The genetics of speciation: insights from Fisher's Geometric Model

Christelle Fraïsse[1,2,3], P. Alexander Gunnarsson[3], Denis Roze[4,5,], Nicolas Bierne[1,2], John J. Welch[3*]

1. Université Montpellier, Institut des Sciences de l'Évolution, UMR 5554, Montpellier Cedex 05, France.

2. CNRS, Institut des Sciences de l'Évolution, UMR 5554, OREME Station Marine, Sète, France.

3. Department of Genetics, University of Cambridge, Downing Street, Cambridge, UK.

4. CNRS, UMI 3614, Evolutionary Biology and Ecology of Algae, Roscoff, France.

5. Sorbonne Universités, UPMC University Paris VI, Roscoff, France.

∗ To whom correspondence should be addressed (j.j.welch@gen.cam.ac.uk).

Running title:

Fisher's Geometric Model and Speciation

Key words:

Reproductive isolation, Dobzhansky-Muller incompatibilities, snowball effect, hybridization

Statistics:

- Figures: 5 (main text) + 14 (supp. info.)
- Tables: 2 (main text) + 4 (supp. info.)
- Conflicts of interest: None

Abstract

Research in speciation genetics has uncovered many robust patterns in intrinsic reproductive isolation, and fitness landscape models have been useful in interpreting these patterns. Here we examine fitness landscapes based on Fisher's geometric model. Such landscapes are analogous to models of optimizing selection acting on quantitative traits, and have been widely used to study adaptation and the distribution of mutational effects. We show that, with a few modifications, Fisher's model can generate all of the major findings of introgression studies (including "speciation genes" with strong deleterious effects, complex epistasis and asymmetry), and the major patterns in overall hybrid fitnesses (including Haldane's Rule, the speciation clock, heterosis, hybrid breakdown, and male-female asymmetry in the F1). We compare our approach to alternative modelling frameworks that assign fitnesses to genotypes by identifying combinations of incompatible alleles. In some cases, the predictions are importantly different. For example, Fisher's model can explain conflicting empirical results about the rate at which incompatibilities accumulate with genetic divergence. In other cases, the predictions are identical. For example, the quality of reproductive isolation is little affected by the manner in which populations diverge.

Introduction

Intrinsic reproductive isolation involves regions of the genome that confer high fitness in their original genetic background (or in conspecific matings), but low fitness in hybrid genetic backgrounds (or allospecific matings). Empirical studies have revealed several robust patterns about such reproductive isolation, each of which has been reported in at least two kingdoms (Coyne and Orr 2004; Gavrilets 2004). These patterns are summarised in Table 1, and details are given in Supplemental Table S1.

One set of patterns concerns overall hybrid fitness, and includes suprising differences in hybrid fitness with the sex of the offspring ("Haldane's Rule"; Haldane 1922), or the sex of the parents (Darwin 1859, Ch 8; Turelli and Moyle 2007). Another set of patterns stems from the pioneering experiments of Dobzhansky, which involved introgressing small regions of the genome from one species to another. These studies have shown that very small chromosomal regions ("speciation genes") can have strong negative fitness effects in the heterospecific background, and that introgressions often show complex epistasis.

Dobzhansky also played a central role in a second development: using fitness landscape models to interpret these experimental data (Dobzhansky 1937; Coyne and Orr 2004; Gavrilets 2004). For example, low hybrid fitness between a pair of populations might be taken to imply that one of the diverging populations had "passed through a fitness valley", with the state of low fitness forming an intermediate step in their divergence. Dobzhansky (1937) showed that this inference was mistaken. With a two-locus model, populations can evolve strong reproductive isolation without the low-fitness genotypes appearing during their divergence. The key component is strong negative epistasis between some combination of alleles at the two loci. Dobzhansky's work shows clearly how incorrect assumptions about the fitness landscape can lead to incorrect inferences about the mode of population divergence, and the requirements for reproductive isolation to evolve.

Most subsequent theoretical work has extended Dobzhansky's model to many loci using an "incompatibility-based" approach. In these models, each novel combination of alleles at each pair of loci has some fixed probability of being incompatible; models track the accumulation of incompatibilities over time, and then make assumptions about how these incompatibilities combine to yield overall hybrid fitness (Muller 1942; Orr 1995; Coyne and Orr 2004; Gavrilets 2004; Welch 2004; Turelli and Moyle 2007; Bank *et al*. 2012; Fraïsse *et al*. 2014). This approach has been highly productive, but there are other ways to extend Dobzhansky's two-locus model. For example, "holey landscape" models assign low fitness values to complete genotypes, rather than to combinations of alleles, and they make some importantly different predictions (Gavrilets and Gravner 1997; Gavrilets 2004; Welch 2004).

A third approach is to use fitness landscapes based on "Fisher's geometric model" (Fisher 1930; Orr 1998),

which derives predictions about the fitness effects of mutations using a simple model of optimizing selection acting on multiple quantitative traits. This approach has been used to make testable predictions about the genetics of adaptation and the distribution of mutational fitness effects (Mani and Clarke 1990; Orr 1998; Welch and Waxman 2003; Martin and Lenormand 2006a,b; Martin *et al.* 2007; Tenaillon *et al.* 2007); and some of these predictions show a remarkably good fit to data (e.g., Martin and Lenormand 2006a,b; Martin *et al.* 2007; Manna *et al.* 2011). The approach has also been used to study speciation (Mani and Clarke 1990; Barton 2001; Chevin *et al.* 2014), because Fisher's model naturally generates sign epistasis in fitness (e.g. an allele that increases height might be beneficial in short individuals, but deleterious in tall individuals), but previous analyses have found that this epistasis is usually quite weak (Barton 2001; Martin *et al.* 2007; Chevin *et al.* 2014; see below), which contradicts the speciation genetics data. However, these analyses have used versions of the model where fitness declines slowly with the distance to the optimum (see below), and other authors have studied generalised versions of the model, in which the level of fitness epistasis can be tuned (Peck *et al.* 1997; Martin and Lenormand 2006b; Tenaillon *et al.* 2007; Gros *et al.* 2009). Here, we investigate such a generalised form of Fisher's geometric model, and ask whether it can generate the major qualitative features of the speciation genetics data (Tables 1 and S1). We also compare its predictions to those of incompatibility-based models, to ask whether these predictions are robust.

Model

Fitness landscape

The data summarised in Tables 1 and S1 provide very clear replicated patterns, but are not generally suitable for detailed model fitting. As such, most results below use the simplest version of Fisher's model that might account for these patterns, without adding the complications required for a fully quantitative approach (e.g. Turelli 1985; Zhang and Hill 2003; Martin and Lenormand 2006a; see also Appendix section 3).

With this in mind, we will represent an individual's phenotype as an n-dimensional vector, $\mathbf{z} = \{z_1, z_2, ..., z_n\}$, whose components z_i, are their phenotypic values for each trait. The fitness of this phenotype depends on the optimal value of each trait, which we collect in a second vector $\mathbf{o} = \{o_1, o_2, ..., o_n\}$. We will assume that selection is independent and equally strong on all n traits (isotropic selection), and that the relative fitness of an individual with phenotype $\mathbf{z}$ is

$$w(\mathbf{z}) = \exp\left(-\sigma||\mathbf{z}-\mathbf{o}||^k\right), \tag{1}$$

where $\|\mathbf{z}-\mathbf{o}\| \equiv \sqrt{\sum_{i=1}^{n}(z_i-o_i)^2}$ is the Euclidean distance of the phenotype from the optimum, σ is the overall

strength of selection, and k determines the curvature of the fitness function. The parameter k is closely related to measures of fitness epistasis (Peck *et al.* 1997; Gros *et al.* 2009), and so can be used to tune the level of epistasis. Mutations are assumed to shift the phenotype relative to its present value in an additive fashion (i.e. translational mutation), and so, denoting the vector of mutational changes as $\mathbf{r}$, the selection coefficient of the mutation, s, is defined as follows.

$$s \equiv \frac{w(\mathbf{z}+\mathbf{r})}{w(\mathbf{z})} - 1 \tag{2}$$

Again, for simplicity, we will assume that mutations are equally likely to occur in all possible "directions" in phenotypic space (i.e. isotropic mutation), and that they affect all of the n traits under consideration (i.e. universal pleiotropy). Despite this last assumption, most results below can be applied to models with restricted pleiotropy (e.g., Welch and Waxman 2003; Chevin *et al.* 2010; Lourenço *et al.* 2011) by interpreting n as referring to "pleiotropy" for results that concern single mutations/introgressions, but as the total number of traits (sometimes called "complexity"), for results that concern total fitness. In both cases, however, we stress that the z_i cannot generally be equated with the values of any measurable quantitative trait, at least in any simple or direct way (see, e.g., Waxman and Welch 2005; Johnson and Barton 2005; Martin and Lenormand 2006a; Lourenço *et al.* 2011; Martin 2014). As such, n is best viewed as a phenomenological parameter of a fitness landscape model, which is flexible and small (few parameters).

Simulation methods

We simulated the evolution of reproductive isolation under the above model, again choosing the simplest assumptions that can generate the speciation genetic patterns of interest. We consider a population of $2N$ hermaphrodites, which reproduce with discrete, non-overlapping generations. Genomes are haploid, and are assumed to be large enough for each new mutation to occur at a unique site (i.e. infinite sites). We draw the magnitude of each new mutation from some distribution. Several choices were possible here, but our general strategy was to fix the mean strength of selection against deleterious mutations, while other parameters varied (see Appendix section 2.1 for full details). Mutations occur stochastically each generation, but we assume that the mutation rate is low enough to consider the fate of mutations one at a time (Orr 1998; Barton 2001; Chevin *et al.* 2014), and generate only the exponentially-distributed waiting time between them. Each mutation reaches fixation with probability $(1-\exp(-2N_e s/N)) \big/ (1-\exp(-2N_e s))$ (Fisher 1930; Kimura 1983).

At the start of each simulation, the population reached a stochastic mutation-selection-drift equilibrium at the

initial optimum, before splitting into two populations, each of N individuals, which then evolve in allopatry. For most simulations, the phenotypic optima for all traits remained fixed at the origin. However, to simulate divergence via positive selection, we allowed the optimum of one of the n traits to move over time (see below and Appendix section 2.2 for full details). All simulations were written in C and are included as supplementary information.

Results

1. Between-species introgressions

In this section, we ask whether Fisher's geometric model can generate the patterns observed in experimental studies that involve the introgression of small sections of chromosome between species (Tables 1 and S1).

1.1 Evolution of speciation genes A basic observation of speciation genetics is that some alleles generate a very large negative fitness effect in the heterospecific background, so large that they must have been subject to much weaker selection during species divergence (Dobzhansky 1937; Coyne and Orr 2004). To characterise such "speciation genes", we define s_i as the loss of relative fitness associated with introgressing a divergent allele into a heterospecific background; we also define s_d as the maladaptedness of the donor population at the time that this allele became fixed. If $s_i \gg s_d$, then we can say that a population has fixed a speciation gene, without passing through a fitness valley to do so. In Appendix section 1.2, we derive the probability that a mutation will reach fixation under Fisher's geometric model, as a function of the ratio (s_i/s_d) and these results are summarised in Figure 2.

1.1.1 Evolution of speciation genes via genetic drift

Figure 2a shows results when the diverging parental populations are sufficiently well adapted for selection to be ineffective on all fixable mutations ($N_e s_d = 0.1$), such that divergence must take place via genetic drift. In Figure 2a, fixation probabilities are compared to that of a globally neutral mutation ($P_{neut} = 1/N$), and results are shown for different values of k (the curvature of the fitness function; Fig. 1) and n (the number of traits affected by each mutation). Figure 2a suggests that speciation genes can fix with a non-negligible probability, but only when n is sufficiently small, and k is sufficiently large. Most notably, mutations that cause the most serious fitness problems in hybrids ($s_i/s_d \gg 1$) are very unlikely to reach fixation when $k = 2$ (lightest curves). To understand these results further, it is helpful to express the ratio s_i/s_d in terms of other model parameters. If we denote as r, the phenotypic

size of the mutation, and z_d the distance from the optimum of the donor species, measured before the appearance of the mutation, then we find (see Appendix section 1.2):

$$s_i/s_d \approx (r/z_d)^k \tag{3}$$

As such, speciation genes will only appear if some fixations "overshoot" the optimum during divergence (i.e. $r > z_d$). Under Fisher's model, larger mutations are more likely to be deleterious, and less likely to fix via drift (Fisher 1930; Kimura 1983). Furthermore, mutations of any given size are less likely to fix as n increases, and as k increases (see eqs. 31-34). However, the decline with k is more than counterbalanced by its role in eq. 3, which shows that larger values of k allow small overshoots to create much larger fitness differences in the donor and recipient backgrounds, and this leads to the pattern shown in Figure 2a.

Equation 3 also shows that the rate at which speciation genes evolve will be difficult to predict. This is because the distribution of phenotypes that appear during divergence (the z_d values) and the sizes of the mutations that reach fixation (the r values), depend in a complex way on all of the model parameters, including the rate and distribution of mutations that arise, and the strength and efficacy of selection acting on these mutations (see, e.g., Poon and Otto 2000; Lourenço *et al.* 2011; Roze and Blanckaert 2014 and Appendix section 5). However, the important point here is that such mutations are expected to fix occasionally under a wide range of conditions.

These conclusions are confirmed by simulations of population divergence in allopatry, followed by introgressions of divergent alleles (Figure 3a, top row). For example, Figure 3a shows that when $n = 5$ and $k = 6$, fixations may occur via drift, with very small selection coefficients ($|s| \sim 1/N = 0.002$), and yet cause fitness losses of 50% or more ($-s_i > 0.5$) when introgressed into the heterospecific background. These findings are robust to varying the distribution of mutation sizes (Figure S2), but results are much less impressive when $n = 50$ (Figs. S1d, S2d, S7c), and when k is reduced (Figs. S1, S2). Indeed, speciation genes never evolve when $k = 2$ (Figs. S1a,c, S2a,c, S7b). Results are also robust to using the model of Martin and Lenormand (2006a,b) which relaxes the assumptions of isotropic selection and mutation (see Appendix section 3, and Figs. S9-S10).

1.1.2 Evolution of speciation genes via positive selection

Several studies indicate that loci involved in reproductive isolation diverged via positive selection rather than drift (Coyne and Orr 2004; Presgraves 2010a). Because their negative effects in hybrids are larger than the effects of most benefical mutations (e.g. Sella *et al.* 2009), $s_i/s_d \gg 1$ should still hold for these alleles. Figure 2b plots

fixation probabilities via positive selection as a function of s_i/s_d, and compares these fixation probabilities to that of the best possible mutation (which takes the population directly to the optimum, and so fixes with probability $P_{best} \approx 2(N_e/N)s_d$ (e.g. Fisher 1930). Overall, results in Figure 2b closely resemble those in Figure 2a, and again suggest that speciation genes can fix with a non-negligible probability as long as n is sufficiently small and k sufficiently large.

To illustrate these results, we repeated our simulations with a larger population size (reducing the effects of drift), and moving the optimal value of one of the n traits in a smooth cyclic fashion, simulating introgressions when the optimum had returned to its initial value (see eq. 54, Table S2 and Appendix section 2.2 for full details). Figure 3b shows an example where almost all of the divergence was due to positive selection, and where some of the fixed mutations caused strongly deleterious introgressions when introgressed between the backgrounds.

As with divergence under drift, the rate at which these speciation genes evolve depends in a complex way on most details of the model (Figs. S3, S4, S8), but with a moving optimum, one crucial factor is the population's ability to track that optimum (Chevin *et al.* 2014). In general, faster environmental change leads to the fixation of larger-effect mutations, which are more likely to act as speciation genes (eq. 3; Fig. S8). By contrast, if the rate of environmental change is held constant, then speciation genes evolve at very similar rates when the two diverging populations adapt to the same new optima (as in "mutation-order speciation"; Mani and Clarke 1990; Gavrilets 2004), or to different optima (as in "ecological speciation", Coyne and Orr 2004; Gavrilets 2004; Fig. S8a). The ability of a population to track the optimum is also greatly affected by n (Orr 2000; Poon and Otto 2000; Welch and Waxman 2003; Lourenço *et al.* 2011; Roze and Blanckaert 2014; Matuszewski *et al.* 2014; see also Appendix section 5). In our simulations large n populations were often severely maladapted, and this greatly increased the rate at which speciation genes evolved (Figs. S3-S4, S8) (this was not apparent from the analytical results, which assumed that the recipient populations were phenotypically optimal). In all cases, maladapted recipient populations will be closer to regions of very low fitness, and this makes it more likely that introgressions will have severe negative effects (compare, e.g., Figure 3 to Figure S6 where the same simulated introgressions are made into optimal backgrounds).

Adding environmental change to the model brings one final complication: when environments can vary, it becomes important to distinguish between reproductive isolation that is extrinsic (environment-dependent) rather than intrinsic (Chevin *et al.* 2014). We neglect this complication here, by measuring all introgression effects in a common environment, however, our results are qualitatively unchanged if we follow a stricter definition of intrinsic isolation (Chevin *et al.* 2014), and consider only fitness effects associated with the $n-1$ traits whose optima never change - although all selective effects are, predictably, reduced in this case (Figure S5).

1.2 Asymmetry of introgression fitness effects An introgression is said to be asymmetric if an allele introgressed from species 1 to species 2 has a strong negative effect on fitness, while its orthologue introgressed from species 2 to species 1 has little or no effect. Relevant data are understandably scarce, but the best-studied cases do show strong asymmetry (Table S1; Coyne and Orr 2004, Ch. 7), although a largely symmetric incompatibility was uncovered segregating in *Drosophila melanogaster* (Corbett-Detig *et al.* 2013; see their supplemental Fig. 4b).

To understand the predictions made by Fisher's model, we will measure asymmetry via the ratio of log fitnesses,

$$A = \frac{\ln w_{i+}}{\ln w_{i-}} \tag{4}$$

where the $w_{i\pm}$ denote the fitness of genotypes where orthologous alleles have been introgressed in opposite directions. Strong asymmetry is implied if $|\ln A| \gg 0$. The distribution of $\ln A$ is easiest to derive if we consider a mutation of size $r \equiv \|\mathbf{r}\|$, and assume that both pure species populations have the same suboptimal phenotype (realised by different genes), whose Euclidean distance from the optimum is denoted $z \equiv \|\mathbf{z}-\mathbf{o}\|$. With these assumptions, we show in Appendix section 1.3 that

$$E\left[|\ln A|\right] \approx 1.6\frac{k}{\sqrt{n}}\frac{(r/z)}{1+(r/z)^2}, \tag{5}$$

$$\mathrm{CV}\left[|\ln A|\right] \approx 0.76 \tag{6}$$

where CV[.] is the coefficient of variation (standard deviation divided by the absolute mean). It follows that asymmetry can be strong when $k/\sqrt{n}$ is large, and this is illustrated in Figure 4. It also follows that asymmetry is maximised when $r = z$, but declines towards zero if either quantity is much larger than the other (Fig. 4). As such Fisher's model predicts that introgressions will never be asymmetric when mutations have very small phenotypic effects ($r \to 0$), and nor when the recipient species are phenotypically optimal ($z \to 0$) - a fact that follows directly from the symmetry of the fitness function.

This would seem to imply that many incompatibilites will be symmetrical, but in fact, mutations of very small phenotypic effect ($r \ll z$) are unlikely to have strong fitness effects when introgressed in either direction, while mutations of very large phenotypic effect ($r \gg z$) are unlikely to fix during population divergence. The resulting bias toward asymmetrical incompatibilities is confirmed by simulations (Figure 3, middle row), which show a strong tendency for the most deleterious introgressions to be most asymmetric. Again, this result is robust to varying

several aspects of the model, but consistent with eq. 5, asymmetry is much weaker when $k/\sqrt{n}$ is small (Figs. S1-S4, S9 middle rows). Figure S6 also confirms that asymmetry is non-existent when the recipient population is perfectly adapted.

1.3 Multifactor introgressions and complex epistasis When multiple chromosomal regions are introgressed, these often show strong fitness epistasis with each other, and this can result in "complex incompatibilities", where severe loss of fitness occurs only if multiple factors are introgressed together (Table S1; Coyne and Orr 2004; Fraïsse *et al.* 2014). Such interactions can be described by another ratio of selection coefficients:

$$C = \frac{|s_{i,double}|}{\max(|s_{i,single}|)} \tag{7}$$

where $s_{i,double}$ is the fitness loss caused by introgressing two mutations together, and $\max(|s_{i,single}|)$ is the maximum loss of fitness caused by introgressing either mutation on its own. Complex incompatibilities are defined by very high values of C. To explore the behaviour of C under Fisher's model, consider the introgression of two mutations with sizes r_1 and r_2 into a recipient background with phenotypic distance z, and assume that the first introgression is more deleterious, and that $r_1 > r_2$. In this case, we have the following approximations (see Appendix section 1.4).

$$E[C] \approx \frac{\left(z^2 + r_1^2 + r_2^2\right)^{k/2} - z^k}{\left(z^2 + r_1^2\right)^{k/2} - z^k} \tag{8}$$

$$\mathrm{Std}[C] \propto \frac{k}{\sqrt{n}} \tag{9}$$

(We note that these crude approximations also apply when we relax our assumption of universal pleiotropy; Welch and Waxman 2003; Chevin *et al.* 2010; Lourenço *et al.* 2011). From these results, it follows that the mean and variance of C will both increase with k, suggesting that complex incompatibilities are more likely to evolve when $k > 2$. They are also more likely when the recipient background is well adapted ($z \ll r$), and when both introgressions are of similar size ($r_1 \approx r_2$); and in this case, we have $E[C] \approx 2^{k/2}$.

These conclusions are again confirmed by simulations (Figures 3, S1-S4, and S9 bottom rows). In Figure 3 (bottom row), the effects of double and single introgressions are strongly correlated, but there are several cases that would be described as complex incompatibilities, such as pairs of introgressions which individually reduce fitness

by ~10% or less, but which cause an ~80% reduction when introgressed together. Results are less impressive with higher values of n, even in parameter regimes where single introgressions of large effect are generated (e.g., Fig. S4d). This is not only because of the reduced variance in C (eq. 9), but also due to the greater maladaptation of the recipient genotypes reducing $E[C]$ (eq. 8; c.f. Fig. S6).

A final, and important implication of these results is that introgressions with very strong fitness effects are even more likely to evolve if "single-factor introgressions" actually contain many divergent sites. This possibility is supported by fine-scale dissection studies (Davis and Wu 1996; Maside and Naveira 1996), and evidence that individual speciation genes have undergone high rates of amino-acid substitution (Presgraves 2010a). With multi-factor introgressions, very large fitness effects can appear with higher values of n, and with lower values of k (Figs. S1-S4, S7c, S8c). However, with $k = 2$, the weak epistasis means that strongly deleterious introgressions very rarely appear, even when the introgressions include very large numbers of divergent sites (Fig. S7b, S10a, c).

1.4 The snowball effect Incompatibility-based models predict a "snowball effect", in which the probability that an introgression reduces fitness increases with the genetic distance (number of substitutions) separating the donor and recipient species (Orr 1995). This prediction stems directly from the model's assumptions about mapping fitnesses to genotypes, and so the snowball effect is not predicted by holey landscape models (Gavrilets and Gravner 1997; Gavrilets 2004; Welch 2004), and nor by Fisher's model. Under Fisher's model, the probability of an introgression being deleterious depends solely on the mutation's phenotypic effect and the recipient's pre-introgression phenotype, and so the genetic distance between the donor and recipient species plays no role.

Evidence for the snowball effect in nature is contradictory (Presgraves 2010b; Table S1). Strong evidence of snowballing has come from direct crossing studies, following Matute *et al.* (2010; Table S1). Analyses showing no snowball effect have followed the bioinformatic approach of Kondrashov *et al.* (2002; Table S1); this approach identifies single amino-acid mutations with known deleterious effects in one species, that are fixed as the wild type in other species - implying that a single amino-acid introgression between these species would have a deleterious effect.

Although the direct crossing approach is much closer to natural hybridisation, both approaches are valid tests of the snowball effect, and so their contradictory results - now replicated in several systems (Table S1) - are surprising. To explain this contradiction, it has been hypothesized that almost all amino acid substitutions occur in epistatically interacting pairs (Kondrashov *et al.* 2002; Welch 2004; Presgraves 2010b), but Fisher's model suggests another possible explanation. Under Fisher's model, introgressions containing many divergent sites will tend to have larger phenotypic effects, and so be more deleterious, than introgressions containing fewer divergent sites (see, e.g., eqs.

3 and 8). When introgressions include multiple sites, the number of divergent sites in each introgression will tend to increase with the overall genetic divergence between the species, and so this could lead to an "apparent snowball effect". The effect is "apparent", because it depends on the level of divergence within each introgressed segment of chromosome, and not on the divergence in the remainder of the genome. As such, Fisher's model predicts an apparent snowball effect for introgression studies involving small chromosomal regions (such as Matute *et al.* 2010), but no snowball effect for bioinformatic studies, where the "introgressions" always comprise a single divergent amino acid (such as Kondrashov *et al.* 2002).

To illustrate this possibility, Figure 5 shows results from simulated introgressions. When we simulated introgressions that comprised single divergent mutations (replicating the bioinformatic studies), then the proportion of these introgressions that were strongly deleterious remained constant, regardless of the genetic divergence between the species (Fig. 5a). In Figure 5b, exactly the same simulated data were used, but we assigned each of the divergent sites, at random, to one of 5000 genomic segments, and then simulated the introgression of each segment as a whole (replicating the direct crossing studies). In this case, the proportion of the 5000 introgressions that were deleterious increased rapidly with the genetic divergence - in accordance with predictions of a snowball effect. These results were confirmed by formal tests, which yielded evidence of a snowball effect for multi-site introgressions, but no such effect for single-mutation introgressions (see Table S3 and Appendix section 6 for full details). The same results were observed with different parameter regimes and variants of the model (see Figs. S7, S8, S10 and Table S3), and with different numbers of genomic segments, although the rate of accumulation starts to slow when most genomic segments cause strongly deleterious effects (Fig. 5c).

Fisher's model shows how the conflicting experimental results might be explained, without there being a true snowball effect, and without the assumption that all amino acid substitutions occur in coupled pairs.

2. Overall hybrid fitness

In this section, we consider data from overall hybrid fitness, usually in diploids (Tables 1 and S1). With Fisher's model, divergence with diploidy is complicated by high levels of transient overdominance (Barton 2001; Manna *et al.* 2011; Sellis *et al.* 2011; Connallon and Clark 2014; Draghi and Whitlock 2015). Nevertheless, following Barton (2001), simple arguments can usually show whether the model can account for these data.

2.1 Recombinant hybrids The simplest predictions from Fisher's model concern recombinant hybrids. These predictions derive from a haploid model, but also apply to the F2 cross in diploids (Barton 2001; Chevin *et al.* 2014). For parental species that differ by d substitutions, we show in the Appendix section 4, that

$$\mathrm{E}\left[\ln w_{\mathrm{F2}}\right] \quad \propto \quad -d^{k/2} \tag{10}$$

$$\mathrm{CV}\left[\ln w_{\mathrm{F2}}\right] \quad \approx \quad \frac{k}{\sqrt{2n}}, \qquad n \gg k \tag{11}$$

and so, for all values of k, the model predicts a decline in mean hybrid fitness with the genetic distance, d (Barton 2001; Chevin *et al.* 2014), but no change in the coefficient of variation in log fitness (see also Figure S11, and eq. 61, which relaxes the condition that $n \gg k$). Quantitative tests of these predictions are difficult with existing data, but the qualitative prediction of eq. 10 is well supported by observations of an "F2 speciation clock" (Edmands 2002; Table S1).

2.2 The F1 cross: parental sex asymmetry A remarkable pattern in the initial F1 cross is the common asymmetry in the fitnesses of male-female and female-male crosses of the same species pair (Table S1; Darwin 1859, Ch. 8; Turelli and Moyle 2007). Such asymmetry even appears in simultaneous hermaphrodites, such as the freshwater snail *Physa* (=*Physella*) *acuta*, and the sea squirt complex *Ciona intestinalis*/*robusta*, where some divergent populations produce fit hybrids when one species takes the male role, and sterile hybrids when the same species takes the female role (Escobar *et al.* 2008; Sato *et al.* 2014; Bouchemousse *et al.* 2016).

Such asymmetry implies the existence of "parent-of-origin" effects on fitness, and in Fisher's framework, these can be modelled as distinct paternal and maternal contributions to the phenotype. With additivity, any phenotype can be written as the sum of maternal, paternal and biparental contributions. If we further neglect genetic variation within the parental species (labelled P1 and P2), then their phenotypes can be written as

$$\mathbf{z}_{\mathrm{P1}} \quad \equiv \quad 2\mathbf{b}_1 + \mathbf{m}_1 + \mathbf{p}_1 \tag{12}$$

$$\mathbf{z}_{\mathrm{P2}} \quad \equiv \quad 2\mathbf{b}_2 + \mathbf{m}_2 + \mathbf{p}_2 \tag{13}$$

where $\mathbf{m}_i$, $\mathbf{p}_i$, and $\mathbf{b}_i$ denote the phenotypic effects of maternally, paternally and biparentally inherited loci from parental species i. We will now introduce new notation to describe the midparental phenotype, $\mathbf{z}_{mp}$, and the phenotypic differences between species that are encoded by uniparental loci, $\mathbf{r}_u$.

$$\mathbf{z}_{mp} \equiv \frac{1}{2}(\mathbf{z}_{\mathrm{P1}} + \mathbf{z}_{\mathrm{P2}}) \tag{14}$$

$$\mathbf{r}_u \equiv \frac{1}{2}((\mathbf{m}_1 - \mathbf{m}_2) + (\mathbf{p}_2 - \mathbf{p}_1)) \tag{15}$$

With this notation, it becomes clear that the phenotypes of the reciprocal F1s (labelled $\mathrm{F1}_{12}$ and $\mathrm{F1}_{21}$), have equal, but opposite deviations from the midparent.

$$\begin{aligned} \mathbf{z}_{\mathrm{F1}_{12}} &\equiv \mathbf{b}_1 + \mathbf{b}_2 + \mathbf{m}_1 + \mathbf{p}_2 \\ &= \mathbf{z}_{mp} + \mathbf{r}_u \end{aligned} \tag{16}$$

$$\begin{aligned} \mathbf{z}_{\mathrm{F1}_{21}} &\equiv \mathbf{b}_2 + \mathbf{b}_1 + \mathbf{m}_2 + \mathbf{p}_1 \\ &= \mathbf{z}_{mp} - \mathbf{r}_u \end{aligned} \tag{17}$$

It follows that our earlier results for the asymmetry of single introgressions (eqs. 5-6; Fig. 4) also apply to overall hybrid fitness, if we (re)define $z \equiv \|\mathbf{z}_{mp}\|$ and $r \equiv \|\mathbf{r}_u\|$. To express these results in terms of genetic divergence (the number of substitutions), we note that phenotypic changes at the uniparental loci (which contribute to $\mathbf{m}_i$ or $\mathbf{p}_i$) can always be "compensated" by changes elsewhere in the genome(s) that contribute to $\mathbf{b}_i$. As such, the evolution of $\mathbf{r}_u$, considered without the compensatory changes, can be approximated as a random walk in phenotypic space (see, e.g., Appendix eq. 60), and so we have

$$E\left[r^2\right] \propto d_u \tag{18}$$

$$\mathrm{CV}\left[r^2\right] \approx \sqrt{2/n} \tag{19}$$

where d_u is the genetic divergence (number of substitutions) at uniparental loci. Putting these results together, Fisher's model predicts that parental sex asymmetry can be highly stochastic, particularly when n is small, but on average, it should increase with initial parental divergence, before peaking when $r = z$, and then declining again at high divergences (eq. 5; Figure 4). Curves of this form have been observed in data from toads (Brandvain *et al.* 2014) and centrachid fishes (Bolnick and Near 2005; see also Turelli and Moyle 2007; Table S1).

We note that parent-of-origin effects, as modelled here, encompass a wide range of phenomena, including true sex linkage, uniparentally-inherited cytoplasmic elements (organelles or symbionts), and imprinting or silencing of various kinds (Turelli and Moyle 2007). All of these phenomena have been clearly implicated in cases of asymmetric reproductive isolation (Turelli and Moyle 2007; Vrana 2007; Greiner *et al.* 2011; Burton and Barreto 2012; Lima 2014), and their relative importance is difficult to assess - particularly given uncertainty about the taxonomic distribution of imprinting (Ohlsson 2013). Nevertheless, examples involving organelle genomes seem particularly common (Greiner *et al.* 2011; Burton and Barreto 2012); it is suggestive, for example, that hermaphroditic animals showing strong asymmetry at low genetic distances also have mitochondria that are unusually rapidly evolving (Gissi *et al.* 2004; Nolan *et al.* 2014; Sato *et al.* 2014; Bouchemousse *et al.* 2016).

2.3 The F1 cross: optimal outbreeding A second, very common pattern, is for the F1 to have higher fitness than the parental lines (heterosis or hybrid vigour), with low fitness appearing only in the F2 and later crosses (hybrid breakdown; Edmands 2002; Table S1). Barton (2001) showed that this pattern is readily generated by Fisher's model, for simple geometrical reasons (see below and Figs. S12-S13). However, Barton's model predicted heterosis regardless of the genetic distance between the parents (Barton 2001), while data show a steady decline in F1 fitness at higher divergences. The overall pattern yields an intermediate level of "optimal outbreeding", and at larger divergences, an "F1 speciation clock" (Table S1; Edmands 2002; Coyne and Orr 2004, Ch. 2).

Here we show that the observed patterns in F1 fitness are predicted by Fisher's model, once we have included parent-of-origin effects - whose presence can be inferred from the widespread phenomenon of parental-sex asymmetry, as explained above. To see this, we define as $\mathbf{r}_{tot}$ the total phenotypic difference between the parental species.

$$\mathbf{r}_{tot} \equiv \frac{1}{2}\left(\mathbf{z}_{\mathrm{P1}} - \mathbf{z}_{\mathrm{P2}}\right) \tag{20}$$

The magnitude of this vector, $r_{tot} \equiv \|\mathbf{r}_{tot}\|$ can be contrasted with the differences in phenotype due to uniparental loci alone, $r \equiv \|\mathbf{r}_u\|$, as used above (see eq. 15). With this notation, we can now compare the fitnesses of the parents and hybrids as follows.

$$H \equiv \frac{\ln w(\mathbf{z}_{\mathrm{P1}}) + \ln w(\mathbf{z}_{\mathrm{P2}})}{\ln w(\mathbf{z}_{\mathrm{F1}_{12}}) + \ln w(\mathbf{z}_{\mathrm{F1}_{21}})} \approx \left(\frac{z^2 + r_{tot}^2}{z^2 + r^2}\right)^{k/2} \tag{21}$$

$$= (r_{tot}/r)^k, \qquad z \to 0 \tag{22}$$

Heterosis appears when $H > 1$, and low F1 fitness when $H < 1$. To interpret eqs. 21-22, we need to understand how the vector magnitudes, r and r_{tot}, are likely to change over evolutionary time. Most importantly, in the initial stages of divergence, it is likely that $r_{tot} > r$, since r_{tot} includes contributions from divergence at all loci, while r includes only the uniparental loci; in this case, the model predicts heterosis. However, as genetic divergence increases, r will continue to grow (eqs. 18-19), but r_{tot} will remain small, because selection acts to keep both parental phenotypes close to the phenotypic optimum. As such, at higher levels of genetic divergence, it becomes likely that $r > r_{tot}$, and F1 fitness will tend to decline, as observed (Table S1; Edmands 2002).

2.4 Haldane's Rule The best-known pattern in speciation genetics is "Haldane's Rule": the tendency for sex-specific hybrid problems to affect the heterogametic sex (Haldane 1922; Coyne and Orr 2004, Ch. 8; Schilthuizen *et al.* 2011). Barton (2001) showed that this pattern could be generated by Fisher's model, but argued that this happens only if all derived mutations, differentiating the parental species, have complete phenotypic dominance (Barton 2001). As such, Haldane's Rule would appear only thanks to the action of "Haldane's sieve" (the preferential fixation of beneficial mutations that are dominant). However, Haldane's sieve cannot apply to mutations that fix via drift (as in Figs. 2a or 3a) and may not apply to many cases of adaptive evolution either (e.g. Orr and Betancourt 2001; Gerstein *et al.* 2014). Here we show that Fisher's model predicts Haldane's Rule under a much wider range of plausible conditions. These conditions are summarised in Table 2 and illustrated in Figure S12; there, and below, we refer to an XO system for simplicity, but all results apply also to XY or ZW systems.

Species with heteromorphic chromosomes tend to show similar levels of high fitness for pure-species individuals of both sexes, despite their different complement of X-linked alleles. Complete phenotypic dominance could explain this, since it predicts identical phenotypes for homozygous or hemizygous genotypes (Barton 2001). However, the same results can be obtained by dosage compensation (Mank *et al.* 2011). Dosage compensation can be included in Fisher's framework by modelling changes in expression as changes in the length of the phenotypic vectors. For example, if hemizygous X-linked alleles are upregulated (Mank *et al.* 2011; Deng *et al.* 2011) - modelled by a doubling of the phenotypic vector associated with these alleles - then this leads to (i) identical phenotypes

for pure species males and females, (ii) high fitness (heterosis) for homogametic F1 offspring, and (iii) a potential loss of fitness for heterogametic hybrids - all consistent with Haldane's Rule (Figure S12a). The same result holds if both copies of derived X-linked alleles are downregulated in the homogametic sex (Mank *et al.* 2011; Deng *et al.* 2011; Table 2, Fig. S12b). By contrast, Haldane's Rule is not predicted when dosage compensation involves silencing X-linked alleles from the XO parent (Mank *et al.* 2011; Table 2, Fig. S12c); in this case, all genotypes are functionally heterogametic, and so hybrids of both sexes are predicted to be equally unfit.

These results apply to traits that are expressed equally in both sexes. However, results are complicated when we consider traits that are sex-specific, such that dosage compensation is not required (Table 2; Fig. S12). This seems to apply to many components of fertility that are implicated in reproductive isolation, such as failures of spermatogenesis (Wu and Davis 1993; Coyne and Orr 2004, Ch. 8). In this case, Haldane's Rule is predicted for traits expressed solely in heterogametic offspring (Table 2). However, violations of Haldane's Rule are also predicted, but only for loci with early X silencing, which code for traits expressed solely in XX offspring (Table 2; Fig. S12c-d). These necessary conditions do apply plausibly to most known violations of Haldane's Rule (Sawamura 1996; Watson and Demuth 2012; Moran *et al.* under review), including the best-studied case in *Xenopus* (Malone *et al.* 2007).

Finally, we note that Fisher's model also predicts the severe loss of fitness that is observed in "unbalanced" hybrid genotypes beyond the F1. For example, low fitness is predicted with classic unbalanced homogametic hybrids (which contain two copies of the X from one parental species in an otherwise F1 background; Coyne and Orr 2004, Ch. 8; Fig. S13). The lowest possible fitness is predicted to apply to crosses that are double homozygotes (i.e., where both alleles at each locus come from one or other parental species; Fig. S13); and again, this prediction has some support in a range of systems (see Table S1 for details).

Discussion

Experimentalists have generated a very large body of data on intrinstic reproductive isolation (Table S1), and fitness landscape models can be useful in interpreting these data. Following Barton (2001) and Chevin *et al.* (2014), we have shown that all of the major patterns in the empirical literature (Table 1) can be generated by a simple form of Fisher's geometric model, which is analogous to optimizing selection acting on multiple quantitative traits. In particular, we have shown that such models can generate "genetic incompatibilities" or "speciation genes", i.e. alleles that confer very low fitness in foreign genetic backgrounds, but high fitness in their original backgrounds (sections 1.1-1.4). As such, Fisher's model is a legitimate way of modelling the "Dobzhansky-Muller mechanism",

where strong negative epistasis appears only in hybrids (Coyne and Orr 2004; Gavrilets 2004). However, this success requires a modification of the simplest model, namely, an increase in the curvature of the fitness function, k, to allow for strongly negative fitness interactions (Fig. 1). This explains why Barton (2001) and Chevin *et al.* (2014) did not observe classic speciation genes, since they considered models that are roughly equivalent to our model with $k = 2$.

We have also shown that the model can successfully account for the major patterns observed in hybridisation studies (sections 2.1-2.4), but again, this requires modifications, namely the inclusion of parent-of-origin effects (to generate parental sex asymmetry, but also yielding optimal outbreeding in the F1), and certain forms of dosage compensation (to generate Haldane's Rule for traits expressed by both sexes).

Differences from incompatibility-based models Most previous modelling work on speciation genetics has assigned fitnesses to genotypes in a quite different way: counting all possible combinations of two or more alleles, and assuming that each combination has a fixed probability of being incompatible, and thus conferring low fitness (Muller 1942; Orr 1995; Turelli and Orr 1995; Coyne and Orr 2004; Gavrilets 2004; Welch 2004; Turelli and Moyle 2007; Bank *et al*. 2012; Fraïsse *et al*. 2014). The predictions of Fisher's model and these incompatibility-based models differ in some important ways.

One difference concerns complex incompatibilities - i.e. epistatic fitness interactions between introgressed factors (section 1.3). Incompatibility-based models (Muller 1942; Orr 1995; Fraïsse *et al.* 2014), make no predictions about how common complex incompatibilities will be, since this enters as a free parameter, and yet they always predict that they will be difficult to detect. This is because the model assumptions imply that the introgression of a single factor alone is far more likely to reveal an incompatibility, than is the introgression of multiple factors together, providing that the introgressed regions are small (Welch 2004); this follows because a single introgression is involved in far more novel combinations of alleles, and so many more potential incompatibilities. This prediction does seem to contradict the data, because complex incompatiblities are often detected by introgressing multiple factors (Orr 1995; Coyne and Orr 2004; Fraïsse *et al.* 2014). By contrast, Fisher's geometric model predicts that multiple introgressions will often uncover complex incompatibilities, and particularly when recipient genotypes are well adapted (section 1.3). The difference arises because Fisher's model assigns fitnesses to complete genotypes, rather than to their component combinations of alleles; indeed, holey landscape models (Gavrilets and Gravner 1997; Gavrilets 2004), which also assign fitnesses to complete genotypes, make similar predictions to Fisher's model (Welch 2004).

A second, related difference concerns the rate at which incompatibilities (i.e. introgressions with large delete-

rious effects) accumulate with the genetic divergence between the parental species. When fitnesses are assigned by counting combinations of alleles, the number of incompatibilities grows very rapidly with divergence, as a direct result of combinatorial explosion (Orr 1995; Welch 2004). No such effect is predicted by Fisher's model, where fitnesses are assigned based on trait values. This means that Fisher's model does not predict a "snowball effect" in the number of incompatibilities (Orr 1995; Section 1.4). Nevertheless, we have shown that Fisher's model can generate an "apparent snowball effect" when the introgressions themselves contain a larger number of divergent sites (section 1.4; Figure 5). This might explain the conflicting empirical results about the snowball effect in nature, since the snowball effect has not been observed for single amino acids, but has been observed for larger introgressions (Presgraves 2010b; Table S1).

A third difference, discussed by previous authors (Gavrilets 2004, Ch. 6; Turelli and Moyle 2007), is that a gradual decline in F1 fitness (section 2.3) is difficult to explain using models where the number of potential incompatibilities grows explosively. Incompatibility-based models generate a gradual decline only if there is strong positive epistasis between incompatibilities (Gavrilets 2004; Welch 2004; Turelli and Moyle 2007; Gourbière and Mallet 2010), or if intermediate levels of mean hybrid fitness result from genetically variable populations in which individual hybrids have either very high or very low fitness (Gavrilets 2004, Ch. 5). However, there is little evidence of positive epistasis between strongly deleterious mutations (Charlesworth *et al*. 2004; Turelli and Moyle 2007), and widespread evidence that individual F1 hybrids have intermediate fitness levels (Table S1).

A final difference concerns asymmetry. Regarding single introgressions, incompatibility-based models make a stronger prediction, because, under Dobzhansky's model of two biallelic loci, incompatibilities must be asymmetric for divergence to take place via single substitutions (Dobzhansky 1937; Orr 1995). However, this constraint does not apply to incompatibilities involving three or more loci (Fraïsse *et al.* 2014), and nor to any number of loci with multiple alleles (Nei *et al.* 1983). Fisher's model, by contrast, includes a symmetric fitness function (Fig. 1), and so cannot generate asymmetry when introgressions are made into backgrounds that are phenotypically optimal (Figure S6). Nevertheless, we have shown that strong asymmetry is generated in high fitness, but suboptimal backgrounds (e.g. Fig. 3; middle row). Note also that, with Fisher's model, the asymmetry in the fitness of the F1 (section 2.2) appears quite independently of the symmetry or asymmetry of individual introgressions, and so there is no requirement that nuclear-nuclear incompatibilities "line up in the same direction" (Turelli and Moyle 2007).

Similarities to incompatibility-based models Despite these differences, several predictions are shared by both modelling approaches. We can be more confident that these shared predictions are not due to the simplifying assumptions of any one approach.

One important area of agreement concerns Haldane's Rule, where Fisher's model reproduces all of the qualitative predictions of the "dominance theory" (section 2.4; Muller 1942; Turelli and Orr 1995; Barton 2001; Coyne and Orr 2004, Ch. 8; Watson and Demuth 2012). In a sense, this tells us only that optimizing selection produces the dominance relations posited by the dominance theory (e.g. Manna *et al.* 2011) - just as it produces the dominance relations needed to generate heterosis in the F1 (section 2.3). However, modelling Haldane's Rule without identifying individual incompatibilities, suggests alternative and complementary ways of understanding the data. In Figure S12, for example, the low fitness of the heterogametic hybrids is most naturally explained by their lack of X-linked derived alleles that complement the autosomal alleles in their homospecific background; this modelling approach also places more emphasis on the requirement that conspecific males and females have similarly high fitness (Barton 2001; Watson and Demuth 2012).

The second area of agreement concerns the lack of connection between the mode of divergence of populations, and the quality of the reproductive isolation that emerges (section 1.1; Orr 1995; Gavrilets 2004; Chevin *et al.* 2014; Fraïsse *et al.* 2014). All of the patterns listed in Table 1 can arise as a passive consequence of genetic divergence, however this divergence took place. Indeed, we have argued that previous attempts to make inferences about divergence from patterns in reproductive isolation may have been premature. In particular, the lack of a "snowball effect" in the accumulation of incompatibilities involving single amino acids (Orr 1995; Kondrashov *et al.* 2002; Table S1), need not imply that all amino acid substitutions occur in epistatically interacting pairs (section 1.4; Fig. 5; Kondrashov *et al.* 2002; Gavrilets 2004; Welch 2004; Presgraves 2010b). Similarly, Haldane's Rule need not imply the action of Haldane's sieve - the preferential fixation of dominant mutations (section 2.4; Fig. S12; Barton 2001; Orr and Betancourt 2001; Gerstein *et al.* 2014). This prediction implies that the data summarised in Table 1 cannot be used to draw strong inferences about the process of speciation, or - equivalently - that the near ubiquity of these patterns is quite consistent with speciation taking place in a great variety of ways (Mayr 2001).

Comparison with other uses of Fisher's model

We have shown that the basic phenomena of speciation genetics can be generated by Fisher's geometric model, but in several cases, this requires that $k > 2$ (Table 1). An attractive feature of Fisher's geometric model is that this assumption can be "cross-checked" with other sources of data, unconnected to speciation genetics. This is an issue for our analysis, because the most successful quantitative predictions with Fisher's model have used $k = 2$ (e.g. Martin and Lenormand 2006a,b; Martin *et al.* 2007; Manna *et al.* 2011), and some of these analyses are evidence against higher k values. Perhaps the strongest evidence comes from observations that the mean strength of selection against new deleterious mutations, $E[s]$, is relatively unaffected by changing the environment in which

the effects are measured (Martin and Lenormand 2006b). The relevant data come from several species, including model organisms for speciation genetics (see Table 2 of Martin and Lenormand 2006b; see also Bank *et al.* 2014), and they support $k \approx 2$ because, for several variants of Fisher's model, it is straightforward to show that $E[s]$ is independent of z (the distance from the optimum) only when $k = 2$ (Martin and Lenormand 2006b; see also Appendix sections 1.1 and 3).

These apparently conflicting results might be reconciled in two non-exclusive ways. First, it might be the case that results from the speciation-genetic literature concern fixations at a small subset of loci, characterised by a high-k landscape, while the majority of spontaneous mutations are better described by $k \approx 2$. Second, the fitness landscape might take some more complicated form, with varying levels of curvature (see, e.g. Peck *et al.* 2012). More complex landscapes might also be generated by adding incompatibilities to a trait-based model with $k = 2$ (Manna *et al.* 2011; Chevin *et al.* 2014). This approach has been proposed by Manna *et al.* (2011) to account for data on the dominance coefficients of new mutations, because strong recessivity, like strong negative epistasis, does not otherwise appear when $k = 2$. Whether such a model can also account for the speciation genetic data, will depend on whether it inherits the difficulties of existing incompatibility-based approaches, as described above. And this will depend on details of the model that need further exploration (see also Appendix section 7). One promising direction might be to consider fitness functions that are asymmetrical (Williams 1992, Ch. 5; Urban *et al.* 2013), since much existing work on Fisher's model has relaxed the assumption of isotropy, while retaining the assumption of symmetry on some system of coordinates.

Acknowledgements

This work was funded by the Agence Nationale de la Recherche (HYSEA project, ANR-12-BSV7- 0011). We are very grateful to Lindell Bromham, Suzanne Edmands, Tatiana Giraud, Sébastien Gourbière, Guillaume Martin, Peter Moran, Ole Seehausen, Rike Stelkens, Frédérique Viard and David Waxman for their assistance, and to Ricardo Azevedo, Thomas Lenormand and two anonymous reviewers, whose detailed comments greatly improved the manuscript.

References

- Bank, C., R. Bürger, and J. Hermisson. 2012. The Limits to Parapatric Speciation: Dobzhansky–Muller Incompatibilities in a Continent–Island Model. *Genetics* **191**:845–863.

- Bank, C., R. T. Hietpas, A. Wong, D. N. Bolon, and J. D. Jensen. 2014. A Bayesian MCMC approach to assess the complete distribution of fitness effects of new mutations: uncovering the potential for adaptive walks in challenging environments. *Genetics* **196**:841-852.

- Barton, N. H. 2001. The role of hybridization in evolution. *Mol. Ecol.* **10**:551–568.

- Bolnick, D. I., and T. J. Near. 2005. Tempo of Hybrid Inviability in Centrarchid Fishes (Teleostei: Centrarchidae). *Evolution* **59**:1754–1767.

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

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

- Burton, R. S., and F. S. Barreto. 2012. A disproportionate role for mtDNA in Dobzhansky–Muller incompatibilities? *Mol. Ecol.* **21**:4942–4957.

- Charlesworth, B., H. Borthwick, C. Bartolomé, and P. Pignatelli. 2004. Estimates of the genomic mutation rate for detrimental alleles in *Drosophila melanogaster*. *Genetics* **167**:815–826.

- Chevin, L.-M., G. Martin, and T. Lenormand. 2010. Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evolution* **64**:3213–3231.

- Chevin, L.-M., G. Decorzent, and T. Lenormand. 2014. Niche Dimensionality and the Genetics of Ecological Speciation. *Evolution* **68**:1244–1256.

- Connallon, T., and A. G. Clark. 2014. Balancing selection in species with separate sexes: insights from Fisher's geometric model. *Genetics* **197**:991–1006.

- Corbett-Detig, R. B., J. Zhou, A. J. Clark, D. L. Hartl, and J. F. Ayroles. 2013. Genetic incompatibilities are widespread within species. *Nature* **504**:135–137.

- Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Sunderland, M.A.

- Darwin, C. 1859. On The Origin of Species by Means of Natural Selection. John Murray, London, U.K.

- Davis, A.W., and C.-I. Wu. 1996. The broom of the sorcerer's apprentice: the fine structure of a chromosomal region causing reproductive isolation between two sibling species of Drosophila. *Genetics* **143**:1287–1298.
- Deng, X., J. B. Hiatt, D. K. Nguyen, S. Ercan, D. Sturgill, L. W. Hillier *et al.* 2011. Evidence for compensatory upregulation of expressed X-linked genes in mammals, *Caenorhabditis elegans* and *Drosophila melanogaster*. *Nat. Genet.* **43**:1179–1185.
- Dobzhansky, T.G. 1937. Genetics and the Origin of Species. Columbia Univ. Press, New York, N.Y.
- Draghi, J., and M. C. Whitlock. 2015. Overdominance interacts with linkage to determine the rate of adaptation to a new optimum. *J. Evol. Biol*. **28**:95–104.
- Edmands, S. 2002. Does parental divergence predict reproductive compatibility? *Trends Ecol. Evol.* **17**:520–527.
- Escobar, J. S., A. Nicot, and, P. David. 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. Clarendon Press, Oxford, U.K.
- Fraïsse, C., J. A. D. Elderfield, and J. J. Welch. 2014. The genetics of speciation: are complex incompatibilities easier to evolve? *J. Evol. Biol.* **27**:688–699.
- Gavrilets, S. 2004. Fitness landscapes and the Origin of Species. Princeton Univ. Press, Princeton, N.J.
- Gavrilets, S., and J. Gravner. 1997. Percolation on the fitness hyper-cube and the evolution of reproductive isolation. *J. Theor. Biol.* **184**:51–64.
- Gerstein, A. C., A. Kuzmin, and S. P. Otto. 2014. Loss-of-heterozygosity facilitates passage through Haldane's sieve for *Saccharomyves cerevisiae*. *Nat. Commun.* **5**:3819.
- Gissi, C., F. Iannelli, and G. Pesole. 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity (Edinb).* **101**:301–320.
- Gourbière, S., and J. Mallet. 2010. Are species real? The shape of the species boundary with exponential failure, reinforcement, and the "missing snowball". *Evolution* **64**:1–24.
- Greiner, S., U. W. E. Rauwolf, J. Meurer, and R. G. Herrmann. 2011. The role of plastids in plant speciation. *Mol. Ecol.* **20**:671–691.

- Gros, P.-A., H. L. Nagard, and O. Tenaillon. 2009. The Evolution of Epistasis and Its Links With Genetic Robustness, Complexity and Drift in a Phenotypic Model of Adaptation. *Genetics* **182**:277–293.

- Haldane, J.B.S. 1922. Sex-ratio and unisexual sterility in hybrid animals. *J. Genetics* **12**:101–109.Johnson, T., and N. Barton. 2005. Theoretical models of selection and mutation on quantitative traits. Philos. *Trans. R Soc. Lond. B Biol. Sci.* **360**:1411–1425.

- Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge Univ. Press, Cambridge, U.K.

- Kondrashov, A. S., S. Sunyaev, and F. A. Kondrashov. 2002. Dobzhansky–Muller incompatibilities in protein evolution. *Proc. Natl. Acad. Sci. U.S.A.* **99**:14878–14883.

- Lima, T. G. 2014. Higher levels of sex chromosome heteromorphism are associated with markedly stronger reproductive isolation. *Nature commun.* **5**:4743.

- Lourenço, J., N. Galtier, and S. Glémin. 2011. Complexity, pleiotropy, and the fitness effect of mutations. *Evolution* **65**:1559–1571.

- Malone, J. H., T. H. Chrzanowski, and P. Michalak. 2007. Sterility and gene expression in hybrid males of *Xenopus laevis* and *X. muelleri*. *PloS One* **2**:e781.

- Mani, G. S., and B. C. Clarke. 1990. Mutational order: a major stochastic process in evolution. *Proc. R. Soc. Lond. B* **240**:29–37.

- Mank, J. E., D. J. Hosken, and N. Wedell. 2011. Some inconvenient truths about sex chromosome dosage compensation and the potential role of sexual conflict. *Evolution* **65**:2133–2144.

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

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

- Martin, G., S. F. Elena, and T. Lenormand. 2007. Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nat. Genet.* **39**:555–60.

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

- Martin, G., and T. Lenormand. 2006b. The fitness effect of mutations across environments: a survey in light of fitness landscape models. *Evolution* **60**:2413–2427.Maside, X. R., and H. F. Naveira. 1996. On the difficulties of discriminating between major and minor hybrid male sterility factors in Drosophila by examining the segregation ratio of sterile and fertile sons in backcrossing experiments. *Heredity* **77**:433–438.
- Matute, D. R., I. A. Butler, D. A. Turissini, and J.A. Coyne. 2010. A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science* **329**:1518–1521.
- Matuszewski, S., J. Hermisson, and M. Kopp. 2014. Fisher's geometric model with a moving optimum. *Evolution* **68**:2571–2588.
- Mayr, E. 2001. Wu's genic view of speciation. *J. Evol. Biol.* **14**:866–867.
- Moran, P., M. G. Ritchie, and N. W. Bailey. A rare exception to Haldane's Rule: Are X chromosomes key to hybrid incompatibilities? Under review.
- Muller, H. 1942. Isolating mechanisms, evolution and temperature. *Biol. Symp.* **6**:71–125.
- Nei, M., T. Maruyama, and C.- I. Wu. 1983. Models of evolution of reproductive isolation. *Genetics* **103**:557–579.
- Nolan, J. R., U. Bergthorsso, and C. M. Adema. *Physella acuta*: atypical mitochondrial gene order among panpulmonates (Gastropoda). *J. Molluscan Stud.* **80**:388–389.
- Ohlsson, R. 2013. Genomic imprinting: an interdisciplinary approach (Vol. 25). Springer, Berlin, DE.
- Orr, H. A. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* **139**:1805–1813.
- Orr, H. A. 1998. The Population Genetics of Adaptation: The Distribution of Factors Fixed during Adaptive Evolution. *Evolution* **52**:935–949.
- Orr, H. A. 2000. Adaptation and the cost of complexity. *Evolution* **54**:13–20.
- Orr, H. A., and A. J. Betancourt. 2001. Haldane's sieve and adaptation from the standing genetic variation. *Genetics* **157**:875–884.
- Peck, J. R., G. Barreau, and S. Heath. 1997. Imperfect genes, Fisherian mutation and the evolution of sex. *Genetics* **145**:1171–99.

- Peck, J. R., D. Waxman, and J. J. Welch. 2012. Hidden epistatic interactions can favour the evolution of sex and recombination. *PLoS One* **7**:e48382.
- Poon, A., and S. P. Otto. 2000. Compensating for our load of mutations: Freezing the mutational meltdown. *Evolution* **54**:1467–1479.
- Presgraves, D. C. 2010a. The molecular evolutionary basis of species formation. *Nat. Rev. Genet.* **11**:175–180.
- Presgraves, D. C. 2010b. Speciation genetics: Search for the missing snowball. *Curr. Biol.* **20**:R1073–R1074.
- Roze, D., and A. Blanckaert. 2014. Epistasis, pleiotropy, and the mutation load in sexual and asexual populations. *Evolution* **68**:137–149.
- Sato, A., S. M. Shimeld, and J. D. Bishop. 2014. Symmetrical reproductive compatibility of two species in the *Ciona intestinalis* (Ascidiacea) species complex, a model for marine genomics and developmental biology. *Zoolog. Sci.* **31**:369–374.
- Sawamura, K. 1996. Maternal Effect as a Cause of Exceptions for Haldane's Rule. *Genetics* **143**:609–611.
- Schilthuizen, M., C. W. G. Giesbers, and L. W. Beukeboom. 2011. Haldane's rule in the 21st century. *Heredity* **107**:95–102.
- Sella, G., D. A. Petrov, M. Przeworski, and P. Andolfatto. 2009. Pervasive natural selection in the Drosophila genome. *PLoS Genet.* **5**:e1000495.
- Sellis, D., B. J. Callahan, D.A. Petrov, and P.W. Messer. 2011. Heterozygote advantage as a natural consequence of adaptation in diploids. *Proc. Natl. Acad. Sci. USA* **108**:20666–20671.
- Tenaillon, O., O. K. Silander, J.-P. Uzan, and L. Chao. 2007. Quantifying Organismal Complexity using a Population Genetic Approach. *PLoS One* **2**:e217.
- Turelli, M. 1985. Effects of pleiotropy on predictions concerning mutation-selection balance for polygenic traits. *Genetics* **111**:165–195.
- Turelli, M., and L. C. Moyle. 2007. Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics* **176**:1059–1088.
- Turelli, M., and H. A. Orr. 1995. The dominance theory of Haldane's rule. *Genetics* **140**:389–402.

- Urban, M. C. , R. Bürger, and D. I. Bolnick (2013). Asymmetric selection and the evolution of extraordinary defences. *Nat. Comms.* **4**:2085.
- Vrana, P. B. 2007. Genomic imprinting as a mechanism of reproductive isolation in mammals. *J. Mammal.* **88**:5–23.
- Watson, E. T., and J. P. Demuth. 2012. Haldane's Rule in marsupials: What happens when both sexes are functionally hemizygous? *J. Hered.* **103**:453–458.
- Waxman, D., and J. J. Welch. 2005. Fisher's Microscope and Haldane's Ellipse. *Am. Nat.* **166**:447–457.
- Williams, G. C. 1992. Natural selection: Domains, levels and challenges. Oxford Univ. Press, Oxford, U.K.
- Welch, J. J. 2004. Accumulating Dobzhansky-Muller incompatibilities: reconciling theory and data. *Evolution* **58**:1145–1156.
- Welch, J. J., and D. Waxman. 2003. Modularity and the cost of complexity. *Evolution* **57**:1723–1734.
- Wu, C., and A.W. Davis. 1993. Evolution of postmating reproductive isolation: the composite nature of Haldane's rule and its genetic basis. *Am. Nat.* **142**:187–212.
- Zhang, X.-S., and W. G. Hill. 2003. Multivariate Stabilizing Selection and Pleiotropy in the Maintenance of Quantitative Genetic Variation. *Evolution* **57**:1761–1775.

Figure Captions

Figure 1

The fitness function used in this work, illustrated with $n = 1$ trait, and curves, from light to dark, showing $k = 2, 4, 6$. Higher values of k lead to a function that is increasingly "table-like", with a plateau of high fitness values around the optimum ($z = 0$) and a precipitous drop at extreme trait values.

Figure 2

The probability of fixation of speciation genes under Fisher's geometric model (see Appendix eqs. 31 and 36). Results are given as a function of the ratio s_i/s_d where s_i is the deleterious effect that would be caused by the mutation were it to be introgressed into an optimal heterospecific background, and s_d is the maladaptation of the diverging parental population when the mutation originally reached fixation. $s_i/s_d \gg 1$ implies that reproductive isolation has evolved without either population needing to "cross a fitness valley". Panel (a) shows results when the diverging populations were sufficiently well adapted for positive selection to be ineffective, such that all mutations fix via genetic drift (eqs. 31-34), and in this case results are compared to the fixation probability of a truly neutral mutation ($P_{neut} = 1/N$). Panel (b) shows results when between-population divergence takes place via positive selection (eq. 36), and the fixation probabilities are compared to that of the best possible mutation ($P_{best} \approx 2(N_e/N)s_d$). In both panels, colours light-to-dark indicate results for $k = 2, 4, 6$, as in Figure 1; and line types indicate results with $n = 5$ (solid lines) and $n = 10$ (dashed lines).

Figure 3

Illustrative simulations of the evolution of speciation genes under Fisher's geometric model, with $k = 6$ and $n = 5$. The upper panels compare the selection coefficients of mutations that reach fixation during population divergence (s), to their negative fitness effects when introgressed into the heterospecific genetic background (s_i). The middle panels compare the effects of reciprocal introgressions, that is, swapping orthologous alleles (ancestral and derived) between a species pair. Points that depart from the 1:1 diagonal show a high degree of asymmetry. The lower panels compare the effects of introgressing single alleles ($s_{i,single}$), to the effects of introgressing both alleles together ($s_{i,double}$), for 1000 randomly chosen pairs of divergent alleles. The diagonals show the expectations for introgressions into an optimal background, when the introgressed mutations have the same magnitude ($2^{k/2}$:1), and when their magnitudes are very different (1:1; see eq. 8). Each column presents results from a single simulation run. For column (a), populations diverged via drift and mutation pressure in an unchanging environment, and plots show

results after $N\mu t = 10^7$ mutations had appeared in each population, of which ~6000 reached fixation. For column (b) divergence tooks place via positive selection, as each population adapted to cyclical change in the optimal value of one of the n traits (eq. 54); plots show results after 100 cycles of change, during each of which $N\mu t_{cycle} = 10^4$ mutations appeared, of which ~500 fixed in each population. In both cases, mutations were generated from eq. 51, with $\sigma = 0.5$ and $\bar{r} = 0.1$, such that deleterious mutations in an optimal background reduced fitness by ~1% on average (eq. 52).

Figure 4

Patterns of asymmetry in the fitness effects of reciprocal introgressions under Fisher's geometric model. Results are plotted as a function of r, which is defined as the magnitude of the phenotypic effect associated with the introgressed factor. Results also apply to the asymmetry of male-female versus female-male F1 hybrids, if we treat r as the magnitude of the difference in phenotypic effects coded at "uniparentally inherited" loci (see text and eq. 15). For each value of r, 500 pairs of fitnesses, representing reciprocal introgressions or crosses, were simulated by generating a random direction in phenotypic space, using results shown in eqs. 24 and 37. Panels (a) and (c) show the mean (black line) and 95% quantiles (grey lines) of overall asymmetry (eq. 4), and this is compared to the analytical approximation of eq. 5 (dotted line). Panels (b) and (d) show the corresponding fitness values of the pair of genotypes. The two black lines show results for the simulation replicate whose level of asymmetry was closest to the mean value across replicates, while the grey lines show the replicate whose level of asymmetry was closest to the upper 95% quantile. Also shown, (dotted line) is the analytical approximation for the mean log fitness (eq. 42), which is indistinguishable from the simulated values averaged across directions and replicates (not shown). Simulations use $n = 16$ and results are compared with $k = 6$ (panels (a)-(b)) and $k = 2$ (panels (c)-(d)), confirming that stronger asymmetry appears with higher values of $k/\sqrt{n}$. Other parameters were set at $z = 0.8$, and $\sigma = 0.5$. The vertical dotted lines show the region of maximum expected asymmetry ($r = z$).

Figure 5

Simulations of the accumulation of genetic incompatibilities (between-species introgressions with strong deleterious effects) under Fisher's geometric model. The x-axis shows the total number of substitutions that had accrued between the diverging populations, and the y-axis shows the proportion of introgressions that were strongly deleterious. Panel (a) shows results from simulating introgressions that comprised single divergent sites, thus replicating bioinformatic tests of the snowball effect. In the remaining panels, substitutions were assigned at random to one of (b) 5000 or (c) 1000 genomic segments, and these segments were then introgressed as a whole, thus replicating

direct experimental tests. The same simulation runs were used for all three panels, and they used the same parameter values as Figure 3a ($k = 6$, $n = 5$ and $N = 500$). Curves show averaged results across 50 replicate simulations, and points (black and grey) show results from the single illustrative run, also shown in Fig. 3a, with introgressions in opposite directions. Incompatibilities were defined as introgressions that reduced fitness by at least 50% (red curves), 25% (green curves and points), or 15% (blue curves).

Table 1: Patterns in speciation genetic data

	Pattern		Generated by Fisher's model?	Section
Introgression studies:	Speciation genes	Large loss of fitness can be caused by the introgression of small chromosomal regions	Yes, if $k > 2$	1.1
	Introgression asymmetry	Reciprocal introgressions have very different effects on fitness	Yes, if $k > \sqrt{n}$	1.2
	Complex epistasis	Effects of single introgressions can be strongly affected by presence of additional introgressions	Yes, if $k > 2$	1.3
	Snowball effect	The probability that an introgression is strongly deleterious increases with genetic distance between the backgrounds	No, but "apparent snowball effect" appears with multi-site introgressions	1.4
Overall hybrid fitness:	F2 speciation clock	A steady decline in the mean fitness of recombinant hybrids with parental genetic distance	Yes	2.1
	F1 parental sex asymmetry	F1 fitness depends strongly on which species takes the female role during hybridization	Yes, with "parent-of-origin" effects, and $k > \sqrt{n}$	2.2
	Heterosis/hybrid breakdown	F1 hybrids at low genetic distances can be fitter than their parents, with low fitness appearing in the F2 or later.	Yes	2.3
	Optimal outbreeding, and F1 speciation clock	F1 hybrids become increasingly unfit with increasing parental genetic distance	Yes, with "parent-of-origin" effects	2.3
	Haldane's Rule	If hybrids of one sex only are unfit, it will be the heterogametic sex	Yes, with some forms of dosage compensation, or sex-specific traits	2.4

Table 2: Predicted occurrences of Haldane's Rule under Fisher's Geometric Model

	Traits expressed in:		
Dosage compensation:	Both sexes	XO	XX
None	na	HR	–
X down-regulation in XX individuals	HR	HR	–
X up-regulation in XO individuals	HR	HR	–
X silenced from XO parent	–	HR	V
X silenced from XX parent	na	–	V

Note: HR: Haldane's Rule predicted; V: Violation of Haldane's Rule predicted; na: highly unequal fitness for pure species males and females. Note that low fitness for hybrids of both sexes, or of neither sex, can occur under all scenarios, so that Haldane's Rule may be reciprocal (applying to crosses in both directions) or non-reciprocal. Predictions are shown for an XO system, but apply equally to XY and ZW systems.

Online Appendix for "The genetics of speciation: insights from Fisher's Geometric Model" by Fraïsse, Gunnarsson, Roze, Bierne and Welch.

1. Results relating to single introgressions

1.1 Distribution of selection coefficients Given the fitness function (eq. 1) if a population is an Euclidean distance $z \equiv \|\mathbf{z}-\mathbf{o}\| \equiv \sqrt{\sum_{i=1}^{n}(z_i-o_i)^2}$ from the optimum, and if a mutation has size $r \equiv \|\mathbf{r}\| \equiv \sqrt{\sum_{i=1}^{n} r_i^2}$ then the selection coefficient of this mutation, s (eqs. 1-2), can be defined via

$$\begin{aligned} u &\equiv \ln(1+s) = \ln\left(w(\mathbf{z}+\mathbf{r})\right) - \ln\left(w(\mathbf{z})\right) \\ &= \sigma\left(z^k - \left(\sqrt{z^2+r^2+2zr\cos\theta}\right)^k\right) \end{aligned} \tag{23}$$

where θ is the angle between the vectors $\mathbf{z}-\mathbf{o}$ and $\mathbf{r}$ (see, e.g., Welch and Waxman 2003). To derive the distribution of selection coefficients, we assume isotropic mutation, and then use the standard approximation for the pdf (probability density function) of the quantity $y \equiv -\sqrt{n}\cos\theta$, denoted $f(y)$

$$\begin{aligned} f(y) &= \frac{\Gamma(n/2)}{\sqrt{n\pi}\Gamma((n-1)/2)}\left(1-\frac{y^2}{n}\right)^{(n-3)/2}, \qquad y^2 < n \\ &\approx \sqrt{\frac{1}{2\pi}}e^{-y^2/2},\ n \gg 1 \end{aligned} \tag{24}$$

(Fisher 1930; Leigh 1987; Welch and Waxman 2003). We can then use a change of variables to derive the distribution of selection coefficients for mutations with size r:

$$\varphi(u|r) = \frac{\sqrt{n}}{k\sigma rz}\left(z^k - u/\sigma\right)^{2/k-1} f\left(\frac{\sqrt{n}}{2rz}\left(r^2+z^2-(z^k-u/\sigma)^{2/k}\right)\right) \tag{25}$$

This is bounded above at the largest possible beneficial mutation, which occurs when $u = \sigma z^k = \sigma r^k$ since $s \leq e^{\sigma z^k} - 1$, which occurs when $\mathbf{z} = -\mathbf{r}$, and the mutation takes the population directly to the optimum.

We note here that the mean of this distribution varies with z (the distance of the phenotype from the optimum), except in the special case of $k = 2$ (see also Waxman and Welch 2005; Martin and Lenormand 2006b). This explains

why the constancy of $E[s]$ across different enviroments, implying very different levels of population adaptation, are most consistent with $k \approx 2$ (Martin and Lenormand 2006b; see also Discussion).

1.2. Fixation probabilities For a new mutation with selection coefficient s, the fixation probability is approximately

$$P \approx \frac{1-e^{-\frac{2N_e s}{N}}}{1-e^{-2N_e s}} \approx \begin{cases} 2\frac{N_e}{N}s, & N_e s > 1 \\ \\ \frac{1}{N}, & N_e |s| \ll 1 \end{cases} \tag{26}$$

(Fisher 1930; Kimura 1983). Figure 2 plots this probability as a function of s_i (the effect of introgressing the mutation into an optimal background), s_d (the maladaptedness of the population before the mutation appeared), and the model parameters n and k. To do this, we first write s_i and s_d as functions of the other model parameters. Making a small s approximation, and using eq. 23, we find

$$-s_i \approx \ln(1-s_i) = \ln w(\mathbf{r}) = -\sigma r^k \tag{27}$$

$$-s_d \approx \ln(1-s_d) = \ln w(\mathbf{z}) = -\sigma z^k \tag{28}$$

from which eq. 3 follows directly. We can then rewrite u as

$$u = s_d \left(1 - \left(1 - \left(\frac{y}{\rho} - 1\right)(s_i/s_d)^{2/k}\right)^{k/2}\right) \tag{29}$$

where we have used Fisher's "characteristic length", ρ (Fisher 1930; Orr 1998).

$$\rho \equiv \frac{\sqrt{n}r}{2z} = \frac{\sqrt{n}}{2}(s_i/s_d)^{1/k} \tag{30}$$

Combining eqs. 24-30, we derive the result for fixations under drift.

$$P_{drift} \approx \frac{1}{N}\int_{-1/N_e}^{1/N_e} \varphi(u|r)du \tag{31}$$

$$\approx \frac{1}{N}\int_{h}^{\infty} f(x)dx \tag{32}$$

$$\approx \frac{1}{N}\frac{e^{-h^2/2}}{\sqrt{2\pi}h}, \qquad h > 1 \tag{33}$$

where

$$h = \frac{\sqrt{n}}{2}\frac{1+(s_i/s_d)^{2/k}-\left(1+(N_e s_d)^{-1}\right)^{2/k}}{(s_i/s_d)^{1/k}} \tag{34}$$

$$= \frac{\sqrt{n}}{2}\frac{1+(r/z)^2-\left(1+(N_e\sigma z^k)^{-1}\right)^{2/k}}{(r/z)} \tag{35}$$

Equation 31 is plotted in Figure 2a, while the further approximations of eq. 33 assume that $N_e s_d < 1$, such that all fixable mutations are under ineffective selection. Equations 33-34 show that the fixation probability will decline with n and k as discussed in the main text.

We use the same approach to derive the fixation probability under positive selection. Again, combining eqs. 25-30 we find

$$P_{seln} \approx \frac{2N_e}{N}\int_0^{\infty} u\varphi(u|r)du$$

$$= \frac{2N_e}{N}s_d\int_0^{(s_i/s_d)^{-2/k}} \left(1-\left(1-(s_i/s_d)^{2/k}x\right)^{k/2}\right)\rho f(\rho(1+x))dx \tag{36}$$

Equation 36, which is plotted in Figure 2b, generalises previous results (Kimura 1983; Orr 1998; Waxman and Welch 2005), and shows that, regardless of k, mutations very rarely reach fixation via positive selection when $\rho = \sqrt{n}(r/z)/2 > 3$. As such, large effect mutations, with $r > z$, will almost never fix when $\sqrt{n} > 6$. Together with eq. 28, this explains why speciation genes will never evolve with large values of n, unless mutations are introgressed into a heterospecific background that is phenotypically suboptimal (see, e.g., Figs. 2b, 3b and S6).

The dependency of the fixation probability on ρ also explains why populations tend to track a moving optimum less closely when n increases, and this leads, in the simulations, to the occasional fixation of more strongly beneficial mutations (Figs. 3b, S3-S4; see also Matuszewski *et al.* 2014).

1.3 Asymmetry in introgressions As in the main text, we assume that both parental backgrounds have the same suboptimal phenotype, with Euclidean distance z from the optimum (which allows us to apply the results directly to the overall fitness of the F1 cross). To calculate the asymmetry in the effects of reciprocal introgressions, we note that we reverse the direction of a vector by adding π radians to its angle, θ, and that $\cos(\theta+\pi) = -\cos(\theta)$. As such,

$$\begin{aligned} \ln w(\mathbf{z}+\mathbf{r}) &= -\sigma\left(z^2+r^2-\frac{2zry}{\sqrt{n}}\right)^{k/2} \\ \ln w(\mathbf{z}-\mathbf{r}) &= -\sigma\left(z^2+r^2+\frac{2zry}{\sqrt{n}}\right)^{k/2} \end{aligned} \tag{37}$$

Then, using the definition of A (eq. 4), we find

$$\begin{aligned} A &= \frac{\ln w(\mathbf{z}+\mathbf{r})}{\ln w(\mathbf{z}-\mathbf{r})} \\ &= \left(\frac{1+(r/z)^2-2(r/z)y/\sqrt{n}}{1+(r/z)^2+2(r/z)y/\sqrt{n}}\right)^{k/2} \end{aligned} \tag{38}$$

We now take logarithms and a first order series expansion in y to yield:

$$\ln A \approx y\frac{2k}{\sqrt{n}}\frac{(r/z)}{(1+(r/z)^2)} \tag{39}$$

From the distribution of y (eq. 24), $\ln A$ will be approximately normal with mean zero, and a standard deviation that increases with k. Results for the absolute value of this quantity, eqs. 5-6, follow from standard results for the half-normal distribution, namely $E[|y|] = \sqrt{2/\pi}$, and $\mathrm{Var}[|y|] = 1-2/\pi$.

Results for the F1 cross take the same approach. Using the notation of eqs. 12-17 and 20, the parental species have phenotypes $\mathbf{z}_{mp}+\mathbf{r}_{tot}$ and $\mathbf{z}_{mp}-\mathbf{r}_{tot}$, and the reciprocal F1s have phenotypes $\mathbf{z}_{mp}+\mathbf{r}_u$ and $\mathbf{z}_{mp}-\mathbf{r}_u$. We can

then derive the expected log fitness of F1 hybrids:

$$E\left[\ln w_{\mathrm{F1}}\right] = \frac{1}{2}E\left[\ln w(\mathbf{z}_{mp}+\mathbf{r}_u)+\ln w(\mathbf{z}_{mp}-\mathbf{r}_u)\right] \tag{40}$$

$$= E_y \left[-\frac{\sigma}{2}\left(\left(z^2+r^2-\frac{2zry}{\sqrt{n}}\right)^{k/2}+\left(z^2+r^2+\frac{2zry}{\sqrt{n}}\right)^{k/2}\right)\right] \tag{41}$$

$$= -\sigma\left(z^2+r^2\right)^{k/2}\left[1+\frac{k(k-2)}{2n}\frac{r^2z^2}{(r^2+z^2)^2}\right]+O\left(y^3\right) \tag{42}$$

Equations 21-22 use the same approach, neglecting terms in y^2 (the square brackets in eq. 42).

1.4 Complex epistasis To derive results for C - which quantifies the epistatic interaction between introgressions - we start from the definition of eq. 7 and use a small s approximation:

$$C \approx \frac{\ln w(\mathbf{z}+\mathbf{r}_1+\mathbf{r}_2)-\ln w(\mathbf{z})}{\ln w(\mathbf{z}+\mathbf{r}_1)-\ln w(\mathbf{z})}$$

$$= \frac{\|\mathbf{z}+\mathbf{r}_1+\mathbf{r}_2\|^k-z^k}{\|\mathbf{z}+\mathbf{r}_1\|^k-z^k} \tag{43}$$

where $\mathbf{z}$ is the recipient phenotype before the introgression, and $\mathbf{r}_1$ ($\mathbf{r}_2$) are the phenotypic effects of the more (less) deleterious introgression.

To complete the expression, we require the angles between the vectors $\mathbf{z}$ and $\mathbf{r}_1$, and between $\mathbf{z}+\mathbf{r}_1$ and $\mathbf{r}_2$. We can again use the normal approximation of eq. 24, but it follows from the geometry that the two random variables are correlated.

$$\mathrm{Cor}\left(y_{\mathbf{z},\mathbf{r}_1}, y_{\mathbf{z}+\mathbf{r}_1,\mathbf{r}_2}\right) \approx \sqrt{\frac{1}{2}} \tag{44}$$

Nevertheless, either variable can be expressed as the sum of uncorrelated normal variables, and this yields the

following approximation.

$$C \approx \frac{\left(z^2+r_1^2-\frac{2zr_1}{\sqrt{n}}y_0+r_2^2-2r_2\sqrt{\left(z^2+r_1^2-\frac{2zr_1}{\sqrt{n}}y_0\right)}\frac{\sqrt{\frac{1}{2}}y_0+\sqrt{\frac{1}{2}}y_1}{\sqrt{n}}\right)^{k/2}-z^k}{\left(z^2+r_1^2-\frac{2zr_1}{\sqrt{n}}y_0\right)^{k/2}-z^k} \tag{45}$$

where y_0 and y_1 are uncorrelated standard normals. Expanding in y_0 and y_1 then yields eqs. 8-9. We note that these approximations assume that n is sufficiently large to guarantee that all introgressions are deleterious, despite the maladaptation of the recipient genotype. We note further that the expansion means that results still apply approximately when mutations affect different subsets of the traits (relaxing the assumption of universal pleiotropy), since in that case, all vectors will be orthogonal. A fuller expression for $\mathrm{CV}[C]$ is easier to obtain when the recipient genotype is optimal ($z = 0$), and in this case we find

$$\mathrm{CV}[C] \approx \frac{k}{\sqrt{n}}\frac{(r_2/r_1)}{1+(r_2/r_1)^2}, \qquad z = 0. \tag{46}$$

2. Detailed simulation methods

2.1 Mutation model Given our simplifying assumptions that the effects of new mutations are translational and isotropic, there are two commonly used approaches to model variation in their sizes (e.g. Orr 2000; Poon and Otto 2000). First, we could assume that the mutational change in each trait is drawn independently from a normal distribution with a fixed variance, v, so that the mutated value of trait i is given by

$$z_i^{'} = z_i + \sqrt{v}\xi_i \tag{47}$$

where ξ_i is a random number drawn from a standard normal distribution. This simple model can be considered as a localised approximation of a more general mutation function, and is a special case of the approach used by Martin and Lenormand (2006a) and Chevin *et al.* (2014), who allowed for differences in mutational covariances (see Appendix section 3 below). However, unless n is very small, this model also has the property of suppressing mutations of very large and/or very small effects (Wingreen *et al.* 2003; Johnson and Barton 2005). This is because, as n increases, it becomes increasingly unlikely that all n random numbers will have very small or very large magnitudes. For this reason, we also follow Orr (1998), and constrain each mutation to have a known size, r.

$$z_i^{'} = z_i + r\frac{\xi_i}{\sqrt{\sum \xi_i^2}} \tag{48}$$

We can then vary r by treating it as a random number drawn from some distribution, $p(r)$. The multivariate normal approach (eq. 47) follows as special case when we set $p(r)$ to be a chi-distribution, with parameters n and v.

$$p(r;n,v) = \frac{2^{1-\frac{n}{2}}(r/\sqrt{v})^{n-1}e^{-r^2/2v}}{\Gamma(n/2)} \tag{49}$$

In this case, we have $E[r] \approx \sqrt{vn}$ and $\mathrm{Var}[r] \approx v/2$, but distributions with higher variance, and no dependence on n can also be chosen. To understand the effects of this choice, we can consider the distribution of deleterious mutations in an optimal genotype. As above, this is easiest to calculate using the approximation $s \approx u$, and in this case, since $u = -\sigma r^k$, we find

$$\varphi_{z=0}(u) = \frac{p\left((-u/\sigma)^{1/k}\right)}{k\sigma(-u/\sigma)^{1-1/k}} \tag{50}$$

For our main simulation results, we chose to keep this distribution constant for different values of k, as such we assumed that the quantity $r^* \equiv r^k$ was drawn from an exponential distribution with mean $\bar{r}$

$$p^*(r^*;\bar{r}) = \exp(-r^*/\bar{r})/\bar{r} \tag{51}$$

In this case, the mean and coefficient of variation of the selective effects of new mutations in an optimal background are

$$E[u] = -2\sigma\bar{r}^2 \tag{52}$$

$$\mathrm{CV}[u] = \frac{\sqrt{80}}{2} \approx 4.5 \tag{53}$$

For reported simulations, we chose $\bar{r} = 0.1$ and $\sigma = 0.5$, such that mutations had a mean deleterious effect of around 1% in an optimal background.

2.2 Environmental change To simulate divergence via positive selection (Figs. 3b, S3-S5, S8), we induced adaptive change by moving the optimum phenotype of one of the n traits in a cyclic fashion, while the remaining $n-1$ optima remained fixed at the origin. Specifically, the moving optimum was set at the origin at the beginning and end of each cycle (when the simulated introgressions took place), and in between, it moved to a new optimum

a unit distance away. Since log fitness varies according to z^k, where z is the distance from the optimum, we varied the pattern of environmental change so that the decline in fitness experienced by a non-evolving population would not depend on k. In particular, we used the following function to choose the optimal value of the trait.

$$o(t) \quad = \quad \left(\frac{1}{2}\left[1+\sin\left(2\pi t-\pi/2\right)\right]\right)^{2/k} \tag{54}$$

This function, which generalises the function used by Charlesworth (1993) and Zhang (2012), moves smoothly, so that populations never experience a substantial drop in fitness. In general, we moved the optima in identical ways in the two diverging populations (Table S2a), but in Figure S8a (lighter curves), we also explored two other scenarios, and these are described in Table S2b-c.

3. Simulations with anisotropic selection and mutation

Martin and Lenormand (2006a) introduced an important variant of Fisher's geometrical model, which relaxes the assumption of isotropy for both selection and mutation. Their model treats both selection and mutation as multivariate normal distributions, with arbitrary covariance matrices, **S** and **M**. Under this model, Martin and Lenormand (2006a) show that fitness can be written as

$$\ln w(\mathbf{z}) = -\sigma\sum_{i}^{n}\lambda_i z_i^2, \tag{55}$$

where λ_i is the i^{th} eigenvalue of the matrix dot product $\mathbf{S}\cdot\mathbf{M}$, and the z_i are the components of the phenotype in a transformed space. The n components of each new mutation are then generated from a standard unit normal, so that the selection coefficient of a mutation can be determined via

$$\ln(1+s) = \ln w(\mathbf{z}+\mathbf{r}) - \ln w(\mathbf{z}) = \sigma\left[\sum_{i}^{n}\lambda_i z_i^2 - \sum_{i}^{n}\lambda_i\left(r_i+z_i\right)^2\right], \tag{56}$$

where $r_i \sim N(0,1)$. Martin and Lenormand (2006b) introduce a natural extension of this model to higher k.

$$\ln w(\mathbf{z}) = -\sigma\sum_{i}^{n}\lambda_i\left|z_i\right|^k \tag{57}$$

Note that, even in the case of istropic mutation and selection (such that all λ_i are equal), this model is not identical to our eq. 1 (Tenaillon et al. 2007), except in the special case of $k=2$. Nevertheless, both models capture the same basic idea that fitness declines more rapidly with higher k. Under this model, for an optimal phenotype ($\mathbf{z}=\mathbf{0}$), we

have

$$E[s] \approx E[\ln(1+s)] = -\sigma 2^{k/2} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi}} \sum_{i}^{n} \lambda_i, \tag{58}$$

and so, using this result, the matrices can be scaled so as to fix the mean strength of selection at a given value. For our simulations, we generated **S** and **M** for each replicate as random Wishart matrices with n degrees of freedom (see Martin and Lenormand 2006a), and then used eq. 58 to scale the matrices such that $E[s] \approx 0.01$. We first ran simulations with $n = 5$, but correlations reduce "the effective number of traits", n_e (Orr 2000; Waxman and Welch 2005), as calculated from

$$n_e = \frac{n}{1 + CV^2(\lambda)} \tag{59}$$

(eq. 3 of Martin and Lenormand 2006a). Using Wishart matrices with n degrees of freedom implies that $n_e \approx n/3$ (Martin and Lenormand 2006a; see their eqs. 8 and A1), but this is a slight overestimate unless n is large. Accordingly, to match our isotropic simulations, which used $n = 5$ (e.g. Figure 3), we ran simulations both with $n = 5$ (yielding $n_e \approx 1.6$) and with $n = 16$ (yielding $n_e \approx 5$). The simulation results are shown in Figures S9 and S10. These plots show that the qualitative results of the simple isotropic model are also generated by the more general model, and in neither case do we observe introgressions with large fitness effects when $k = 2$. The sole complication is that asymmetric introgressions (albeit of very small effect) can appear with $k = 2$ and $n = 5$ (Fig. S9a middle row), perhaps because $k > \sqrt{n_e}$ holds, even though $k < \sqrt{n}$ (Table 1).

4. Results relating to recombinant hybrid fitness

Under Fisher's model, the log of recombinant hybrid fitness can be written as

$$\begin{aligned} \ln w_{\mathrm{F2}} &= -\sigma \left[\sum_{i}^{n} \left(\sum_{j}^{d} r_{ij} \right)^2 \right]^{k/2} \\ &\equiv -\sigma Z^{k/2} \end{aligned} \tag{60}$$

where d is the number of fixed mutations that differentiate the parental genotypes, and r_{ij} is the i^{th} component of the j^{th} fixed mutation. If we assume that the fixed mutations occur in random directions in phenotypic space, then the components can be treated as normal random variables, and the term in square brackets, Z, will follow a chi-squared distribution, with n degrees of freedom. Central-limit-type behaviour means that this approximation

works well, even if there are dependencies among mutation components, e.g. from fixing the total mutation length, as in eq. 48, or among overall mutation directions, e.g. resulting from environmental change (Waxman and Welch 2005).

Equation 10 follows directly from eq. 60, and the constant of proportionality depends on the size of the mutations fixed. This factor, like d, cancels from the coefficient of variation, which is found to be

$$\begin{aligned} \mathrm{CV}\left[\ln w_{\mathrm{F2}}\right] &\approx \frac{\sqrt{E\left[Z^k\right] - E^2\left[Z^{k/2}\right]}}{E\left[Z^{k/2}\right]} \\ &= \frac{\sqrt{\Gamma\left(k+\frac{n}{2}\right)\Gamma\left(\frac{n}{2}\right) - \Gamma^2\left(\frac{k}{2}+\frac{n}{2}\right)}}{\Gamma\left(\frac{k}{2}+\frac{n}{2}\right)} \\ &= \sqrt{\frac{2}{n}}, \quad k=2 \end{aligned} \tag{61}$$

(see Fig. S11). Equation 11 then follows as a further approximation. This prediction of a constant $\mathrm{CV}\left[\ln w_{\mathrm{F2}}\right]$, contrasts with incompatibility-type models, which predict a steady decrease in $\mathrm{CV}\left[\ln w_{\mathrm{F2}}\right]$ with parental divergence. To see this, consider a threshold-type model in which a proportion q of the hybrids are inviable or infertile and a proportion $(1-q)$ have optimal fitness (e.g. Gavrilets 2004, Ch. 6). In this case, we find

$$\mathrm{CV}\left[\ln w_{\mathrm{F2}}\right] = \sqrt{(1-q)/q} \tag{62}$$

which decreases with q, and therefore with d (Gavrilets 2004, Ch. 6). Testing these different predictions is difficult, however, since they require the fitnesses of many individual recombinant genotypes measured with high precision, and such data are - understandably - scarce (though see Xu and He 2011). Equations 18-19 also follow from eq. 60 using a similar approach.

5. Maladaptedness of the parental populations

Many results above depend on the maladaptedness of the pure species populations, before an introgression or cross. This quantity affects the probability of observing speciation genes (main text section 1.1; Figure 3 upper panels), the asymmetry of individual introgressions (main text section 1.2; Figure 3 middle panels), and the sex asymmetry of F1 fitness (main text section 2.2; Figure 4). Adapting previously published work (Poon and Otto 2000; Tenaillon *et al*. 2007; Zhang 2012; Roze and Blanckaert 2014), we can make some comments about the typical values of parental maladaptedness at mutation-selection-drift equilibrium.

Results are easiest to derive when evolution is mutation limited ($N\mu \ll 1$), as we assumed in our simulations. In this case, Tenaillon *et al*. (2007), following Poon and Otto (2000), showed the equilibrium fitness to be

$$\hat{w} \approx \exp\left(-\frac{n}{2kN_e}\right) \tag{63}$$

From this, we can approximate the equilibrium value of the parental phenotype as

$$E[z] \approx \left(\frac{n}{2k\sigma N_e}\right)^{1/k}, \qquad N\mu \ll 1 \tag{64}$$

Notice that the level of maladaptation increases with n, explaining why simulated populations were further from the optimum with $n = 50$ than with $n = 5$ (Figs. 3, S1-S4). Since the term in brackets is likely to be small in all but the very smallest populations, the degree of parental maladaptation will generally increase with k, and this will further increase the ability of the model to reproduce the basic observations of speciation genetics when $k > 2$ and $N\mu \ll 1$.

With higher mutation rates ($N\mu > 1$), we must account for standing genetic variation, and this makes results more sensitive to details of the model, such as the distribution of mutation effect sizes (Zhang 2012; Roze and Blanckaert 2014). We can say, however, that the dependency on n does remain in some conditions. For example, with multivariate normal mutations with variance v (eq. 47), and very weak selection, we find $E[z] \approx \sqrt{vn}/2$ (which follows by further approximation from eq. 8 of Roze and Blanckaert 2014).

6. Testing for the snowball effect

Incompatibility-based models predict that two-locus incompatibilities will increase with the genetic divergence, d, as $\binom{d}{2} = (d^2 - d)/2$ (Orr 1995), and so an appropriate test of the snowball effect is a quadratic regression forced through the origin:

$$E[I] = \beta_1 d + \beta_2 d^2 \tag{65}$$

where I is the number of introgressions that have a strongly deleterious fitness effect. For most real world data, this approach should not be applied directly, due to the phylogenetic non-independence of the data points, e.g. the use of a common focal species meaning that some incompatibilities are counted multiple times (Wang *et al.* 2013). Our simulated data (Figures 5, S7, S8 and S10), differ from most real-world data sets, in that they comprise a single pair of populations sampled over time. This implies that incompatibilities can both appear and disappear

over time, as the populations continue to evolve, and also that existing phylogenetic methods (Wang *et al.* 2013) are not appropriate. For our simulated data, an appropriate approach is to use datapoints that represent the additional incompatibilities accumulated (or lost) over a given period of divergence, denoted as $\triangle d \equiv d_2 - d_1$.

$$\begin{aligned} E[I_2 - I_1] &= \beta_1(d_2 - d_1) + \beta_2(d_2^2 - d_1^2) \\ &= \beta_1 \triangle d + \beta_2 \triangle d(\triangle d + 2d_1) \end{aligned} \tag{66}$$

If we choose datapoints such that $\triangle d$ is constant, i.e., if we simulate introgressions after a given number of substitutions have occurred, then we can fit a regression of the form:

$$E[I_2 - I_1] = \beta_0 + \beta d_1 \tag{67}$$

A value of β that is significantly greater than zero is consistent with a snowball effect, since it implies that the rate of accumulation of incompatibilities increases with divergence (or, equivalently, that the number of incompatibilities accumulates more rapidly than linearly with divergence).

Table S3 contains the results of fitting eq. 67 to our simulated data. We chose to simulate introgressions at 20 equally-spaced points, giving our regression tests 18 degrees of freedom. The model was fit by least squares, and the significance of the coefficients β_0 and β determined with standard t-tests. In all cases, we defined an incompatibility as an introgression that reduced fitness by at least 25%, and took as our data the results from 50 replicated simulations runs, thus approaching the high levels of genetic divergence that would characterise a real experiment. As such, our datapoints are equivalent to the green lines shown in Figures 5, S7, S8, and S10.

Results in Table S3 show that evidence of a snowball effect was never obtained when introgressions comprised single divergent sites. By contrast, evidence of an apparent snowball effect ($\beta > 0$) was regularly found from the same data when introgressions comprised multiple sites. The sole exceptions were the two sets of simulations that included adaptation to a rapidly moving optimum (Figure S8a and b). This seems to be due to the fixation of large effect mutations in this parameter regime, such that introgressions containing only a single divergent site were frequently strongly deleterious. Indeed, the same data did show evidence of an apparent snowball effect when we used a larger fitness cut-off, and defined "incompatibilities" as introgressions causing a reduction in fitness of at least 50% (not shown).

7. Approaches to modelling dominance

Manna *et al.* (2011) discuss the use of Fisher's geometric model to fit data on the dominance coefficients of new mutations. Their key results concern mutations which are weakly deleterious in homozygous form, but they also include mutations that are strongly recessive, and have very strong effects as homozygotes. Such mutations are analogous to those considered in the speciation genetics literature, since they have small fitness effects in one genetic background (i.e. when they appear as heterozygotes), and very strong negative effects in another background (i.e. when they appear as homozygotes). Manna *et al.* (2011) consider modelling such mutations with an incompatibility-based approach (see their Fig. 10B), and using a region of the fitness landscape with high curvature (see their Fig. 10A), which is roughly analogous to the high-k approach studied here. Manna *et al.* argue that an incompatibility-based approach is preferable for their data, because it more naturally generates the observation that the mean heterozygous effects of mutations remains very similar for mutations with small versus large homozygous effects.

Here, we suggest that this argument is compelling, but not decisive, because the same results can be obtained by using contrived fitness landscapes, but without including incompatibilities. To see this, let us define the following Fisher-type fitness landscape.

$$w(\mathbf{r}) = \begin{cases} e^{-\sigma||\mathbf{r}||^2} - 1, & g(\mathbf{r}) \leq 1 \\ \\ 0, & g(\mathbf{r}) > 1 \end{cases} \tag{68}$$

where

$$g(\mathbf{r}) \equiv \sum_{i=1}^{n} \left(\frac{r_i}{c_i} \right)^2 < 1 \tag{69}$$

As such, mutations are lethal if they lead to a phenotype falling outside of an ellipse, defined by $g(\mathbf{r})$, but otherwise have fitness values assigned by an isotropic landscape, with $k = 2$. Such a landscape is illustrated for $n = 2$ in Figure S14, and roughly resembles the landscape shown in Fig. 10A of Manna *et al.* (2011).

Assuming that the wild-type phenotype is optimal, the fitness effects of a mutation in heterozygous and homozygous forms are defined as follows.

$$hs \equiv w(\mathbf{r}) - 1 \tag{70}$$

$$s \equiv w(2\mathbf{r}) - 1 \tag{71}$$

To explore the behaviour of this model, we simulated 5×10^5 mutations with $n = 5$ and $\sigma = 0.5$. The z_i were drawn from a multivariate normal distribution, with mean 0 and variance $v = 1/(2n) = 0.1$ (see eq. 47). In each case, we compared the mean and standard deviation of the heterozygous effects for simulated mutations that were lethal versus non-lethal in homozygous form (any heterozygous lethals were discarded). The proportion of mutations in each category is not reported since this can be varied by introducing directional bias (although this was not done in these simulations). Results, shown in Table S4, show that the distribution of *hs* can be relatively independent of *s* when we incorporate very strong heterogeneity in the c_i, that is, by relaxing the assumption of isotropy for the strongly-deleterious mutations.