results also imply that the gene flow, which by definition reduces d, may have an unpredictable effect on F1 fitness. In some cases, homogenisation of the parental genomes will increase the fitness of their F1, but in other cases, it will lead to a switch in the direction of the gene flow (as the F1 becomes adapted to the other parental habitat), or even to a substantially lower

F1 fitness. In the latter case, the outcome would therefore resemble reinforcement selection, albeit via a completely different route.

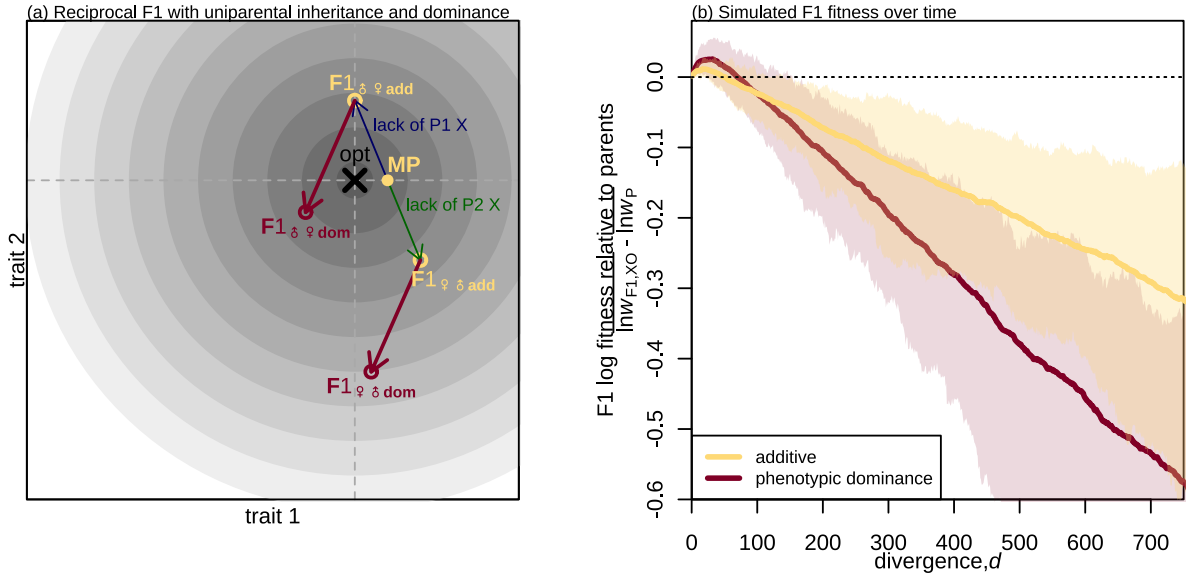

Figure 4: Uniparental effects interact with dominance to determine F1 fitness. Both phenotypic dominance and uniparental inheritance/expression cause the F1 to deviate from the midparent (MP). **(a)**: An illustration of heterogametic (XO) offspring, from a reciprocal F1 cross: i.e. including both the male-female and female-male cross directions of the same parental lines. Yellow points show the reciprocal F1 under additivity ($\mathrm{F1}_{♀♂,\mathrm{add}}, \mathrm{F1}_{♂♀,\mathrm{add}}$) and dark red points show the reciprocal F1 with phenotypic dominance ($\mathrm{F1}_{♀♂,\mathrm{dom}}, \mathrm{F1}_{♂♀,\mathrm{dom}}$). The dark blue and green lines indicate the deviations due to the absent paternal X, and are equal and opposite for the two cross directions. The dark red lines indicate the dominance deviations due to heterozygosity on the autosomes, and are identical for both cross directions. Both sets of deviations are expected to grow with the divergence, d. **(b)**: Simulation results of log fitness in XO F1 hybrids, averaged across the two cross directions. Results show that the F1 clock appears under additivity with uniparental effects. Plotting conventions and simulation runs are identical to those shown in Figure 1c, except that a proportion $x = 1/4$ of the divergent sites were randomly assigned to the X-chromosome before forming the hybrids. As a result, optimal outbreeding and the F1 clock appear not only with dominance (dark red curves) but also with additivity (yellow curves)

.

4. The interaction of phenotypic dominance with uniparental inheritance

In addition to dominance, the F1 might differ from the midparent for a second reason: the uniparental inheritance or expression of divergent alleles, for example on sex chromosomes or mitochondria (Fraïsse et al., 2016; Simon et al., 2018). In the parental lines, the complete set of alleles may be co-adapted in their additive effects, e.g. due to compensatory changes. When hybrids inherit a complete set of parental alleles, these co-adapted relationships maintain intact; but when hybrids inherit some parts of their genome from only one parent, they may lack some of the co-adapted alleles from the other parent. As we show below, dominance and uniparental inheritance have similar effects on hybrid fitness, but with one key difference—under uniparental inheritance the direction of phenotypic deviation from the midparent is opposite for the two cross directions.

The effects of uniparental inheritance are illustrated in Figure 4a. In this figure, and throughout this section, we assume a concrete example of uniparental effects, although results generalize easily to other cases. Our example involves sex chromosomes, where the heterogametic sex are effectively XO (i.e. where males carry only the maternal X, and the Y is either missing or highly degenerate). We further assume a common form of dosage compensation where X-linked alleles have identical effects in homozygous and hemizygous state. This last assumption ensures that non-hybrid offspring of both sexes are phenotypically identical (i.e., that male and female offspring from a P1 $\times$ P1 cross are equally fit).

In the case described, the absence of the paternal X causes the heterogametic F1 to deviate from the midparent, even under phenotypic additivity. These deviations are equal and opposite for the two cross directions (see the blue and green arrows in Fig. 4a). If there is phenotypic dominance for alleles on the autosomes, then this leads to further deviations that apply identically to both cross directions (see the dark red arrows in Fig. 4a). In the sections below, we show how uniparental effects and phenotypic dominance combine to determine F1 fitness.

4.1. The F1 clock under dominance and uniparental inheritance

If we consider any single F1, then uniparental effects and phenotypic dominance have essentially the same consequences. As such, they represent alternative, but non-exclusive explanations of the F1 clock and optimal outbreeding. To see this, let us consider the expected log fitness of an XO F1 hybrid. If we assume that a fraction, x, of the d divergent alleles are X-linked, then from eq. 5 and published results (Fraïsse et al., 2016; Simon et al., 2018), we find:

$$E\left(\ln w_{\mathrm{F1,XO}}\right) = \frac{\overline{\ln w_{\mathrm{P}}}}{2}(1+x^2) - dnV_{a\delta}(1-x) - dnV_a x(1-x) \tag{16}$$

Equation 16 still predicts optimal outbreeding, but there are now two terms that reduce F1 fitness. The second term captures the dominance deviations from the autosomes, and the third term captures the lack of co-adaptation between paternal autosomes and the absent paternal X.

Figure 4b illustrates these results. We returned to our simulated data, and formed heterogametic F1 in the way described. Results show that optimal outbreeding and the F1 clock now appear under the additive model too (yellow curves; Fraïsse et al., 2016; see also Fig. S5c-d). Dominance simply accelerates the fitness decline (dark red curves).

4.2. Haldane's rule

If we consider fitness differences between heterogametic and homogametic hybrids, then the uniparental effects and phenotypic dominance tend to push in different directions. If the parents are well adapted, then the additive model predicts lower fitness for the heterogametic sex, in accord with Haldane's Rule (Haldane, 1922; Barton, 2001; Fraïsse et al., 2016; Simon et al., 2018). This is due to the loss of co-adaptation, described above (see the third term of eq. 16). However, phenotypic dominance has the opposite effect, because it tends to makes heterozygosity deleterious, and heterogametic F1, being hemizygous for the X, lack the deleterious heterozygosity on the X (to see this, compare the second terms of eqs. 7 and 16).

To see how these two effects balance, let us consider the difference in log fitness between heterogametic and homogametic F1. Using eqs. 7, 8 and 16, we find:

$$E\left(\ln w_{\mathrm{F1,XO}} - \ln w_{\mathrm{F1,XX}}\right) = \frac{\overline{\ln w_{\mathrm{P}}}}{2}x^2 - dnxV_a\left(1 - x - V_\delta\right) \tag{17}$$

Haldane's Rule holds on average if eq. 17 is negative. If parents are well adapted, or divergence is substantial, then the second term in eq. 17 dominates and we expect Haldane's Rule on the condition that

$$V_\delta < (1-x). \tag{18}$$

Note that this equation was derived under the restrictive assumptions of eq. 8, but it holds much more

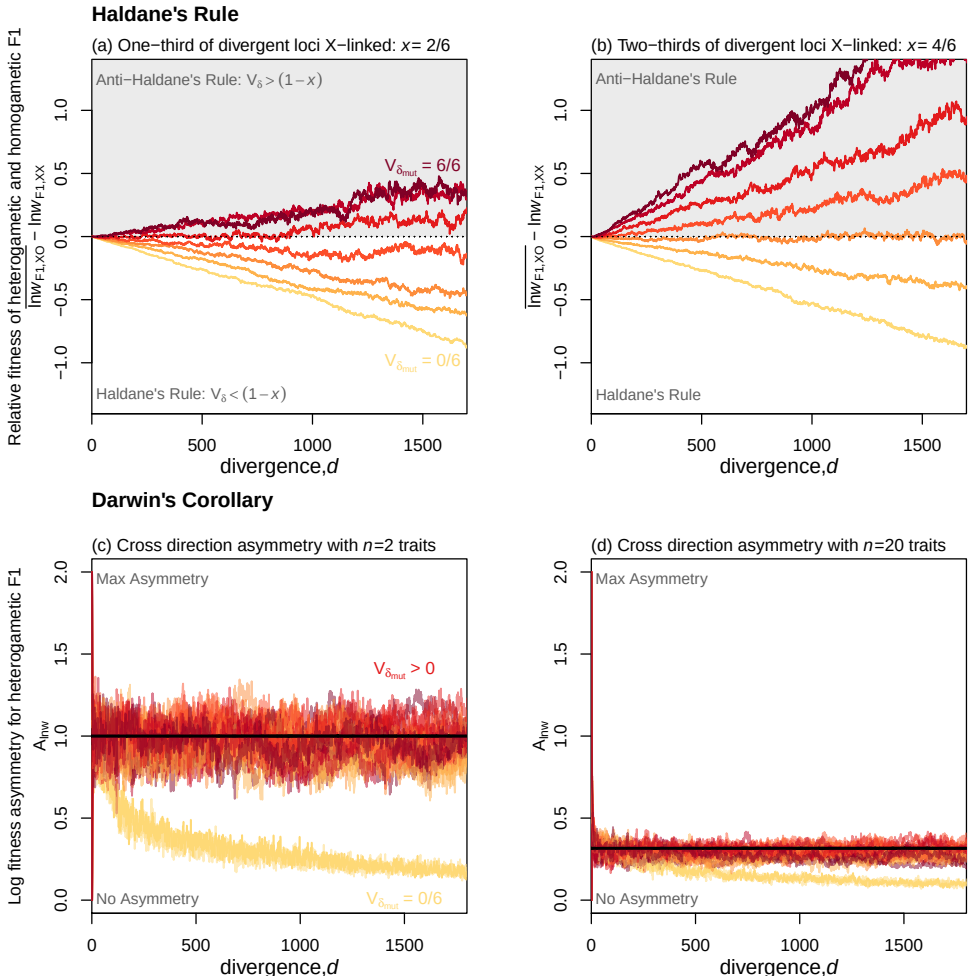

Figure 5: Differences between F1 cross types with uniparental inheritance and variable phenotypic dominance. **(a)–(b)**: Phenotypic dominance can lead to violations of Haldane's Rule. Curves show the difference in mean log fitness between heterogametic and homogametic F1, at various levels of parental divergence. Curves below the horizontal dotted line indicate that heterogametic F1 are less fit, such that Haldane's Rule holds. Colours correspond to those in Figure 2, and show the same 7 values of $V_{\delta_{\rm mut}}$. Haldane's Rule always holds with additivity (yellow lines; $V_{\delta_{\rm mut}} = 0$), but can be violated if dominance coefficients are highly variable (darker colours), or if the X chromosome is very large – compare (a) to (b). **(c)–(d)**: Phenotypic dominance enhances log fitness asymmetry between reciprocal F1. Curves show the asymmetry measure of eq. 19, measured for XO hybrid offspring from the two cross directions. Results under additivity (yellow), differ qualitatively from those with phenotypic dominance (all other colours), and the latter all show approximately the same result: $A_{\ln w} \approx \sqrt{\frac{2}{n}}$ (black horizontal lines). Plots include 28 sets of simulations, using two populations sizes: $N = 100, 1000$; two sizes of X chromosome: $x = \frac{1}{3}, \frac{2}{3}$; and the 7 $V_{\delta_{\rm mut}}$ values shown in Fig. 2a). Each curve represents the mean across 100 simulation replicates, all with $U = \bar{s} = 0.01$; (a)–(b) used $n = 20$ and $N = 100$. Note that the divergence was simulated under strictly biparental inheritance, to ensure that $V_\delta \approx V_{\delta_{\rm mut}}$.

generally, if we define V_δ as the relative sizes of the dominance and additive effects (i.e., as $V_\delta \equiv \mathrm{V}_{a\delta}/V_a$, where the latter two quantities are variances averaged across traits).

Equation 18 confirms that Haldane's Rule will always hold under additivity (when $V_\delta = 0$), but that variable phenotypic dominance can lead to violations (i.e., 'anti-Haldane's Rule'), especially when the dominance coefficients are highly variable, or if the X is very large. This is confirmed by the simulations shown in Figure 5a–b. These plots use the simulation runs from Fig. 2b, for which the dominance coefficients match the mutational input ($V_\delta \approx V_{\delta_{\mathrm{mut}}}$). The results show that Haldane's Rule is indeed violated when $V_{\delta_{\mathrm{mut}}} > (1 - x)$.

4.3. Darwin's Corollary

Let us now consider the fitness differences between the F1 cross directions, that is a cross where P1 is dam (P1♀×P2♂) compared to a cross where P1 is sire (P2♀×P1♂). In stark contrast to Haldane's Rule, phenotypic dominance works together with uniparental effects to generate this pattern. To see this, we will use the following measure of log fitness asymmetry, where the absolute difference between the cross directions is normalized by the absolute mean:

$$A_{\ln w} \equiv \frac{\left|\ln w_{\mathrm{F1,♀♂}} - \ln w_{\mathrm{F1,♂♀}}\right|}{\frac{1}{2}|\ln w_{\mathrm{F1,♀♂}} + \ln w_{\mathrm{F1,♂♀}}|} \tag{19}$$

Here, ♀♂ and ♂♀ denote the two cross directions, and the statistic is bounded at $0 \leq A_{\ln w} \leq 2$.

Fraïsse et al. (2016) studied F1 asymmetry under the additive version of Fisher's model, and showed that asymmetry could appear only if the midparental phenotype was maladapted. This reason is clear from Figure 4a. When the midparent matches the optimum, the equal and opposite deviations (blue and green arrows), will lead to identical fitness loss.

Furthermore, even if the midparent is maladapted, its distance from the optimum must be large relative to the deviations for substantial asymmetry to appear. But, in a stable environment, the parents (and therefore the midparent) remain close the optimum, while the deviations grow with d (e.g. eq. 7), such that the F1 in both cross directions are expected to become less fit, and the difference between their fitnesses smaller. This implies that any asymmetry between the cross directions will decline with d under the additive model. This is confirmed by the yellow curves in Figure 5c–d.

Adding phenotypic dominance to the model qualitatively changes this result. Because the dominance deviations also grow with d, levels of asymmetry can remain large, even at high divergence. An illustrative case is shown in Figure 4a. Moreover, with dominance, we find that to a rough approximation:

$$E\left(A_{\ln w}\right) \approx \sqrt{\frac{2}{n}}, \qquad \text{if } V_\delta > 0 \tag{20}$$

so the normalized asymmetry in log fitness is predicted to remain roughly constant at all levels of divergence, and to depend solely on the number of traits, n. This is confirmed by the simulations with dominance shown in Figure 5c–d. For all of the parameter values we simulated, as long as there was phenotypic dominance ($V_{\delta_{\text{mut}}} > 0$), then asymmetry levels remained roughly constant, and were close to the prediction of eq. 20 (see also Figure S5e–f).

The arguments in this section show that under Fisher's geometric model, Darwin's Corollary can be explained as another "lucky" beneficial effect of dominance, where the good luck for one cross direction is balanced by bad luck for the other. The consequence is that – with phenotypic dominance, but not with additivity – Darwin's Corollary is predicted at high levels of genomic divergence.

Discussion

During evolutionary divergence between populations, the additive effects of the divergent alleles act together in homozygous genotypes, and so they will often by co-adapted – i.e. selected to function well together in the current environment. By contrast, the dominance effects will act together only in early generation hybrids, and so they are more likely to wander erratically in phenotypic space. We have presented a systematic study of these dominance effects and their consequences for F1 hybrid fitness. We have shown that the effects are usually deleterious, reducing hybrid fitness; but there can also be lucky outcomes when, by chance, the novel F1 phenotype is better adapted than the parental lines.

Taking the negative and positive effects together, dominance could have a broad range of consequences for the F1. These consequences include asymmetric and fluctuating gene flow between locally adapted populations (Fig. 3c-d), and F1 hybrid advantage, either in novel habitats (Fig. 3a-b), or during ongoing directional selection (Fig. 2f). The consequences also include well-known patterns, that have been observed in multiple taxa, such as optimal outbreeding and the F1 clock (Waser, 1993; Edmands, 2002; Wei and Zhang, 2018; Dagilis et al., 2019; Coughlan and Matute, 2020; Figs. 1c and 2c).

But while dominance is a possible cause of these patterns, it need not be their actual or only cause. Variable phenotypic dominance has often been measured directly (Lynch and Walsh, 1998, p.485; Thompson et al., 2021); including as transgressive trait variation, which is non-additive on any scale of measurement, and so cannot be easily "transformed away" (Lynch and Walsh, 1998, p.308). However, it does not follow that this dominance is important for determining hybrid fitness. While explicit links between traits and fitness can be made (e.g. Thompson et al., 2021), often the relevant set of traits and the way selection perceives these traits is unknown. In general, the "traits" in Fisher's model need not be equated with standard quantitative traits (Martin, 2014), and an additive model may yield good predictions even if the measurable quantitative traits are highly non-additive.

Moreover, we have shown that some of the empirical patterns generated by dominance also appear under an additive phenotypic model, if some determinants of the phenotype are uniparentally inherited or expressed. Data give strong support for such uniparental effects – not least in the ubiquity of Darwin's Corollary (Kölreuters, 1766; Darwin, 1859, Ch. 8; Muller, 1942; Turelli and Moyle, 2007), in diverse plant (Tiffin et al., 2001), fungus (Dettman et al., 2003), and animal systems (Bolnick et al., 2008; Brandvain et al., 2014), including simultaneous hermaphrodites (Escobar et al., 2008; Sato et al., 2014; Bouchemousse et al., 2016; Fraïsse et al., 2016).

One obvious question, therefore, is whether there are any telltale signs of phenotypic dominance,

i.e. patterns in the data that cannot be generated by additive models. In the current work, we have identified two. The first pattern involves Darwin's Corollary, although it is not the observation of fitness asymmetry *per se* that implicates dominance. We have shown that log fitness asymmetry between cross directions tends to decline with genomic divergence under the additive model, but remains constant if there is variable phenotypic dominance (Fig. 5c-d). Note that the difference between predictions with and without dominance is relatively subtle, and so cannot be inferred directly from published results. For example, the relevant measure of asymmetry (eq. 19) uses log transformed fitness, relative to the maximum obtainable value, and this is difficult to obtain at low divergences – where parental fitness cannot serve as a proxy for the optimum – and at very high divergences – where zeros cannot be log transformed. The pattern can also be noisy (see Fig. S5e-f).

The second pattern that can only be explained by phenotypic dominance is somewhat simpler. Dominance and uniparental effects both allow for transgressive trait variation (i.e trait values that lie outside of the range of parental values; Stelkens and Seehausen, 2009), but only with dominance do the reciprocal cross directions produce transgression in the same phenotypic direction (Fig. 4a). The unique prediction of the dominance model is therefore a particular type of F1 fitness advantage. First, the advantage must apply to both cross directions. Second, it must not be due to the masking of deleterious recessives, nor to bounded hybrid advantage, such as found on an environmental gradient (with hybrids expressing an intermediate phenotype, beneficial in intermediate environments as in Wang et al., 1997); both latter sorts of heterosis are generated by the additive model (Simon et al., 2018; Schneemann et al., 2020; Yamaguchi and Otto, 2020). To demonstrate the pattern of heterosis that is unique to dominance, the reciprocal F1 must outperform both parents in a novel environment, while both parents outperform the F1 in their shared original environment. Equivalently, the two "environments" might involve the presence or absence of novel stressors, to which parental lines are not normally exposed. Some possible examples of such heterosis include increased thermal tolerance in copepod F1 (Pereira et al., 2014), invasive hybrid tiger salamanders (Ryan et al., 2009; Cooper and Shaffer, 2021), and *Senecio* ecotype hybrids in a habitat foreign to both parents (Walter et al., 2020). Such data are the strongest evidence for the importance of phenotypic dominance for F1 hybrid fitness.

This paper has focused on the consequences of dominance for the F1, but of course later-generation recombinant hybrids will experience some effects of dominance as well. While the telltale patterns in the F1 apply whenever dominance effects are present, the key question for later-generation hybrids is more quantitative: how do the deleterious effects of dominance effects compare to the costs of segregation

variance, i.e. to the lack of coadaptation between the additive effects? We have denoted as V_δ the relative sizes of these effects, and shown that under some conditions, V_δ will equal the variance in the dominance coefficients of the fixed effects. The importance of dominance effects for later generation hybrids therefore depends on the size of V_δ.

There are a number of observations that do provide upper bounds for V_δ in nature. The first is the ubiquity of Haldane's Rule (Haldane, 1922; Coyne and Orr, 2004, Ch. 8; Schilthuizen et al., 2011). We have shown that Haldane's Rule is predicted only when V_δ is smaller than the proportion of divergent loci that are biparentally inherited (eq. 18; Fig. 5a-b). Moreover, this bound applies to dominance effects at X-linked loci, which are probably larger than those on the autosomes (Charlesworth et al., 1987). Similar bounds also follow from direct observations of later-generation hybrids. For example, the model predicts that heterozygosity will be beneficial on average if $V_\delta < 1$, and so this bound is supported by studies reporting selection for increased heterozygosity in recombinant hybrids (Lindtke et al., 2012; Simon et al., 2018; Thompson et al., 2022). Easier to observe, and so better supported, is hybrid breakdown—the fact that later-generation hybrids are very often less fit than the initial F1 (Muller, 1940; Vetukhiv, 1954; Fraïsse et al., 2016, Table S1). A simple generalization of eq. 18 (eq. 52) shows that hybrid breakdown between well adapted parental lines will occur only if:

$$V_\delta < \frac{4h(1-h) - p_{12}}{1 - p_{12}}. \tag{21}$$

where h is the hybrid index (the proportion of divergent alleles that come from one the parental lines) and p_{12} is the interpopulation heterozygosity. It follows that F2 breakdown will occur whenever $V_\delta < 1$ (since $4h(1-h) \approx 1$ for an F2). But backcross breakdown sets a stronger bound as it occurs only if $V_\delta < p_{12}$ (since $4h(1-h) = p_{12}(2-p_{12})$ for any backcross), and $p_{12} = 1/2$ on average for the first backcross.

Taken together, then, these bounds do not allow us to conclude that dominance effects will be negligible for any class of hybrid. However, they are at least consistent with a world where $0 < V_\delta \ll 1$. In this case, simple additive phenotypic models would yield good predictions for later generation recombinant hybrids (Fraïsse et al., 2016; Simon et al., 2018), even when dominance qualitatively affects outcomes for the F1.

References

Ala-Honkola, O., Hosken, D. J., Manier, M. K., Lüpold, S., Droge-Young, E. M., Berben, K. S., Collins, W. F., Belote, J. M., and Pitnick, S. (2013). Inbreeding reveals mode of past selection on male reproductive characters in *Drosophila melanogaster*. *Ecology and Evolution* 3 (7):2089–2102.

Arnold, M. L. and Hodges, S. A. (1995). Are natural hybrids fit or unfit relative to their parents? *Trends in Ecology & Evolution* 10 (2):67–71.

Barton, N. H. (2001). The role of hybridization in evolution. *Molecular Ecology* 10 (3):551–568.

Barton, N. (1992). On the spread of new gene combinations in the third phase of Wright's shifting-balance. *Evolution* 46 (2):551–557.

Bateson, P. (1978). Sexual imprinting and optimal outbreeding. *Nature* 273:659–660.

Blair, W. F. (1955). Mating Call and Stage of Speciation in the *Microhyla olivacea-M. carolinensis* Complex. *Evolution* 9 (4):469–480.

Bolnick, D. I., Turelli, M., López-Fernández, H., Wainwright, P. C., and Near, T. J. (2008). Accelerated mitochondrial evolution and "Darwin's corollary": Asymmetric viability of reciprocal F1 hybrids in centrarchid fishes. *Genetics* 178 (2):1037–1048.

Bouchemousse, S., Lévêque, L., Dubois, G., and Viard, F. (2016). Co-occurrence and reproductive synchrony do not ensure hybridization between an alien tunicate and its interfertile native congener. *Evolutionary Ecology* 30:69–87.

Brandvain, Y., Pauly, G. B., May, M. R., and Turelli, M. (2014). Explaining Darwin's corollary to Haldane's rule: the role of mitonuclear interactions in asymmetric postzygotic isolation among toads. *Genetics* 197:743–747.

Butlin, R. (1987). Speciation by reinforcement. *Trends in Ecology & Evolution* 2 (1):8–13.

Charlesworth, B., Coyne, J. A., and Barton, N. H. (1987). The relative rates of evolution of sex chromosomes and autosomes. *The American naturalist* 130:113–146.

Charlesworth, B. (1992). Evolutionary rates in partially self-fertilizing species. *The American Naturalist* 140 (1):126–148.

Chevin, L.-M., Decorzent, G., and Lenormand, T. (2014). Niche dimensionality and the genetics of ecological speciation. *Evolution* 68 (5):1244–1256.

Cooper, R. D. and Shaffer, H. B. (2021). Allele-specific expression and gene regulation help explain transgressive thermal tolerance in non-native hybrids of the endangered California tiger salamander (*Ambystoma californiense*). *Molecular Ecology* 30 (4):987–1004.

Coughlan, J. M. and Matute, D. R. (2020). The importance of intrinsic postzygotic barriers throughout the speciation process. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (1806):20190533.

Coyne, J. A. and Orr, H. A. (2004). *Speciation*. Sinauer Associates.

Crnokrak, P. and Roff, D. A. (1995). Dominance variance: Associations with selection and fitness. *Heredity* 75 (5):530–540.

Dagilis, A. J., Kirkpatrick, M., and Bolnick, D. I. (2019). The evolution of hybrid fitness during speciation. *PLOS Genetics* 15 (5):e1008125.

Darwin, C. (1859). *On the origin of species*. London: John Murray.

Dettman, J. R., Jacobson, D. J., and Taylor, J. W. (2003). A multilocus genealogical approach to phylogenetic species recognition in the model eukaryote *Neurospora*. *Evolution* 57 (12):2703–2720.

Dobzhansky, T. (1937). *Genetics and the origin of species*. New York: Colombia University.

Edmands, S. (2002). Does parental divergence predict reproductive compatibility? *Trends in Ecology and Evolution* 17 (11):520–527.

Escobar, J. S., Nicot, A., and David, P. (2008). The different sources of variation in inbreeding depression, heterosis and outbreeding depression in a metapopulation of *Physa acuta*. *Genetics* 180:1593–1608.

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

Fraïsse, C., Gunnarsson, P. A., Roze, D., Bierne, N., and Welch, J. J. (2016). The genetics of speciation: Insights from Fisher's geometric model. *Evolution* 70 (7):1450–1464.

Frankham, R. (1990). Are responses to artificial selection for reproductive fitness characters consistently asymmetrical? *Genetical Research* 56 (1):35–42.

Griswold, C. K. (2006). Gene flow's effect on the genetic architecture of a local adaptation and its consequences for QTL analyses. *Heredity* 96:445–453.

Guo, H., Makki, B., Alouini, M.-S., and Svensson, T. (2001). A semi-linear approximation of the first-order Marcum Q-function with application to predictor antenna systems. *IEEE Open Journal of the Communications Society* 2:273–286.

Haldane, J. B. S. (1922). Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics* 12 (2):101–109.

Haldane, J. B. S. (1924). A mathematical theory of natural and artificial selection, Part I. *Transactions of the Cambridge Philosophical Society* 23:19–41.

Haldane, J. B. S. (1927). A mathematical theory of natural and artificial selection, Part V: selection and mutation. *Mathematical Proceedings of the Cambridge Philosophical Society* 28:838–844.

Helstrom, C. (1995). *Elements of Signal Detection and Estimation*. Upper Saddle River, NJ: Prentice Hall.

Kapun, M., Fabian, D. K., Goudet, J., and Flatt, T. (2016). Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. *Molecular Biology and Evolution* 33 (5):1317–1336.

Kölreuters, J. (1766). *Dritte Fotsetzung*. Technical report, pages 119–124.

Lee, Y., Marsden, C. D., Norris, L. C., Collier, T. C., Main, B. J., Fofana, A., Cornel, A. J., and Lanzaro, G. C. (2013). Spatiotemporal dynamics of gene flow and hybrid fitness between the M and S forms of the malaria mosquito, *Anopheles gambiae*. *Proceedings of the National Academy of Sciences* 110 (49):19854–19859.

Lindtke, D., Buerkle, C. A., Barbará, T., Heinze, B., Castiglione, S., Bartha, D., and Lexer, C. (2012). Recombinant hybrids retain heterozygosity at many loci: new insights into the genomics of reproductive isolation in *Populus*. *Molecular Ecology* 21 (20):5042–5058.

Lynch, M. and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates.

Mallet, J. and Barton, N. (1989). Inference from clines stabilized by frequency-dependent selection. *Genetics* 122:967–976.

Mani, G. and Clarke, B. (1990). Mutational order: a major stochastic process in evolution. *Proceedings of the Royal Society of London B* 240:29–37.

Manna, F., Martin, G., and Lenormand, T. (2011). Fitness landscapes: An alternative theory for the dominance of mutation. *Genetics* 189:923–937.

Marcum, J. (1950). Table of Q functions. *U.S. Air Force RAND Research Memorandum M-339*. Santa Monica, CA: Rand Corporation.

Martin, G. (2014). Fisher's geometrical model emerges as a property of complex integrated phenotypic networks. *Genetics* 197:237–255.

Martin, G. and Lenormand, T. (2006). A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* 60 (5):893–907.

Matuszewski, S., Hermisson, J., and Kopp, M. (2014). Fisher's geometric model with a moving optimum. *Evolution* 68 (9):2571–2588.

Muller, H. (1940). Bearing of the *Drosophila* work on systematics. *The New Systematics*. edited by J. Huxley. Oxford: Clarendon Press, p. 185–268.

Muller, H. (1942). Isolating mechanisms, evolution and temperature. *Biological Symposium* 6:71–125.

Ono, J., Gerstein, A. C., and Otto, S. P. (2017). Widespread genetic incompatibilities between first-step mutations during parallel adaptation of *Saccharomyces cerevisiae* to a common environment. *PLOS Biology* 15 (1):e1002591.

Orr, H. A. (1991). A test of Fisher's theory of dominance. *Proc. Nati. Acad. Sci. USA* 88:11413–11415.

Orr, H. A. (1998). The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52 (4):935–949.

Orr, H. A. (2000). Adaptation and the cost of complexity. *Evolution* 54 (1):13–20.

Pereira, R. J., Barreto, F. S., and Burton, R. S. (2014). Ecological novelty by hybridization: experimental evidence for increased thermal tolerance by transgressive segregation in *tigriopus californicus*. *Evolution* 68 (1):204–215.

Rieseberg, L. H., Whitton, J., and Gardner, K. (1999). Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152 (2):713–727.

Ryan, M. E., Johnson, J. R., and Fitzpatrick, B. M. (2009). Invasive hybrid tiger salamander genotypes impact native amphibians. *Proceedings of the National Academy of Sciences* 106 (27):11166–11171.

Sato, A., Shimeld, S. M., and Bishop, J. D. D. (2014). Symmetrical reproductive compatibility of two species in the *Ciona intestinalis* (ascidiacea) species complex, a model for marine genomics and developmental biology. *Zoological Science* 31 (6):369–374.

Schiffman, J. S. and Ralph, P. L. (2021). System drift and speciation. *Evolution*.

Schilthuizen, M., Giesbers, M. C. W. G., and Beukeboom, L. W. (2011). Haldane's rule in the 21st century. *Heredity* 107 (2):95–102.

Schluter, D. (2000). *The ecology of adaptive radiation*. Oxford University Press.

Schneemann, H., De Sanctis, B., Roze, D., Bierne, N., and Welch, J. J. (2020). The geometry and genetics of hybridization. *Evolution* (74):2575–2590.

Simon, A., Bierne, N., and Welch, J. J. (2018). Coadapted genomes and selection on hybrids: Fisher's geometric model explains a variety of empirical patterns. *Evolution Letters* 2 (5):472–498.

Stelkens, R. and Seehausen, O. (2009). Genetic distance between species predicts novel trait expression in their hybrids. *Evolution* 63 (4):884–897.

Sztepanacz, J. L. and Blows, M. W. (2015). Dominance genetic variance for traits under directional selection in *Drosophila serrata*. *Genetics* 200 (1):371–384.

Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., and Gow, J. L. (2006). Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology* 15 (2):343–355.

Thompson, K. A., Urquhart-Cronish, M., Whitney, K. D., Rieseberg, L. H., and Schluter, D. (2021). Patterns, Predictors, and Consequences of Dominance in Hybrids. *The American Naturalist* 197 (3):E72–E88.

Thompson, K. A. et al. (2022). Analysis of ancestry heterozygosity suggests that hybrid incompatibilities in threespine stickleback are environment dependent. *PLOS Biology* 20 (1):e3001469.

Tiffin, P., Olson, S., and Moyle, L. C. (2001). Asymmetrical crossing barriers in angiosperms. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268 (1469):861–867.

Turelli, M. and Moyle, L. C. (2007). Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics* 176:1059–1088.

Vetukhiv, M. (1954). Integration of the genotype in local populations of three species of *Drosophila*. *Evolution* 8 (3):241–251.

Wallace, A. (1889). *Darwinism*. Macmillan.

Walter, G. M., Richards, T. J., Wilkinson, M. J., Blows, M. W., Aguirre, J. D., and Ortiz-Barrientos, D. (2020). Loss of ecologically important genetic variation in late generation hybrids reveals links between adaptation and speciation. *Evolution Letters* 4 (4):302–316.

Wang, H., McArthur, E. D., Sanderson, S. C., Graham, J. H., and Freeman, D. C. (1997). Narrow hybrid zone between two subspecies of big sagebrush (*Artemisia tridentata*: Asteraceae). IV. Reciprocal transplant experiments. *Evolution* 51 (1):95–102.

Waser, N. M. (1993). Population structure, optimal outbreeding, and assortative mating in angiosperms. *The natural history of inbreeding and outbreeding*. Section: 9, p. 173–199.

Wei, X. and Zhang, J. (2018). The optimal mating distance resulting from heterosis and genetic incompatibility. *Science Advances* 4 (11):eaau5518.

Welch, J. J. and Waxman, D. (2003). Modularity and the cost of complexity. *Evolution* 57 (8):1723–1734.

Yamaguchi, R. and Otto, S. P. (2020). Insights from Fisher's geometric model on the likelihood of speciation under different histories of environmental change. *Evolution* 74:1603–1619.

Yeaman, S. and Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution* 65 (7):1897–1911.

Supplementary Figures

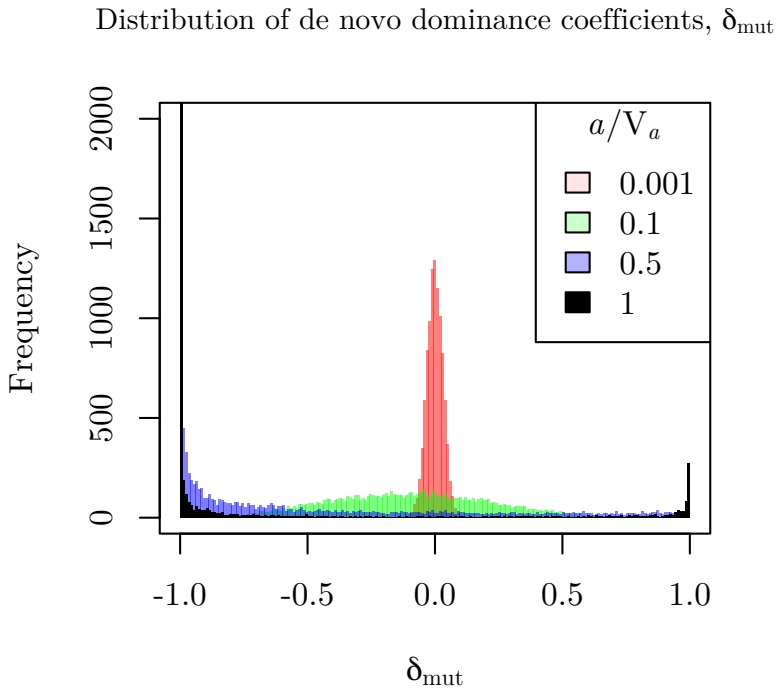

Figure S1: Size-dependent distribution of dominance coefficients. The distribution of dominance coefficients for new mutations as used in the simulations reported in Figures 1c, 4b and Figure S5. The heterozygous effect of a mutation on a given trait is $a_{\text{mut}}(1+\delta_{\text{mut}})$. Here, a_{mut} is the mutation's additive effect, which is normally distributed with vanishing mean and standard deviation $V_{a_{\text{mut}}}$ (eq. 59). The dominance coefficient, δ_{mut}, is drawn from a beta distribution, whose parameters vary with the size of of the additive effect according to eqs. 60-61. The aim is to reflect empirical observations that lethal or highly deleterious mutations are often recessive, while small-effect mutations are more likely to be semi-dominant (Orr, 1991; Manna et al., 2011).

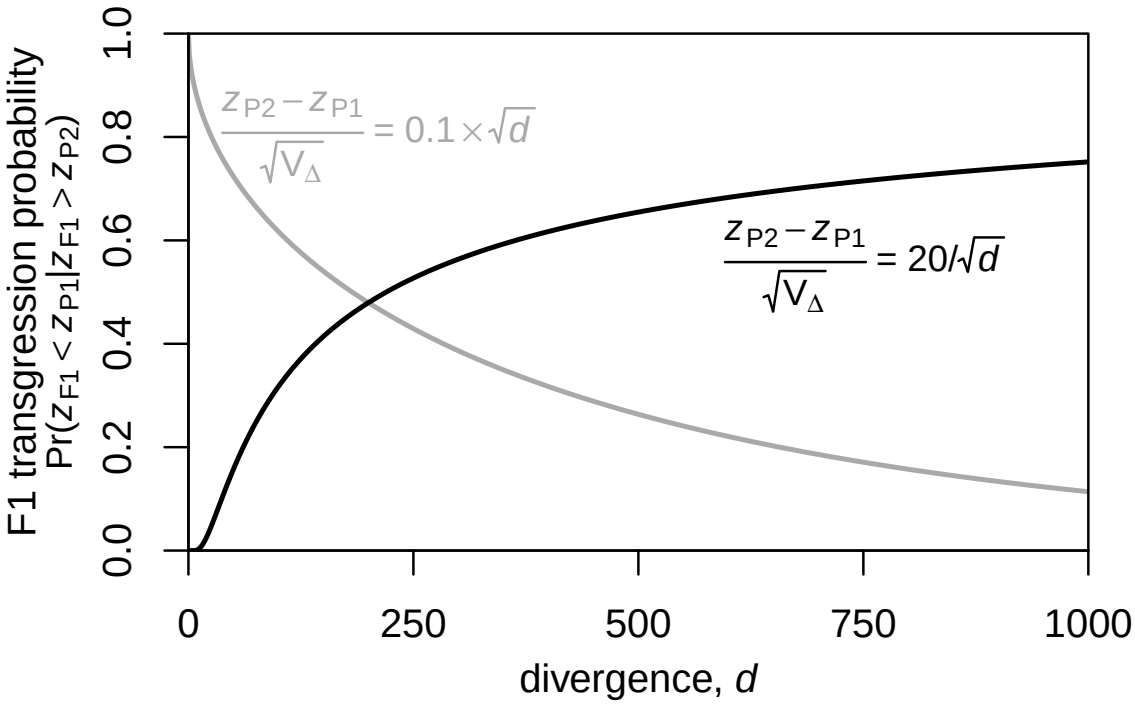

Figure S2: Transgressive F1 trait variation can increase or decrease with genetic divergence. Curves show the probability that the F1 trait value lies outside of the range of values in the two parental lines (eq. 47). The variance in the dominance coefficients, V_Δ increases linearly with genomic divergence, d (e.g. eqs. 6-7). If the divergence in the parental phenotypes also increases with d (e.g. due to directional selection) then the probability of transgressive variation decreases with d (grey curve). If the parental phenotypes remain roughly constant, e.g., due to effective stabilizing selection, then the probability of transgression increases (black curve). Both regimes have been observed in data (Stelkens and Seehausen, 2009).

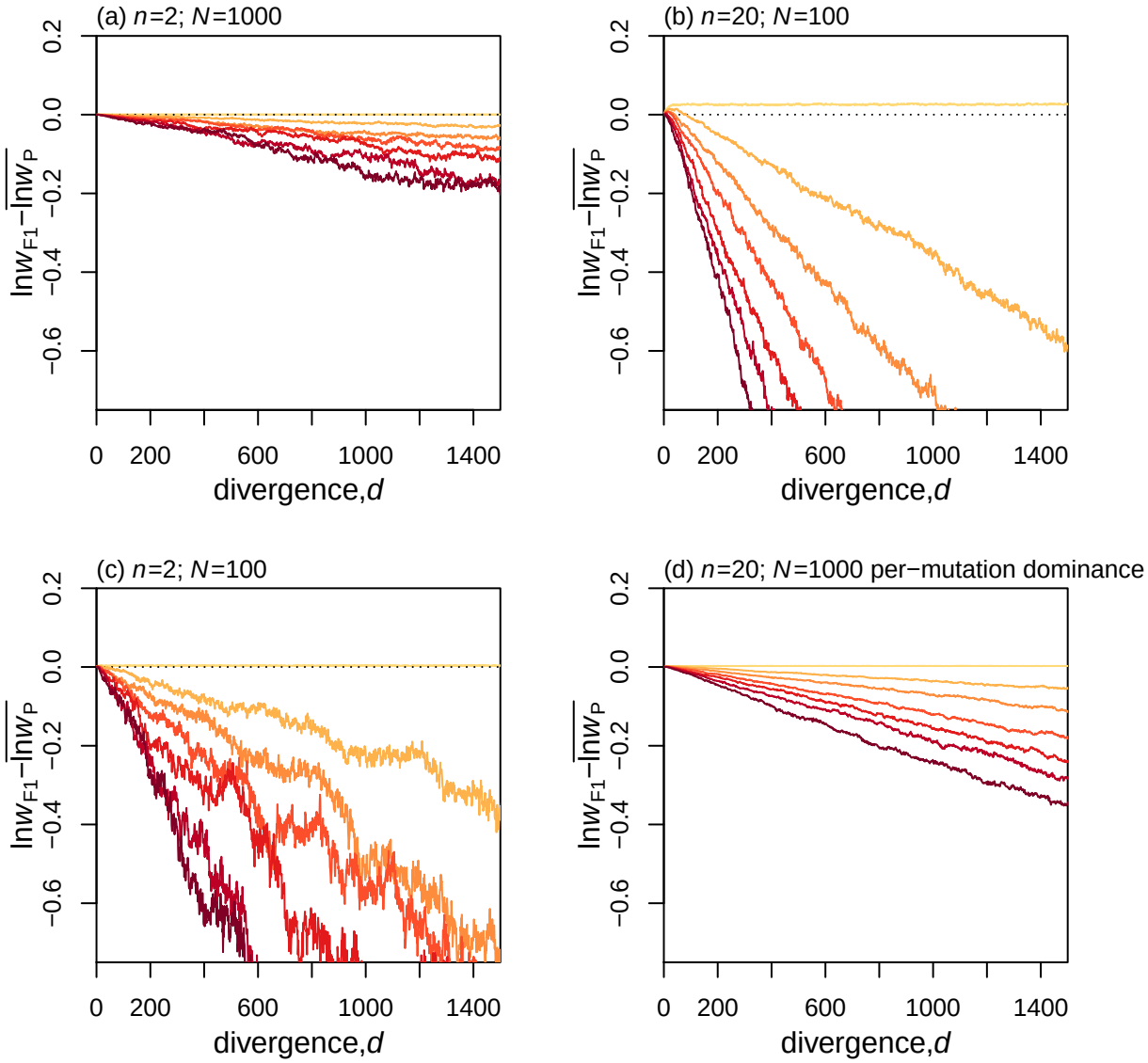

Figure S3: The F1 clock tick rate varies with various model parameters. **(a)-(c)**: Results replicate those shown in Figure 2b, except that the simulation parameters differed in terms of the number of traits (a), the population size (b), or both (c). In all cases, the slopes get steeper with $V_{\delta_{\text{mut}}}$, but the decrease is not always linear, showing that $V_\delta \approx V_{\delta_{\text{mut}}}$ does not always hold. **(d)**: Results also hold in simulations where each mutation had a single value of δ_{mut} applying to all n traits; such that heterozygous effects differed from the homozygous effects in their length but not in their direction. This shows that the effects of dominance reported here do not depend on the variation in direction.

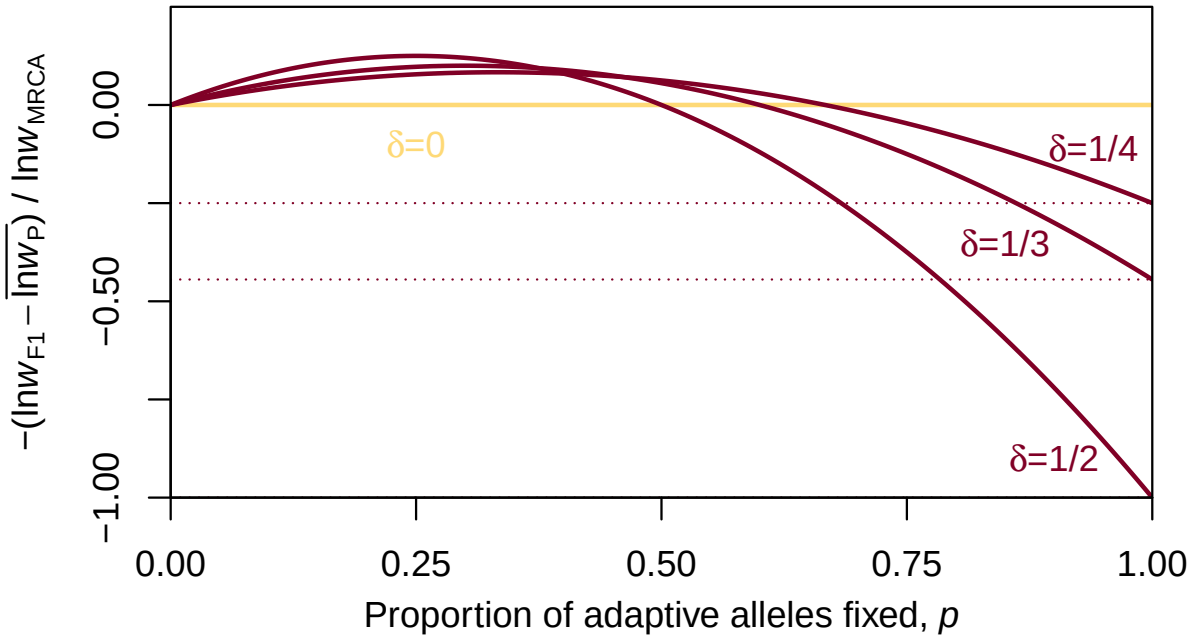

Figure S4: Directional dominance leads to a transient fitness increase and then a permanent decrease. Log F1 fitness relative to current and pre-adaptation parental fitness when parental lines are approaching a new optimum through the fixation of identical substitutions with equally sized additive effects and dominance effects (see eq. 37). The size of the dominance coefficients is denoted δ. The yellow line $\delta = 0$ represents additivity, and the dark red curves represent dominance coefficients of different magnitudes. With dominance, F1 fitness first exceeds parental fitness due to the directional dominance overshoot, but as parental lines get closer to the new optimum this overshoot becomes deleterious and F1 fitness declines. The horizontal dotted lines give the predicted overshoot when the optimum has been reached.

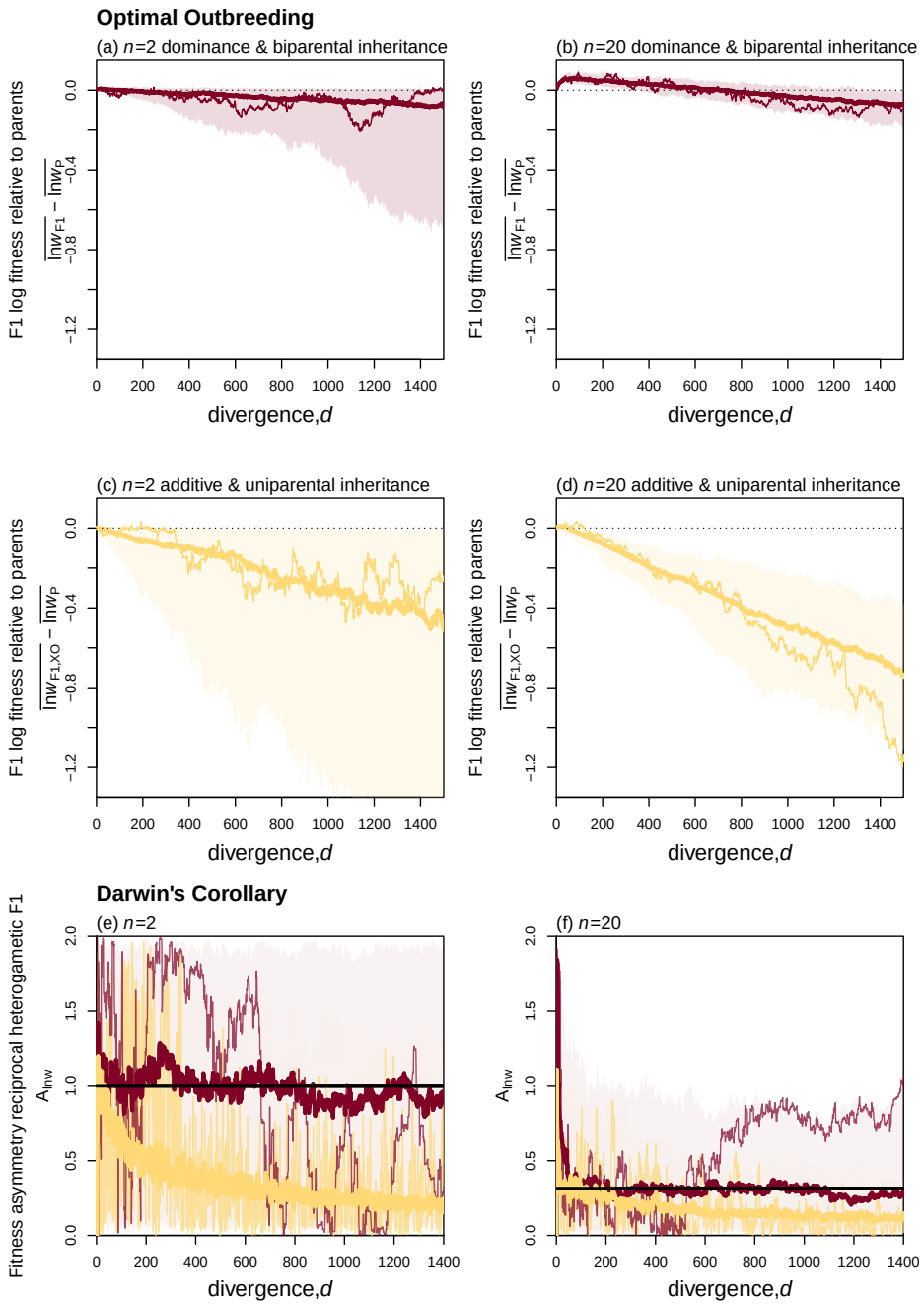

Figure S5: Major patterns of F1 fitness for simulated F1 hybrids using the dominance function illustrated in Fig. S1. Simulation results are reported comparing phenotypic additivity (yellow curves) to phenotypic dominance, using the dominance function shown in Fig. S1, where mutations of larger effect had more extreme levels of dominance (dark red curves). Results correspond to **(a)**-**(b)**: Fig. 1c; **(c)**-**(d)**: Fig. 4b; **(d)**-**(e)**: Fig. 5c-d. In all cases, the average over 100 replicates (thicker solid lines) is compared to a single randomly chosen replicate (thinner lines), to illustrate the stochasticity. Heterogametic F1 in (c)–(f) used ($x = \frac{1}{3}$).

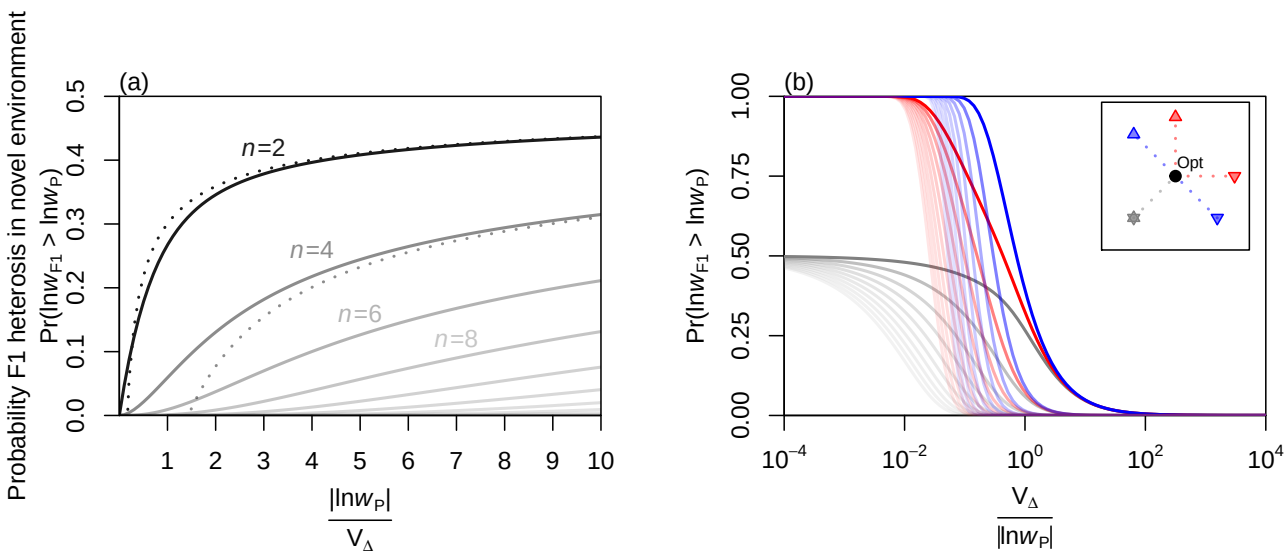

Figure S6: The probability of F1 heterosis in a novel environment depends on the number of traits, and the parental phenotypes. Panel **(a)** and gray lines in **(b)** represent the scenario from Figure 3a, when the two parental lines have similar and equally maladapted phenotypes. **(a)**: a replication of Figure 3b, but without log transformation on the x-axis, highlighting the decline in the probability of heterosis with n. **(b)**: the probability of heterosis increases if the parents are maladapted on different traits (red curves), or in opposite phenotypic directions (blue curves). The positions of the parents relative to the optimum are shown in the inset. Colour intensity indicates the value of n which ranges from 2 (bright lines) to 20 (faint lines), and in all cases, the probability of heterosis declines steeply with n and with $V_\Delta/|\ln w_\mathrm{P}|$. The functions plotted are described in Appendix 1.

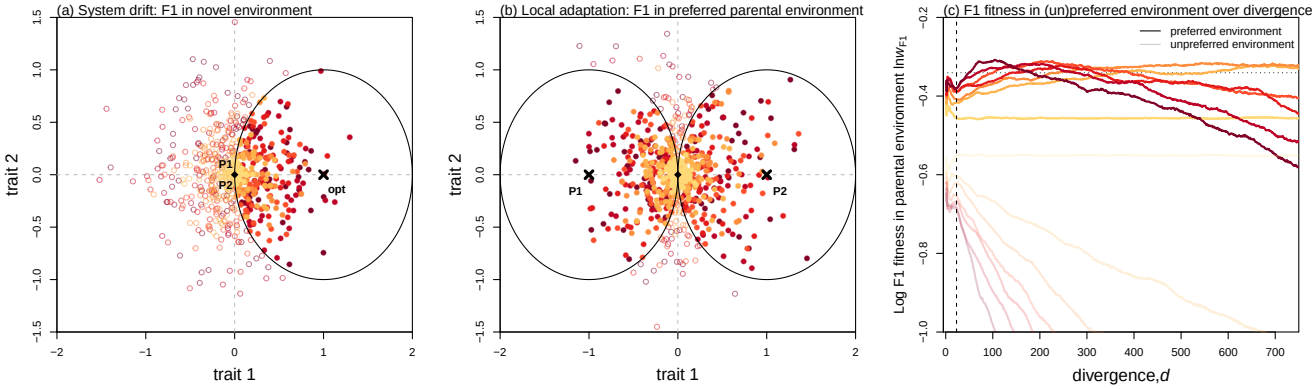

Figure S7: Lucky beneficial effects of dominance in novel environments, with variable levels of mutational dominance. **(a)**–**(b)**: Simulated F1 hybrids after 50 divergent substitutions with different levels of variation in dominance show "Lucky" beneficial effects in a new environment. **(c)**: Simulations of the expected log fitness of the F1 between locally adapted parents, scored in the preferred parental habitat, to which the F1 is better adapted (solid lines) or the unpreferred parental habitat, to which it is less adapted (faint lines). These results match Figure 3d, but include results for all seven different levels of variance among the dominance coefficients of the new mutations, $V_{\delta_{\mathrm{mut}}}$. Colours for the $V_{\delta_{\mathrm{mut}}}$ values match those in Figure 2, with darker colours indicating higher mutational variance.

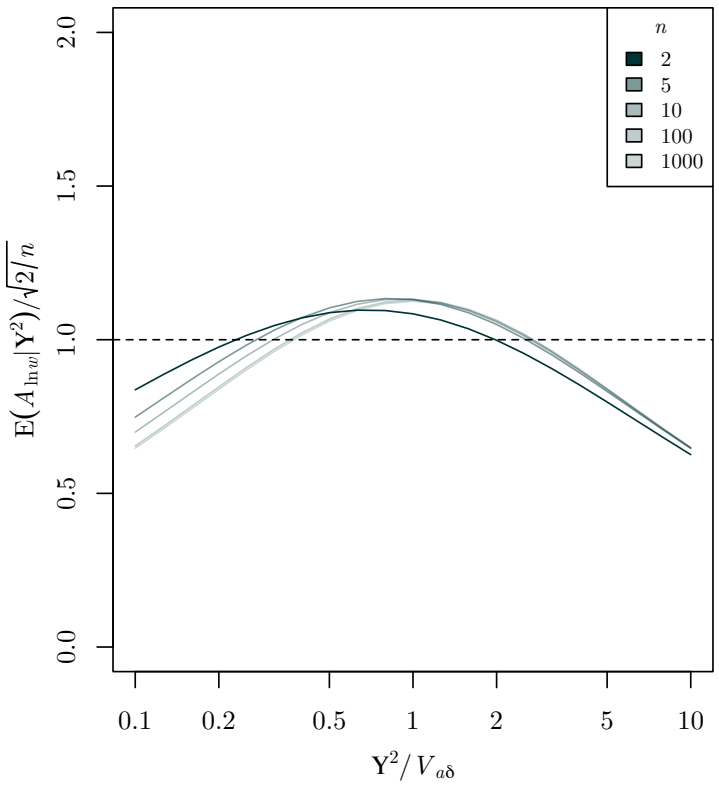

Figure S8: Derivation of the expected levels of log fitness asymmetry between cross directions with variable phenotypic dominance. The expected value of eq. 55 is shown for various values of n, namely (dark-to-light) $n = 2, 5, 10, 100$ and 1000. The expectation (evaluated numerically) is close to the value of the crude approximation $\sqrt{2/n}$ (eq. 20), unless the quantity $\Upsilon^2/V_{a\delta}$ deviates greatly from unity. This quantity compares the deviations due to dominance, and the deviations due to lack of autosomal co-adaptation with the absent paternal X chromosome (see Figure 4a), and depends mainly on the variance in the dominance effects and the proportion of uniparentally inherited loci (eq. 57).

Appendix 1: Derivations

In this Appendix, we derive the analytical results in the main text. The relevant main text equations are listed next to each section.

1 F1 fitness with biparental inheritance

Many results in the main text are given in terms of the total dominance deviations on each trait, that is, the Δ_i as defined in eq. 4. The strongest assumptions that we use are the normality and isotropy of the Δ_i (i.e. that all Δ_i have equal variances). However, neither assumption is required for all results, and the assumption of isotropy is made largely for simplicity.

1.1 Expected log fitness (eqs. 6–11)

s

To calculate the expected log fitness of the F1, we begin by expanding the brackets in eq. 5:

$$\ln w_{\mathrm{F1}} = \ln w_{\mathrm{MP}} - 2\sqrt{|\ln w_{\mathrm{MP}}|}\Delta_1 - \sum_{i=1}^{n} \Delta_i^2 \tag{22}$$

Then we take expectations over the possible values of the dominance deviations, to account for their relative independence of the form of selection acting on the parental lines.

$$E\left(\ln w_{\mathrm{F1}} | \ln w_{\mathrm{MP}}\right) = \ln w_{\mathrm{MP}} - 2\sqrt{|\ln w_{\mathrm{MP}}|}E\left(\Delta_1\right) - \sum_{i=1}^{n} E\left(\Delta_i^2\right) \tag{23}$$

Equations 9–10 follow directly, if we focus only on trait 1.

1.1.1 Stabilizing selection (eqs. 6–8)

Next, if there is no directional dominance, i.e., if all $E(\Delta_i) = 0$, then:

$$E\left(\ln w_{\mathrm{F1}} | \ln w_{\mathrm{MP}}\right) = \ln w_{\mathrm{MP}} - \sum_{i=1}^{n} \mathrm{Var}\left(\Delta_i\right), \qquad E(\Delta_i) = 0 \tag{24}$$

and so eq. 6 follows directly if we use the notation:

$$V_{\Delta} \equiv \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}\left(\Delta_i\right) \tag{25}$$

For eq. 7, we use equivalent notation to refer to the effects of the individual substitutions, which under stabilizing selection, we can treat as increments in a random walk, with variance

$$V_{a\delta} \equiv \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}\left(a_i \delta_i\right) \tag{26}$$

where $a_i \delta_i$ is a randomly chosen dominance effect on trait i, and the expectations are over realizations of the evolutionary process. We can also define an equivalent quantity for the additive effects.

$$V_a \equiv \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}\left(a_i\right) \tag{27}$$

where a_i is a randomly chosen additive effect on trait i. Equation 8 now holds generally if we define the equivalent quantity for dominance coefficients, V_δ as the ratio of the two quantities above

$$V_\delta \equiv \frac{V_{a\delta}}{V_a} \tag{28}$$

such that it indicates the size of the dominance effect relative to the additive effect of substitutions. V_δ is not generally equivalent to the variance in dominance coefficients, but it is equivalent in the special case of isotropy, and no covariance between the squares of the additive and dominance effects. To see this, we can also write $V_{a\delta}$ in terms of moments of additive effects and dominance coefficients, where similar to the above we use δ_i to refer to the dominance coefficient on trait i of a randomly chosen substitution. With exchangeable parental lines, and recalling that the effects of a substitution are defined with respect to P1, irrespective of whether they are ancestral or derived, it follows that $E\left(a_i\right) = E\left(\delta_i\right) = 0$. Looking at the other moments, we find

$$E\left(a_i \delta_i\right) = \mathrm{Cov}\left(a_i, \delta_i\right) \tag{29}$$

$$E\left(a_i^2 \delta_i^2\right) = \mathrm{Cov}\left(a_i^2, \delta_i^2\right) + \mathrm{Var}\left(a_i\right) \mathrm{Var}\left(\delta_i\right) \tag{30}$$

Now we can write $V_{a\delta}$, as

$$V_{a\delta} = \frac{1}{n} \sum_{i=1}^{n} E\left(a_i^2 \delta_i^2\right) - E^2\left(a_i \delta_i\right) \tag{31}$$

without directional dominance $\mathrm{Cov}\left(a_i, \delta_i\right) = 0$, and if we also assume that larger substitutions have no tendency to have more extreme dominance coefficients, then $\mathrm{Cov}\left(a_i^2, \delta_i^2\right) = 0$, and we get

$$V_{a\delta} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}\left(a_i\right)\mathrm{Var}\left(\delta_i\right), \qquad \mathrm{Cov}\left(a_i, \delta_i\right) = \mathrm{Cov}\left(a_i^2, \delta_i^2\right) = 0 \tag{32}$$

Under isotropy, the distribution of substitutions is the same across traits, so that $V_a = \mathrm{Var}\left(a_i\right)$ and $V_\delta = \mathrm{Var}\left(\delta_i\right)$ are both constant. Thus, with these additional assumptions we can again obtain eq. 8, but with a more intuitive intepretation of V_δ. Note, however, that data do not support the assumption that $\mathrm{Cov}\left(a_i^2, \delta_i^2\right) > 0$ (Orr, 1991; Manna et al., 2011).

1.1.2 Directional selection (eqs. 9–11)

Under directional selection on trait 1, we expect additive effects in the direction of the optimum to be more dominant. Therefore, for these adaptive substitutions (denoted by an asterisk $*$), $\mathrm{Cov}\left(a_1^*, \delta_1^*\right)$ will not vanish. However, we can obtain results in a simplified case in which the initial bout of adaptation took place in both parental lineages, taking them both from an old optimum to a new optimum, and where these optima differed only on trait 1. We will assume that the adaptation took place in exactly d_{dir} substitutions, and that these substitutions affected only trait 1. We will further assume that their additive effects and dominance coefficients covary in sign, but not in size, such that $\mathrm{Cov}\left(|a_1^*|, |\delta_1^*|\right) = 0$. In other words, large adaptive substitutions are no more likely to be dominant or recessive than small adaptive substitutions, but substitutions pointing from the MRCA towards the optimum (adaptive substitutions that are derived in P2) have $\delta_1^*, a_1^* > 0$ whereas substitutions pointing towards the MRCA and away from the optimum (adaptive substitutions derived in P1) have $\delta_1^*, a_1^* < 0$.

Taken together, the chain of adaptive substitutions runs from the new optimum to the old optimum and then back, and so its total length is $2d_{\mathrm{dir}}\,|a_1^*|$, and we have $d_{\mathrm{dir}}\,|a_1^*| = \sqrt{|\ln w_{\mathrm{MRCA}}|}$, where w_{MRCA} is the fitness of ancestral genotype (adapted to the old optimum), in the environment characterized by the new optimum. With these assumptions, we can now derive the moments of Δ_1 after both the adaptation, and a further period of $d - d_{\mathrm{dir}}$ substitutions, accrued under stabilizing selection:

$$
\begin{aligned}
E\left(\Delta_1\right) &= E\left[\sum_{j=1}^{d_{\mathrm{dir}}} a_{1j}^{*}\delta_{1j}^{*} + \sum_{j=d_{\mathrm{dir}}+1}^{d} a_{1j}\delta_{1j}\right] \\
&= d_{\mathrm{dir}} E\left[a_1^{*}\delta_1^{*}\right] \\
&= \sqrt{\left|\ln w_{\mathrm{MRCA}}\right|} E\left[\left|\delta_1^{*}\right|\right]
\end{aligned} \tag{33}
$$

$$
\begin{aligned}
E\left(\Delta_1^2\right) &= E\left[\left(\sum_{j=1}^{d_{\mathrm{dir}}} a_{1j}^{*}\delta_{1j}^{*} + \sum_{j=d_{\mathrm{dir}}+1}^{d} a_{1j}\delta_{1j}\right)^2\right] \\
&\approx d_{\mathrm{dir}}^2 E\left[\left(a_1^{*}\delta_1^{*}\right)^2\right] + \left(d - d_{\mathrm{dir}}\right) E\left[\left(a_1\delta_1\right)^2\right] \\
&= \left|\ln w_{\mathrm{MRCA}}\right| V_{\delta_{\mathrm{dir}}} + \left(d - d_{\mathrm{dir}}\right) \mathrm{Var}\left(a_1\delta_1\right)
\end{aligned} \tag{34}
$$

where

$$
V_{\delta_{\mathrm{dir}}} \equiv E\left[\delta_1^{*2}\right] \tag{35}
$$

Here, we used the random-walk assumptions, described above, for the substitutions accrued after the adaptation. Equation 34 also uses the stronger assumption that covariance between the dominance effects of the adaptive substitutions can also be ignored. After making these assumptions, eq. 11 in the main text follows directly.

We can also give a more complete result that applies during the ongoing adaptation to the new optimum. During the adaptation phase i.e. before the parental lines have reached the optimum, we use $p = \frac{z_{\mathrm{MRCA}} - z_{\mathrm{MP}}}{z_{\mathrm{MRCA}}}$ to denote the proportion of the distance to the new optimum the parental lines have already travelled. Considering only the trait under directional selection, we find

$$
E\left(\ln w_{\mathrm{F1}}\right) = \ln w_{\mathrm{MP}} - 2p(1-p)E\left[\left|\delta_1^{*}\right|\right]\left|\ln w_{\mathrm{MRCA}}\right| - p^2\left|\ln w_{\mathrm{MRCA}}\right| V_{\delta_{\mathrm{dir}}} \tag{36}
$$

For the overshoot of the F1, we find

$$
E\left(\frac{\ln w_{\mathrm{F1}} - \ln w_{\mathrm{MP}}}{\ln w_{\mathrm{MRCA}}}\right) = -p^2 V_{\delta_{\mathrm{dir}}} - 2p(1-p)E\left[\left|\delta_1^{*}\right|\right] \tag{37}
$$

These expressions are plotted in Figure S4 for substitutions with dominance coefficients of identical size $|\delta|$.

1.2 Variance in F1 log fitness (eq. 12)

To calculate eq. 12 we also require the variance in log F1 fitness. Its behaviour is easiest to see if we assume normality and independence of the Δ_i. In this case, we find:

$$\mathrm{Var}\left(\ln w_{\mathrm{F1}}\right) = E\left(\mathrm{Var}\left(\ln w_{\mathrm{F1}} \left|\ln w_{\mathrm{MP}}\right.\right)\right) + \mathrm{Var}\left(E\left(\ln w_{\mathrm{F1}} \left|\ln w_{\mathrm{MP}}\right.\right)\right)$$

$$= 4E\left(\ln w_{\mathrm{MP}}\right) V_{\Delta_1} + 2\sum_{i=1}^{n} V_{\Delta_i}^2 + \mathrm{Var}\left(\ln w_{\mathrm{MP}}\right), \qquad E(\Delta_i) = 0 \tag{38}$$

If the midparent remains well adapted, but the dominance deviations are large, then the second term will come to dominate. Because V_Δ grows with the divergence, this will typically hold at large divergences, when $d \gg 1$, and so

$$\mathrm{Var}\left(\ln w_{\mathrm{F1}}\right) \approx 2nV_\Delta^2, \qquad E(\Delta_i) = 0, d \gg 1 \tag{39}$$

Using the same limit in eq. 6 yields eq. 12.

1.3 F1 fitness in preferred parental environment (eqs. 14, 15)

When the parental lines are adapted to two different optima, then the F1 will tend to be fitter in one of the two parental environments (i.e. its "preferred" environment). In that environment, we can write its log fitness using a small modification of eq. 22:

$$\ln w_{\mathrm{F1}_{\mathrm{pref}}} = -\left(\sqrt{|\ln w_{\mathrm{MP}}|} - |\Delta_1|\right)^2 - \sum_{i=2}^{n} \Delta_i^2$$

$$= \ln w_{\mathrm{MP}} - \sum_{i=1}^{n} \Delta_i^2 + 2\sqrt{|\ln w_{\mathrm{MP}}|}|\Delta_1| \tag{40}$$

If we assume that the parental lines are each optimal in one of the habitats, then distance of midparent from either optimum is half of the distance between the optima, such that $\ln w_{\mathrm{MP}} = \ln w_{\mathrm{P}}/4$, where $\ln w_{\mathrm{P}}$ is the log fitness of the maladapted parental line in any given habitat. Next, we assume that Δ_1 is normal, with a vanishing mean, so that $E(|\Delta_1|) = \sqrt{\mathrm{Var}(\Delta_1)2/\pi}$. Equation 14 then follows directly if we assume equal variances for the dominance deviations on all traits, so that $\mathrm{Var}(\Delta_1) = V_\Delta$.

To derive equation 15, we differentiate eq. 14 by V_Δ, and set to zero. This yields the value of V_Δ that maximises the expected log fitness. Substituting that value into eq. 14 yields eq. 15.

As with eq. 39, we can also derive the variance in $\ln w_{\mathrm{F1_{pref}}}$ using the assumptions of normality, independence and vanishing means for the Δ_i:

$$\mathrm{Var}\left(\ln w_{\mathrm{F1_{pref}}}\right) = 2nV_\Delta^2 - 4E\left(\ln w_{\mathrm{MP}}\right)V_{\Delta_1}\left(1-2/\pi\right) - 4E\left(\sqrt{|\ln w_{\mathrm{MP}}|}\right)\sqrt{V_{\Delta_1}2/\pi}$$

$$+\,\mathrm{Var}\left(\ln w_{\mathrm{MP}} + 2\sqrt{|\ln w_{\mathrm{MP}}|}\sqrt{V_{\Delta_1}2/\pi}\right) \tag{41}$$

At large divergences, the first term will come to dominate, supporting the assertion in the main text that eq. 12 still holds for $\ln w_{\mathrm{F1_{pref}}}$.

1.4 Probability of lucky heterosis (eq. 13)

Now let us consider the F1 between phenotypically identical parental lines, that are both maladapted. To calculate the probability of their F1 being heterotic (Fig 3b), we first note that, with isotropy, such that all $V_{\Delta_i} = V_\Delta$, then, from eq. 5, the quantity $|\ln w_{\mathrm{F1}}|/V_\Delta$ is the sum of n squared normal variables, one with a non-zero mean of $\sqrt{|\ln w_{\mathrm{MP}}|/V_\Delta}$. As such, this quantity will follow a non-central chi-squared distribution, with n degrees of freedom, and non-centrality parameter given by $|\ln w_{\mathrm{MP}}|/V_\Delta = |\ln w_{\mathrm{P}}|/V_\Delta$. Here, we have used the fact that the midparent will be identical to the two (identical) parental phenotypes. We can write the cumulative distribution function of this variable as

$$\mathrm{cdf}(|\ln w_{\mathrm{F1}}|/V_\Delta) = 1 - Q_{n/2}(\sqrt{|\ln w_{\mathrm{P}}|/V_\Delta}, \sqrt{|\ln w_{\mathrm{F1}}|/V_\Delta}) \tag{42}$$

where we have used the Marcum Q function (Marcum, 1950). The F1 will be heterotic if it is closer to the optimum than the midparent and so

$$\Pr\left(\ln w_{\mathrm{F1}} > \ln w_{\mathrm{P}}\right) = \Pr\left(\frac{\ln w_{\mathrm{F1}}}{V_\Delta} > \frac{\ln w_{\mathrm{P}}}{V_\Delta}\right)$$

$$= 1 - Q_{n/2}(\sqrt{|\ln w_{\mathrm{P}}|/V_\Delta}, \sqrt{|\ln w_{\mathrm{P}}|/V_\Delta}) \tag{43}$$

This is the exact result plotted in Figure 3b. This result can also be written in terms of modified Bessel functions of the first kind, as follows:

$$1 - Q_{n/2}(\sqrt{u}, \sqrt{u}) = \frac{1}{2} - e^{-u} \left[\frac{I_0(u)}{2} + \sum_{j=1}^{n/2-1} I_j(u) \right] \tag{44}$$

(Helstrom, 1995). Approximating these functions as $I_j(u) = \frac{e^u}{\sqrt{2\pi u}}$ (see, e.g. Guo et al., 2001), yields eq. 13 of the main text.

Figure S6 shows similar results when the parental lines are equally maladapted, but no longer phenotypically identical. To derive these results, we simply change the non-centrality parameter (i.e., the first argument of the Q function in eq. 43), which represents the position of the midparent. When the parents are maladapted in opposite directions, then the midparent matches the optimum, such that this argument is 0 (blue points and lines in Fig. S6). When the parents are maladapted in orthogonal directions (e.g., on different phenotypic traits), then this argument is $\sqrt{4|\ln w_{\mathrm{P}}|/V_\Delta}$ (red points and lines in Fig. S6).

1.5 Transgressive segregation in the F1

The procedure used in the previous section can also be used to calculate the probability of transgressive variation in a single trait: i.e., an F1 trait value that lies outside of the range of values among the parental lines (Stelkens and Seehausen, 2009). If there is no directional dominance, then the mean F1 trait value will be the midparental value, and its variance will be equal to the variance of the dominance deviations.

$$E(z_{\mathrm{F1}}) = \frac{z_{\mathrm{P2}} - z_{\mathrm{P1}}}{2} \tag{45}$$

$$\mathrm{Var}(z_{\mathrm{F1}}) = V_\Delta. \tag{46}$$

where we have ordered the parents by their trait value. Now, assuming normality, the probability of the F1 trait value lying outside of the parental range is

$$\Pr\left(z_{\mathrm{F1}} < z_{\mathrm{P1}} | z_{\mathrm{F1}} > z_{\mathrm{P2}}\right) = 1 - \mathrm{erf}\left(\frac{z_{\mathrm{P2}} - z_{\mathrm{P1}}}{2\sqrt{V_\Delta}}\right) \tag{47}$$

$$\approx \exp\left(-\frac{z_{\mathrm{P2}} - z_{\mathrm{P1}}}{2\sqrt{V_\Delta}}\right) \tag{48}$$

Because V_Δ grows linearly with d, the probability of transgressive variation will increase with d if the divergence between the parental phenotypes also increases with d (at least as fast as $\sqrt{d}$). If, by

contrast, the difference between the parental traits remains static, then the probability of transgressive variation will decrease with d. This is illustrated in Figure S2. Both regimes were observed by Stelkens and Seehausen (2009).

2 Uniparental inheritance and recombinant hybrids

In this section, we present results for a more general type of hybrid, in which some loci are either hemizygous or homozygous. To do this, we will assume that the the parental lines are under stabilizing selection, as described above, such that the dominance deviations undergo a random walk in phenotypic space. We also use existing results for recombinant hybrids, conditional on the F1 fitness (Simon et al., 2018; Schneemann et al., 2020; De Sanctis et al., 2022 in prep.).

2.1 Expected log fitness and bounds on V_δ (eqs. 16–18 and 21)

An arbitrary hybrid can be characterized by its hybrid index, $0 \leq h \leq 1$ (i.e. the proportion of divergent alleles from one of the parental lines) and by its heterozygosity, $0 \leq p_{12} \leq 1$ (the proportion of divergent loci carrying one allele from each parental line). Let us first begin by defining A_i as the analogue to Δ_i for the additive effects.

$$A_i = \sum_{j=1}^{d} a_{ij} \tag{49}$$

With this notation, the midparental log fitness can be written as

$$\ln w_{\mathrm{MP}} = \overline{\ln w_{\mathrm{P}}} + \sum_{i=1}^{n} A_i^2 \tag{50}$$

It then follows from eq. 3, that when the populations are maladapted from their optimum in random directions, such that $E\left(\ln w_{\mathrm{MP}}\right) = \frac{\overline{\ln w_{\mathrm{P}}}}{2}$, then

$$E\left(\sum_{i=1}^{n} A_i^2\right) = -\frac{\overline{\ln w_{\mathrm{P}}}}{2} \tag{51}$$

Now, using the assumptions discussed above for stabilizing selection, we can combine eq. 51 with the published results, to yield the expected log fitness of an arbitrary hybrid.

$$E\left(\ln w_{\mathrm{H}}\right)=\overline{\ln w_{\mathrm{P}}}\left(1-2h(1-h)\right)-p_{12}E\left(\sum_{i=1}^{n}\Delta_i^2\right)-\left(4h(1-h)-p_{12}\right)E\left(\sum_{i=1}^{n}\sum_{j=1}^{d}a_{ij}^2\right)$$

$$=\overline{\ln w_{\mathrm{P}}}\left(1-2h(1-h)\right)-p_{12}dnV_{\delta}V_a-\left(4h(1-h)-p_{12}\right)dnV_a \tag{52}$$

Note that eq. 52 contains three terms, the second capturing the deleterious effects of dominance, and the third capturing the deleterious effects of lack of coadaptation among the additive effects. Equation 52 also reduces to results in the main text for a fully heterozygous F1, for which $h = 1/2$ and $p_{12} = 1$. Equation 21 follows from comparing this F1 to the arbitrary hybrid, when parents are well adapted ($\overline{\ln w_{\mathrm{P}}} = 0$). Equations 16–18 also follow simply. For an XO F1 hybrid, all of the heterozygosity is found on the autosomes, such that $p_{12} = 1-x$ while the hybrid index is $h = (1\pm x)/2$, depending on which parental line contributed the X. It follows, therefore, that $2h(1-h) = (1-x^2)/2$, and $4h(1-h)-p_{12} = x(1-x)$.

2.2 Darwin's corollary (eq. 20)

For the asymmetry of log fitness between the cross directions, we will begin by assuming that the parental phenotypes – and therefore the midparental phenotype – are optimal. This means that the reciprocal XO F1, carrying different X chromosomes, will deviate in equal and opposite extents from the optimum. We will denote these deviations as $\sqrt{dn}\Upsilon$, reflecting their growth with the number of traits, and with divergence (e.g. eq. 52). The dominance deviations, which come from their shared autosomes, will be in consistent directions. As such, we can orientate the trait axes so that these deviations fall on trait 1. This means that the difference in their log fitnesses can be written as:

$$\ln w_{\mathrm{F1,♀♂}}-\ln w_{\mathrm{F1,♂♀}}=\left(\Delta_1+\sqrt{dn}\Upsilon\right)^2-\left(\Delta_1-\sqrt{dn}\Upsilon\right)^2,\qquad \ln w_{\mathrm{MP}}=0$$

$$=4\sqrt{dn}\Upsilon\Delta_1 \tag{53}$$

while the mean across the cross directions yields:

$$\frac{1}{2}\left(\ln w_{\mathrm{F1,♀♂}}+\ln w_{\mathrm{F1,♂♀}}\right)=\frac{1}{2}\left[\left(\Delta_1+\sqrt{dn}\Upsilon\right)^2+\left(\Delta_1-\sqrt{dn}\Upsilon\right)^2+2\sum_{i=2}^{n}\Delta_i^2\right],\qquad \ln w_{\mathrm{MP}}=0$$

$$=dn\Upsilon^2+\sum_{i=1}^{n}\Delta_i^2 \tag{54}$$

We now treat the dominance deviations as normal variables, such that $\Delta_i = \sqrt{d\mathrm{Var}\left(a_i\delta_i\right)}\zeta_i$, where ζ_i is a standard normal random variable, with vanishing mean and unit variance. The measure of asymmetry is therefore

$$A_{\ln w} = \frac{|\ln w_{\mathrm{F1},♀♂} - \ln w_{\mathrm{F1},♂♀}|}{\frac{1}{2}|\ln w_{\mathrm{F1},♀♂} + \ln w_{\mathrm{F1},♂♀}|} = \frac{4\sqrt{n\mathrm{Var}\left(a_1\delta_1\right)}|\zeta_1|}{\sum_{i=1}^{n}\mathrm{Var}\left(a_i\delta_i\right)\zeta_i^2 + n\Upsilon^2}$$

$$= 4\sqrt{n\Upsilon^2/V_{a\delta}}\frac{|\zeta_1|}{\sum_{i=1}^{n}\zeta_i^2 + n\Upsilon^2/V_{a\delta}}, \qquad \mathrm{Var}\left(a_i\delta_i\right) = V_{a\delta} \tag{55}$$

So this quantity is not expected to change in any consistent way with the genetic divergence, d. We have not been able to evaluate this expression analytically, but numerical integration shown in Figure S8, shows that, to a rough approximation

$$E\left(A|\Upsilon^2\right) \approx \sqrt{\frac{2}{n}}, \qquad \Upsilon^2/V_{a\delta} \sim 1 \tag{56}$$

which agrees with the simulation results reported in the main text (Figure 5c-d) and in Figure S5e-f. We can also anticipate the typical value of $\Upsilon^2/V_{a\delta}$, using eq. 16. This shows that

$$E\left(\Upsilon^2/V_{a\delta}\right) = \frac{x(1-x)}{V_\delta}, \qquad \ln w_{\mathrm{MP}} = 0 \tag{57}$$

suggesting that $\Upsilon^2/V_{a\delta} \sim 1$ will often hold for plausible parameter values.

Appendix 2: Simulation procedure

To complement our analytical predictions, we used individual-based simulations of evolutionary divergence under Fisher's geometric model. The simulation procedures were closely based on those reported in Schneemann et al. (2020), and for compatibility with earlier work, we included a constant into the log fitness function, such that:

$$\ln w = -\alpha \left\| \mathbf{z} - \mathbf{o} \right\|^2 \tag{58}$$

All simulations used $\alpha = 1/2$. In all cases, we simulated pairs of allopatric diploid populations, which were initially identical with no segregating variation. The ancestral population was phenotypically optimal for all simulations except for those presented in Figures 2e–f (modelling directional selection) and 3d (modelling locally adapted parents), where the ancestor was placed at unit distance from the optimum, such that the ancestral population had a relative fitness of $e^{-\alpha} \approx 61\%$, and the simulations began with a bout of directional selection towards the optimum. In all cases, the diverging populations were of constant size N individuals ($2N$ chromosomes), and contained simultaneous hermaphrodites with discrete non-overlapping generations.

To produce the next generation, $2N$ prospective parents were sampled with replacement, with probabilities proportional to their fitness. Gametes were produced with free recombination among all loci, and assuming infinite sites, and subject to mutation. The total number of mutations was drawn from a Poisson distribution with mean $2NU$ and distributed randomly across the $2N$ gametes. There was no back mutation. The mutation rate U was chosen such that multiple mutations generally segregated together ($NU \geq 1$). The additive effects of each mutation on each each trait, were drawn independently from a normal distribution, with vanishing means and variance denoted $V_{a_{\text{mut}}}$. We set this mutational variance at:

$$V_{a_{\text{mut}}} = -\frac{\bar{s}\Gamma\left(\frac{n}{2}\right)}{8\alpha\Gamma\left(\frac{n+2}{2}\right)} \tag{59}$$

Because the homozygous effect of a mutation was twice its additive effect, eq. 59 ensures that the mean selection coefficient acting on a homozygous mutation occurring in an optimally fit individual was approximately $-\bar{s}$, irrespective of the number of traits under selection, n.

The heterozygous effect of a mutation was equal to its additive effect, a_{mut}, multiplied by $(1 + \delta_{\text{mut}})$, where δ_{mut} is its dominance coefficient. For simulations under phenotypic additivity, we simply set

$\delta_{\text{mut}} = 0$ for all mutations. With variable phenotypic dominance, the δ_{mut} were drawn from shifted beta distributions, which we bounded at -1 (for a wholly recessive mutation) and 1 (for a wholly dominant mutation).

For the simulations reported in Figures 2, 3d, 5, and S3, we used the same shifted beta distribution for all mutations, regardless of their additive effects. To do this, we set set $E(\delta_{\text{mut}}) = 0$ and $\text{Var}(\delta_{\text{mut}}) = V_{\delta_{\text{mut}}}$. These distributions are shown in Figure 2a. Generating dominance coefficients in this way allowed us to adjust the amount of variance in dominance coefficients of new mutations $V_{\delta_{\text{mut}}}$, which was reflected in V_δ.

For the simulations reported in Figures 1c, 4b and S5, we generated the δ_{mut} values to reflect some of the known properties of large and small effect mutations. Specifically, small effect mutations were close to additive with little variance, whereas large effect mutations were predominantly recessive but with a higher variance, such that some would be dominant (Orr, 1991; Manna et al., 2011). This was achieved by sampling the dominance multipliers from a beta distribution such that

$$\text{E}(\delta_{\text{mut}}) \equiv \mu = 1 - \frac{2}{1 + \exp\left(-2\frac{|a_{\text{mut}}|}{\sqrt{V_{a_{\text{mut}}}}}\right)} \tag{60}$$

and

$$\text{Var}(\delta_{\text{mut}}) = \mu(\mu^2 - 1) \tag{61}$$

The distributions generated by eqs. 60–61 are illustrated in Figure S1.

For each set of parameters, we simulated 200 independent populations that each evolved until they fixed 1000 substitutions. We then randomly paired populations to obtain 100 pairs for the formation of F1 hybrids. We considered the substitutions that had fixed in the number of generations it took for the first population of the pair to reach 1000 substitutions. To produce the F1, for standard biparental inheritance (and also for homogametic hybrids), we took the heterozygous effect of all substitutions. For our heterogametic hybrids, which were our example case of uniparental inheritance, we assigned a random proportion x of substitutions to be on the X chromosome, and considered their homozygous rather than their heterozygous effects. This is equivalent to a form of dosage compensation, involving up-regulation on the X (see main text). Note that these X-linked substitutions were chosen after simulating the divergence. This meant that all sites were biparentally inherited and expressed during the divergence process. This guaranteed that, for our simulations, the variance in dominance coefficients of fixed mutations did not

differ between the X and autosomes, with both equal to the mutational input, such that ($V_\delta \approx \mathrm{V}_{\delta_{\mathrm{mut}}}$) This made it easier to illustrate the conditions for Haldane's Rule, eq. 18 (see Figure 5), but it is unlikely to hold in nature (Charlesworth et al., 1987, see also Discussion).