bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

In search of the Goldilocks zone for hybrid speciation

Alexandre Blanckaert *[1] and Claudia Bank[1]

[1]Instituto Gulbenkian de Ciência, 2780-156 Oeiras, Portugal

February 15, 2018

Abstract

Hybridization has recently gained considerable interest both as a unique window for observing speciation mechanisms and as a potential engine of speciation. The latter remains a controversial topic. It has been hypothesized that the reciprocal sorting of genetic incompatibilities from parental species could result in hybrid speciation, when the hybrid population maintains a mixed combination of the parental incompatibilities that prevents further gene exchange with both parental populations. However, the specifics of the purging/sorting process of multiple incompatibilities have not been examined theoretically.

We here investigate the allele-frequency dynamics of an isolated hybrid population that results of a single hybridization event. Using models of 2 or 4 loci, we investigate the fate of one or two genetic incompatibilities of the Dobzhansky-Muller type (DMIs). We study how various parameters affect both the sorting/purging of the DMIs and the probability of observing hybrid speciation by reciprocal sorting. We find that the probability of hybrid speciation is strongly dependent on the genetic architecture (i.e. the order and recombination distance between loci along chromosomes), the population size of the hybrid population, and the initial relative contribution of the parental populations to the hybrid population. We identify a Goldilocks zone for specific genetic architectures and intermediate recombination rates, in which hybrid speciation becomes highly probable. While an equal contribution of parental populations to the hybrid populations maximizes the hybrid speciation probability in the Goldilocks zone, other genetic architectures yield asymmetric maxima that are unintuitive on first sight. We provide an explanation for this pattern, and discuss our results both with respect to the best conditions for observing hybrid speciation in nature and their implications for patterns of introgression in hybrid zones.

1 Introduction

The role of hybridization for adaptation and speciation is an ongoing question that has been widely investigated and discussed Barton and Bengtsson (1986); Rieseberg (1997); Arnold et al. (1999); Buerkle et al. (2000); Barton (2001); Mallet (2007); Abbott et al. (2013); Servedio et al. (2013); Nieto Feliner et al. (2017); Schumer et al. (2018). On the one hand, hybridization may serve as a source of genetic variation.Various examples of adaptive introgression have been reported, reviewed in Hedrick (2013), and it has been argued that hybridization may provide the fuel for adaptive radiations Seehausen (2013). On the other hand, gene flow between diverging population may slow down or even reverse speciation either by purging isolating barriers or by one population swamping the other(Seehausen et al., 2008; Turissini et al., 2017). Thus, hybridization may act both as an engine of speciation and boost to genetic variation and as a detrimental mechanism that reduces population fitness and promotes extinction. This duality makes hybridization an important subject of study not only from an evolutionary but also a conservation biology point of view.

*ablanckaert@gulbankian.igc.pt

bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Hybrid speciation describes a scenario in which hybridization is essential for the formation of a "daughter" species that is isolated from both its parental species. The term "hybrid speciation" covers different scenarios that can be distinguished by the mechanism responsible for the buildup of reproductive isolation. In the case of polyploidization, the newly formed species consists of a fusion the genome of the two parents. The parents can be of the same species (autopolyploidization, although the hybrid species tends to be outcompeted by the parental diploid (Mallet, 2007)), or different ones (allopolyploidization), resulting in a single-step speciation event. In contrast, homoploid speciation (or recombinational speciation) corresponds to the formation of a hybrid species without a change in the ploidy level. This mechanism requires the existence of genetic barriers between the parental populations and the newly formed hybrid population, while still allowing the formation of sufficiently fit F1 hybrids. Despite this apparent paradox, numerous empirical cases have recently been reported Schwarz et al. (2005); Mavárez et al. (2006); Larsen et al. (2010); Hermansen et al. (2011); Kang et al. (2013); Yakimowski and Rieseberg (2014); Lamichhaney et al. (2018). Whether all of these represent true cases of homoploid hybrid speciation has been subject to debate. This debate has been led mainly around the definition of hybrid speciation and the resulting implications for the reported cases of empirical evidence (Schumer et al., 2014; Nieto Feliner et al., 2017; Schumer et al., 2018). However, to our knowledge there exists little work that has evaluated the probability of hybrid speciation theoretically.

Buerkle et al. (2000) studied the specific case of hybrid speciation via 2 overlapping parental inversions. Their simulations suggested a rather narrow parameter range in which hybrid speciation is possible, and indicated that (among other restrictions) high fertility of F1 hybrids is necessary to produce a stable hybrid population, which , as a consequence, is only poorly isolated from its parental species. Moreover, (Schumer et al., 2015) studied the conditions for reciprocal sorting of parental (Bateson-)Dobzhansky-Muller incompatibilities (DMIs; (Bateson, 1909; Dobzhansky, 1936; Muller, 1942)). A DMI consists of two (individually neutral or beneficial) alleles at different loci that are negatively epistatic, i.e., their combination is deleterious. Because epistasis in a DMI is by definition asymmetric (Orr, 1995), a single DMI is a poor barrier to gene flow and will eventually be purged from the population (Barton and Bengtsson, 1986; Gavrilets, 1997; Bank et al., 2012). However, if multiple DMIs exist between two species, in a hybrid population they might be resolved reciprocally with respect to the parental allelic origin, which can result in a hybrid species that is isolated from both parental populations. This model was proposed in (Schumer et al., 2015). The authors demonstrated via simulations that pairs of DMIs can trigger hybrid speciation when the derived (incompatible) alleles are strongly beneficial in the hybrid population.

Here, we provide a detailed analysis of Schumer's model in which we identify several parameters that greatly influence the probability of hybrid speciation via DMIs when direct selection on the derived alleles is less strong. Specifically, we quantify how the population size, the initial contribution of parental alleles, and the the genetic architecture affect the probability of hybrid speciation. As genetic architecture, we define the relative position of the different loci involved in the hybrid incompatibilities that contribute to the species barriers (see also figure 1). Consistent with (Schumer et al., 2015), we define the hybrid speciation as the successful reciprocal sorting of incompatibilities, independent of the amount of isolation they confer. We discuss both weak and strong isolating barriers and consider recessive and codominant architectures of the DMIs (Turelli and Orr, 2000; Bank et al., 2012), which differ considerably in their sorting patterns. Our results indicate that the genetic arrangement of the DMIs plays an essential role, such that a specific arrangement can make hybrid speciation almost unavoidable, whereas a different one may make hybrid speciation impossible for otherwise unaltered parameter values. Thus, we identify a Goldilocks zone of hybrid speciation, in which an interplay of various factors may make hybrid speciation more likely than previously assumed.

bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

2 Model

We consider a single population model of constant size N in discrete generations.We model four diallellic loci, A_1, A_2, B_1, B_2; the lower-case letter corresponds to the ancestral allele and the upper-case letter to the derived one. Note that we do not detail here the two-locus model as it is fully included in the four-locus model. It can be obtained by keeping only loci A_1 and B_1. Derived alleles at the different loci are under direct selection (soft selection), with α_k the (direct) fitness effect of allele A_k and β_k of allele B_k. Selection happens in the diploid phase of the life cycle. In addition, negative epistasis, ϵ_k, (which determines the strength of hybrid incompatibility) happens in a pairwise fashion between the derived A_k and B_k alleles (with $k \in \{1, 2\}$). Dominance affects only the epistatic interactions. In this manuscript, we focus mainly on two cases of dominance, which have proven representative of the general patterns (Bank et al., 2012): a recessive scenario and a codominant scenario, illustrated in figure 1. We introduce ϕ_k^n a mathematical placeholder used to distinguish between the recessive and codominant scenario at the k DMI, with n the number of pairs of incompatible alleles. Note that $n = 1$, $n = 2$ and $n = 4$ correspond to the H_0, H_1 and H_2 incompatibilities in Turelli and Orr (2000). Therefore, for a codominant DMI, ϕ_k^n is always equal to one $\forall n \in 0, 1, 2, 4$ while for a recessive DMI, the effect of epistasis is masked for the double heterozygote genotype, i.e. $\phi_k^1 = 0$ while $\forall n \in 0, 2, 4$, $\phi_k^n = 1$.

The population is initially composed of two single genotypes, since it results from secondary contact between two parental populations 1 and 2; i_p denotes the contribution of the parental population 1 to the newly formed hybrid population. We assume that the parental population 1 is fixed for the $A_1b_1A_2b_2/A_1b_1A_2b_2$ genotype and the parental population 2 for $a_1B_1a_2B_2/a_1B_1a_2B_2$. The fitness of a genotype composed of haplotypes i and j is given by:

$$\omega^{ij} = \prod_{k=1}^{2} (1+\alpha_k)^{X_k^i + X_k^j} (1+\beta_k)^{Y_k^i + Y_k^j} \left(1 + (\phi_k^{(X_k^i + X_k^j)*(Y_k^i + Y_k^j)} * \epsilon_k)\right)^{(X_k^i + X_k^j)*(Y_k^i + Y_k^j)}, \quad (1)$$

where X_k^i is the number of alleles A_k in haplotype i, Y_k^i the number of alleles B_k in haplotype i.

Mating is random. We assume that the parents generate an infinite pool of gametes, from which zygotes are formed through multinomial sampling $M(2N, p_1, ..., p_{16})$.

As introduced above, hybrid speciation is defined as the fixation of an haplotype that is incompatible both two parental haplotypes, see table 1. Indeed, if an individual homozygous for the $A_1b_1a_2B_2$ haplotype is backcrossed with an individual from, e.g. parental population 1, then the second DMI is expressed either in the F1 generation (codominant case) or in the F2 generation (recessive one). Similar introduction of such $A_1b_1a_2B_2/A_1b_1a_2B_2$ individual in the parental population 2 leads to the expression of the first DMI. This definition corresponds to an early stage mechanism, leading to an hybrid population that is only partially isolated from both parental population. Note that full isolation is impossible in this setting, as the barriers responsible for the full reproductive isolation, will also prevent the formation of the hybrid population in the first place.

We consider all possible genetic architectures formed by the two DMIs; they are illustrated in Figure 2. There are 6 different ways to organize the 4 loci along a single chromosome (assuming the chromosome does not have an orientation). The two DMIs can be "adjacent", "crossed", or "nested" (Fig. 2). Genetic distance between adjacent loci X and Y is given by $0 \leq r_{XY} \leq 0.5$. The distance between non-adjacent loci X and Y, separated by a single locus W, is given as follows $r_{XY} = r_{XW}(1 - r_{WY}) + r_{WY}(1 - r_{XW})$. If the four loci are spread across multiple chromosomes, this represents a special case of the single chromosome scenarios presented above, in which one or more recombination rates are set to 0.5. If not otherwise specified, we assume that all loci are located on different chromosomes, i.e. $r_{XY} = 0.5$.

bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Ancestral haplotype	$a_1b_1a_2b_2$
Parental pop. 1 haplotype	$A_1b_1A_2b_2$
Parental pop. 2 haplotype	$a_1B_1a_2B_2$
Hybrid haplotypes	$A_1b_1a_2B_2$ or $a_1B_1A_2b_2$
Partly diverged haplotypes	$A_1b_1a_2b_2$ or $a_1B_1a_2b_2$ or $a_1b_1A_2b_2$ or $a_1b_1a_2B_2$
1st incompatibility haplotypes	$A_1B_1a_2b_2$ or $A_1B_1A_2b_2$ or $A_1B_1a_2B_2$
2nd incompatibility haplotypes	$a_1b_1A_2B_2$ or $a_1B_1A_2B_2$ or $A_1b_1A_2B_2$
Both incompatibilities haplotype	$A_1B_1A_2B_2$

Table 1: Classification of possible haplotypes for the "adjacent" linkage architecture.

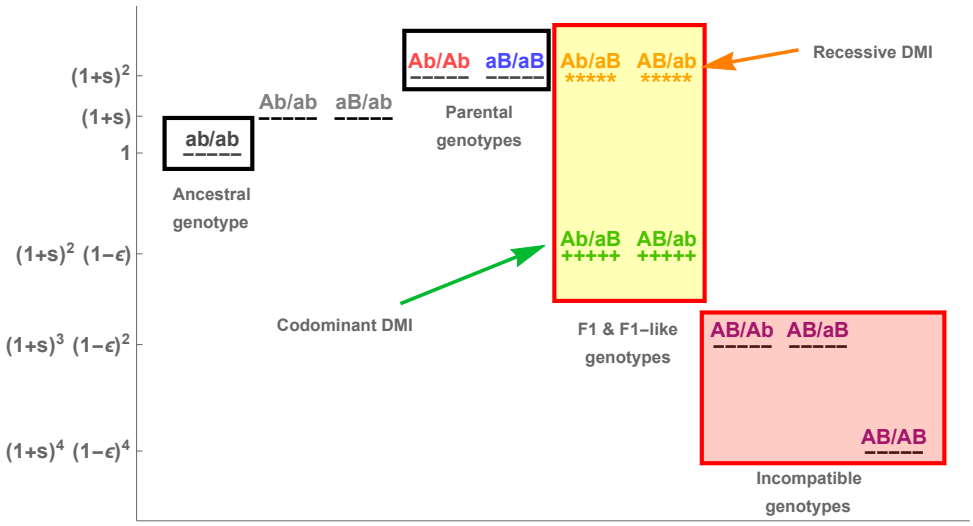


Figure 1: Fitness landscape of the 16 genotypes in the two-locus model, highlighting the effect of dominance of the incompatibility on the fitness of F1 hybrids. For simplicity, we illustrate the case of $\alpha = \beta = s$. Note that there is only 10 genotypes represented here, as we do not distinguish between the parental origin of each haplotype.

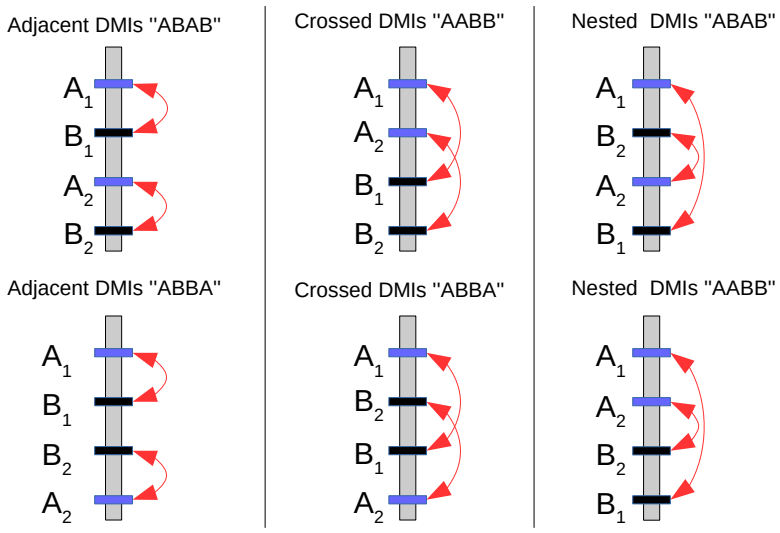


Figure 2: Illustration of all 6 different linkage architectures possible along a single chromosome. The A_k loci are given in blue and B_k in black. Red arrows show the incompatible interactions. The name of each architecture depends on the arrangements of the two incompatibilities and the order of the A and B loci.

bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Simulations are implemented in C++ and end when the population is monomorphic. Note that the deterministic case (i.e., in the absence of drift, $N \to \infty$) can be obtained by skipping the multinomial sampling step during zygote formation.

3 Results

3.1 Resolution of a single DMI

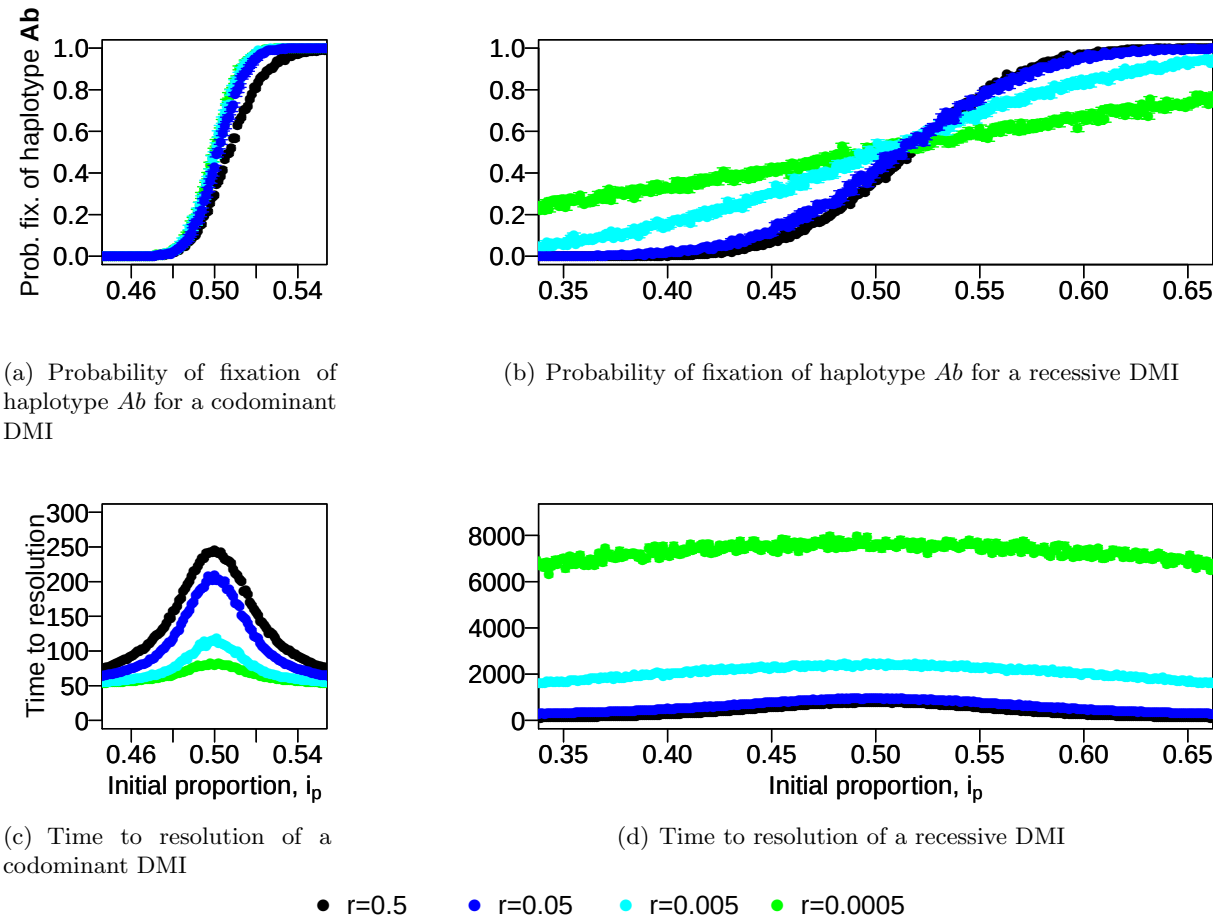
In the first part, we focus on the resolution of a single DMI following the formation of the hybrid population. With a single DMI, hybrid speciation according to our definition is impossible, because one of the negatively interacting partners in the DMI will invariably be lost, which makes the maintenance of a genetic barrier to both parental species impossible. We characterize the resolution of the genetic conflict resulting from the contact between two diverged populations by quantifying: the probability of fixation of the different haplotypes, the time of resolution of the DMI (i.e., the time until at least one of the incompatible alleles is lost) and the time to fixation of a single haplotype. For this section, we only focus on the A_1 and B_1 loci and drop the indices as they do not carry any information.

3.1.1 Dynamics following secondary contact In a single randomly mating population such as the hybrid population we consider here, a DMI cannot be maintained unless directional selection is large as compared with the epistatic effect of the incompatibility Bank et al. (2012). This is because the formation of hybrid individuals initially leads to negative selection against both derived haplotypes. These haplotypes suffer from the incompatibility, either directly by forming an unfit hybrid genotypes or indirectly through the production of unfit offspring. In contrast, the ancestral haplotype has an advantage as soon as it appears and rises in frequency, because it only forms compatible genotypes and produces compatible offspring (assuming that the proportion of incompatible AB haplotypes in the population remains low). As soon as the ancestral type becomes frequent or either of the derived types becomes rare, this marginal advantage disappears, and the ancestral type will either be swamped by the more frequent derived type (in the case of direct selection acting on the derived alleles, i.e., if $\alpha, \beta > 0$), or segregate neutrally (if $\alpha, \beta = 0$). The incompatibility is usually resolved in favor of the more frequent derived allele (if they have similar fitness effects), one main determining factor being the initial frequency ratio between the two derived alleles (Fig. S2). Direct selection, as well as codominance of the incompatibility, reduces the impact of genetic drift (i.e., the outcome converges to the deterministic case). Indeed, once the DMI is resolved, selection increases the probability of fixation of a single derived allele (Haldane, 1927; Kimura, 1962). The codominance of the incompatibility shortens the time required to resolve the DMI (Fig. 3), and therefore reduces the time spent at low frequencies, where loss of the derived alleles because of drift is a likely outcome.

3.1.2 Recombination has opposite effects under different dominance schemes Recombination, because of its dual nature, has a converse impact on the outcome of an hybridization event, depending on the dominance, as illustrated in Figure 3 for haplotype Ab. Indeed, recombination breaks the association between the alleles of the parental haplotype and therefore leads to the formation of both the incompatible haplotype AB and the ancestral haplotype ab. On the one hand, this allows the expression of the incompatibility through the formation of the AB haplotype, leading to a faster sorting of the derived alleles. On the other hand, building a genotype with the ancestral haplotype protects both parental haplotypes from suffering from the genetic incompatibility, leading to a slower sorting of the derived alleles. The balance between these two effects is different for a recessive and a codominant DMI, leading to this opposite behavior.

In the recessive case, recombination is necessary for the expression of the incompatibility. Thus, the need to form the incompatible haplotype overcomes any cost of generating the an-

bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



(a) Probability of fixation of haplotype *Ab* for a codominant DMI

(b) Probability of fixation of haplotype *Ab* for a recessive DMI

(c) Time to resolution of a codominant DMI

(d) Time to resolution of a recessive DMI

Figure 3: Recombination affects in a contrary manner the resolution of a codominant DMI while it speeds it up for a recessive one. We represent the probability of fixation of the *Ab* haplotype (top) for different recombination rates and different dominance schemes (codominant left, recessive right). We also illustrated (bottom) the time to resolve the genetic conflict (i.e., either allele A or B is lost). Each value is estimated over 1000 independent simulations. Note the much larger scale (x30) for panel d). Parameters used: $\alpha = \beta = 0.001, \epsilon = 0.2, N = 5000$.

bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

cestral haplotype. An increase in recombination therefore always accelerates the resolution of a recessive DMI. This reduces the time the derived alleles spend at low frequency, which makes them less susceptible to being lost through genetic drift. This, in turn, reduces the probability that the ancestral haplotype becomes fixed.

In the codominant case, the incompatibility is already expressed in the F1 generation. Recombination is not necessary to express the incompatibility and therefore slows down the resolution of the DMI, as the ancestral haplotype prevents the effective purging of the parental haplotypes through the formation of ab/Ab or ab/AB individuals. In this situation, both derived alleles remain at a lower frequency much longer than in the recessive model, which makes them more likely to be both lost through genetic drift, resulting in the fixation of the ancestral haplotype.

3.2 Resolution of two DMIs and hybrid speciation

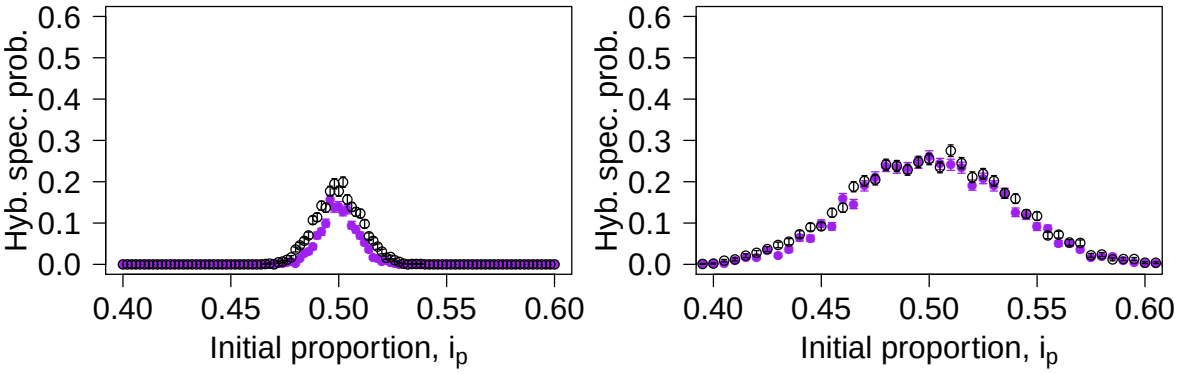
We now focus on the simultaneous resolution of two independent DMIs. Expanding from what we learned above, we now consider what happens when two incompatibilities exist between the parental populations. In contrast to the case of a single DMI, a new evolutionary outcome, namely hybrid speciation, becomes feasible with more than one DMI. By "hybrid speciation", we mean the reciprocal sorting of the two DMIs, i.e. fixation of either alleles A_1 and B_2 or A_2 and B_1. Such a hybrid population will then be genetically isolated from both parental populations.

3.2.1 Isolation of the hybrid population by opposite resolution of two DMIs Hybrid speciation is obtained through the reciprocal sorting of the two DMIs. Given the observed shape of the fixation probability of a derived allele in a single DMI case as a function of the initial contribution of both parental populations (Fig. 3), hybrid speciation should be observable only around a symmetric contact, and this condition should be more stringent for codominant incompatibilities than recessive ones (cf. Fig. 4). In Figure 4, we test this expectation by comparing the probability of hybrid speciation for two DMIs that are located on separate chromosomes (i.e., the "adjacent" architecture from Fig. 2; colored dots in Fig. 4), with the expected probability of resolving two independent single DMIs for opposite derived alleles (e.g. first DMI resolved towards allele A and the second one for allele B; black dots). In the recessive case, the prediction for independent DMIs matches the hybrid speciation probability. In the codominant case, the independent expectation overestimates the probability of hybrid speciation. This can be explained by the faster resolution of the DMIs in the codominant model, which, even in the case of free recombination, leaves insufficient time for the two DMIs to become uncoupled and independently resolved, as the A_1 and A_2 loci start in maximum linkage disequilibrium. In the codominant case, this effect is amplified at low recombination rates as, in that case, the resolution of the DMIs happens even faster (Fig. S4), therefore preserving more of the initial linkage disequilibrium. This leads to a positive correlation between the fixation of the different A_i alleles (as well as B_j alleles), Fig. S3. For the recessive case, the resolution of the two DMIs remains independent as it takes much longer to resolve the DMI.

Figure S4 illustrates the mean time to resolve both DMIs in opposite directions conditioned on the outcome being hybrid speciation. Recombination has the same effect on the resolution of two DMIs than it did for a single one: it accelerates the resolution of recessive DMIs and slows down the resolution of codominant ones. However, the average resolution time is not affected by the initial proportion of the parental species; only trajectories that quickly resolve in the right direction can contribute to hybrid speciation, and we are conditioning on this outcome.

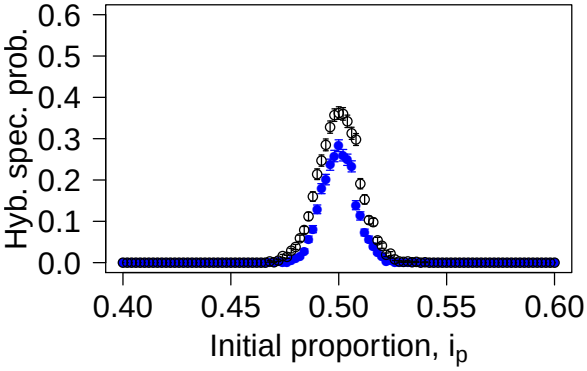
3.2.2 The linkage map determines which alleles survive Fig. 5 illustrates the effect of recombination and the genetic architecture on hybrid speciation, when all loci are on the same chromosome (as opposed to one DMI per chromosome, as illustrated in Fig. 4). As mentioned above, for codominant DMIs, recombination, on the one hand, allows the formation of the hybrid haplotype and helps to reduce the initial linkage disequilibrium. On the other

bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

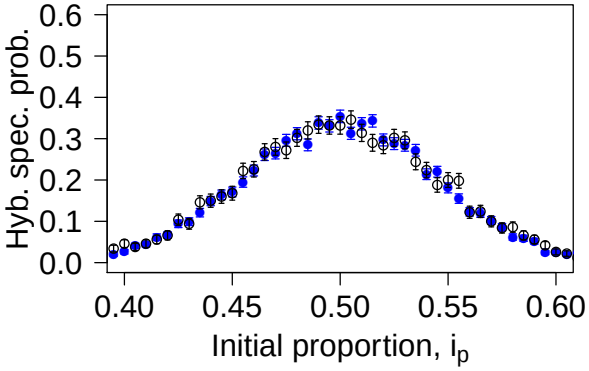


(a) Codominant DMIs, $r = 0.5$

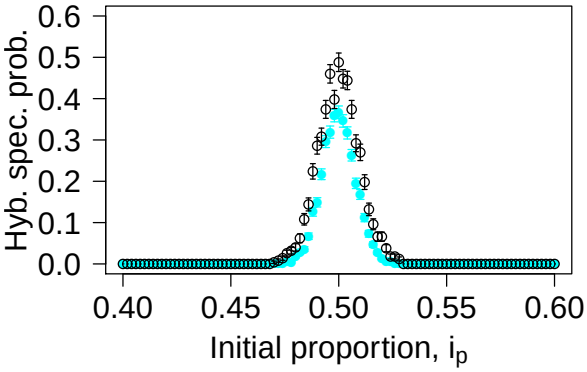
(b) Recessive DMIs, $r = 0.5$



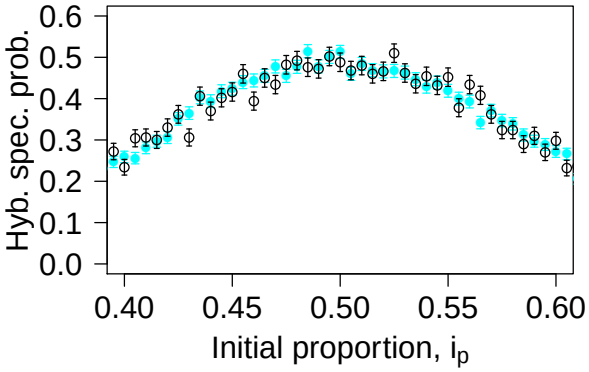
(c) Codominant DMIs, $r = 0.05$



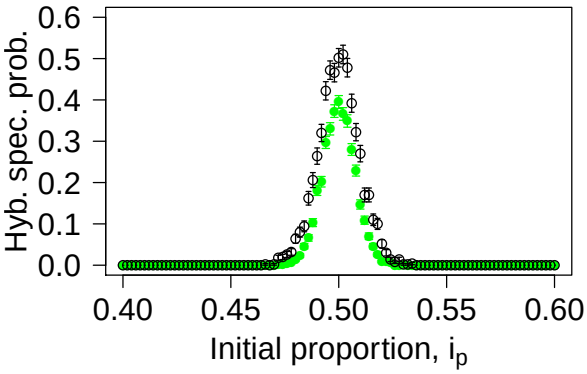
(d) Recessive DMIs, $r = 0.05$



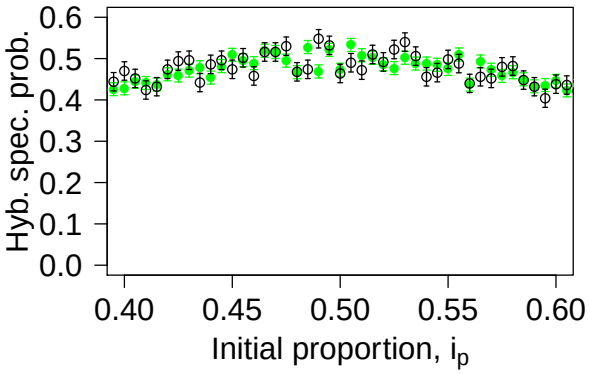
(e) Codominant DMIs, $r = 0.005$



(f) Recessive DMIs, $r = 0.005$



(g) Codominant DMIs, $r = 0.0005$



(h) Recessive DMIs, $r = 0.0005$

Figure 4: Hybrid speciation probability for codominant (left panels) and recessive (right panels) DMIs. The colored dots correspond to the probability of hybrid speciation for two DMIs situated on different chromosomes ($r_{23} = 0.5$). The genetic distance between the interacting loci is indicated below each panel ($r_{12} = r_{34} = r$). The black dots correspond to the predicted hybrid speciation probability based on the resolution of a single DMI. The fast resolution of the codominant DMIs leads to a correlation between their fate, which makes hybrid speciation less likely. Parameters used are $\alpha_i = \beta_i = .001, N = 5000, \epsilon = 0.2$. Each dot is obtained for 1000 replicates.

bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

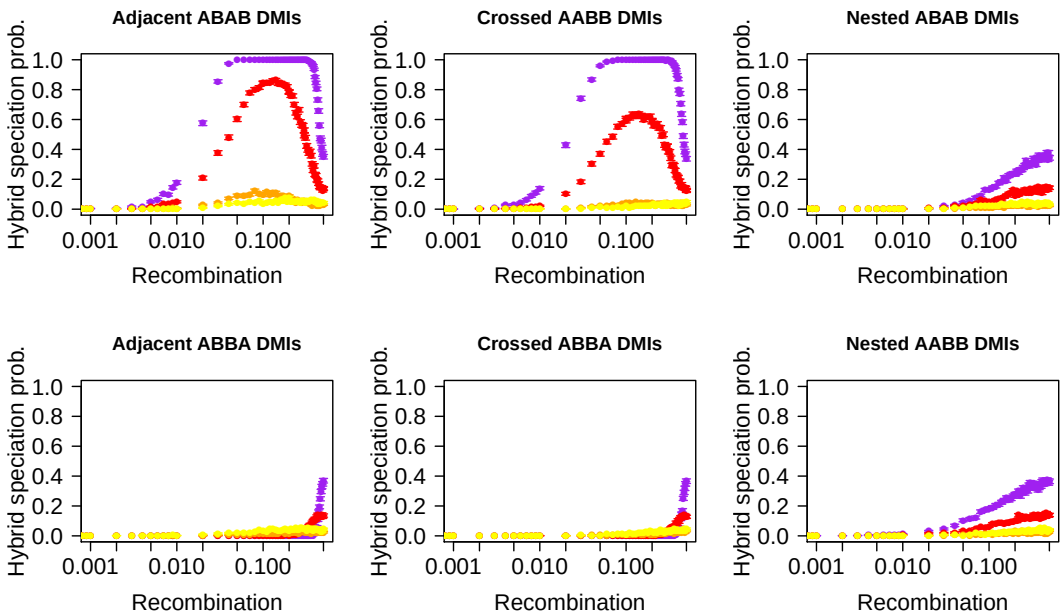


Figure 5: Hybrid speciation probability is a nonlinear function of recombination. We consider that all four loci have the same selective advantage ($\alpha_k = \beta_j = .001$) and are equidistant along a single chromosome. The hybrid speciation probability is plotted for different population sizes: yellow corresponds to $N = 50$, orange to $N = 500$, red to $N = 5000$ and purple to $N = 50000$. Epistasis ($\epsilon = 0.2$) is here codominant but we obtain qualitatively the same results for recessive incompatibilities see Fig. S7. The contribution of both parental populations here is symmetric ($i_p = 0.5$).

hand, it slows down the resolution of the DMI through the formation of compatible haplotypes. Depending on the balance between these two effects, recombination impacts the probability of hybrid speciation differently.

Assuming a symmetric contact, we observe that two of the genetic architectures, the "adjacent ABAB" and "crossed AABB" ones, in which the A_1 and B_2 are on the edge of the chromosome, exhibit a non-monotonic behavior with maximum hybrid speciation probability for intermediate recombination rates. This behavior is most extreme for large population size and indeed corresponds to the deterministic outcome for these two genetic architectures. More precisely, we observe a local maximum of hybrid speciation probability for recombination rates around $r = 0.1$. The "adjacent ABAB" and "crossed AABB" architectures, that show this behavior, are characterized by a higher marginal fitness of the A_2 and B_1 alleles compared to the other alleles in the deterministic case, which promotes hybrid speciation. For all other architectures either either A_1 and A_2 or A_2 and B_2 have the highest marginal fitness, see Fig S5. The higher fitness stems from the production of the $a_1B_1a_2b_2$ and $a_1b_1A_2b_2$ haplotypes (for the "adjacent ABAB" architecture) that are relatively free of epistasis in the F2 generation. The outcomes of a single recombination event per genome for all 6 architectures are given in Table 2 and illustrates how the "adjacent ABAB" and "crossed AABB" architectures stand out in the production of the haplotypes that are needed of hybrid speciation. Importantly, recombination is necessary to generate these haplotypes, but too much recombination will cancel their advantage. Indeed, for $r = 0.5$, all haplotypes are produced in the same frequency in the absence of selection. This dual effect of recombination leads therefore to the observed maximum in the hybrid speciation probability for intermediate recombination rates. When the DMIs are located on two different chromosomes (as in Fig 4), this effect does not appear. Indeed, while recombination still breaks linkage disequilibrium, it no longer generates the relatively "epistasis-free" haplotype and therefore leads to a monotonous increase in the hybrid speciation probability with

bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

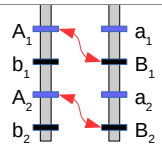
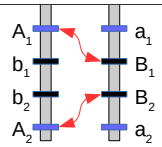
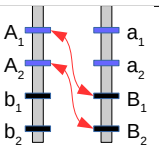
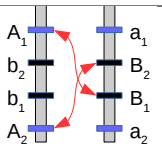
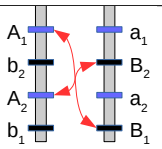
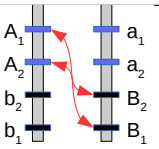
Genetic architecture	F1 hybrid	Single recombination event between: loci 1 and 2	loci 2 and 3	loci 3 and 4
Adj. ABAB		$A_1B_1a_2B_2$ $a_1b_1A_2b_2$	$A_1b_1a_2B_2$ $a_1B_1A_2b_2$	$A_1b_1A_2B_2$ $a_1B_1a_2b_2$
Adj. ABBA		$A_1B_1B_2a_2$ $a_1b_1b_2A_2$	$A_1B_1b_2a_2$ $a_1b_1B_2A_2$	$A_1b_1b_2a_2$ $a_1B_1B_2A_2$
Crossed AABB		$A_1a_2B_1B_2$ $a_1A_2b_1b_2$	$A_1A_2B_1B_2$ $a_1a_2b_1b_2$	$A_1A_2b_1B_2$ $a_1a_2B_1b_2$
Crossed ABBA		$A_1B_2B_1a_2$ $a_1b_2b_1A_2$	$A_1B_2b_1a_2$ $a_1b_2B_1A_2$	$A_1b_2b_1a_2$ $a_1B_2B_1A_2$
Nested AABB		$A_1a_2B_2B_1$ $a_1A_2b_2b_1$	$A_1A_2B_2B_1$ $a_1a_2b_2b_1$	$A_1A_2b_2B_1$ $a_1a_2B_2b_1$
Nested ABAB		$A_1B_2a_2B_1$ $a_1b_2A_2b_1$	$A_1b_2a_2B_1$ $a_1B_2A_2b_1$	$A_1b_2A_2B_1$ $a_1B_2a_2b_1$

Table 2: Haplotypes produced in the F2 breakdown, assuming a single recombination event, explains how different genetics architectures leads to different outcomes for the same loci. By identifying the relatively "epistasis-free" haplotype formed, one can infer whether hybrid speciation may be a likely outcome. In blue, we highlight these "epistasis-free" haplotypes that are important for hybrid speciation and in yellow for fixation of the parental haplotype from population 1.

increasing recombination rate. This behavior, specific to the "adjacent ABAB" and "crossed AABB" genetic architectures is observed both for codominant and recessive DMIs.

As illustrated in Fig. S7 , the recessive case is qualitatively similar to the codominant one. We recover the distinctive pattern between genetic architectures, where the "adjacent ABAB" and "crossed AABB" architectures are more likely to generate hybrid speciation for intermediate recombination rates. However, for the "adjacent ABBA" and "crossed ABBA" genetic architectures, the recessive case differs from the codominant by the existence of two local maxima for the hybrid speciation probability as a function of recombination. These two architectures are characterized by in indirect selective advantage of one of the two parental haplotypes over the other, as shown in table 2, as the partially derived haplotypes $A_1b_1b_2a_2$ and $a_1b_1b_2A_2$ are more likely to form than their counterparts ($a_1B_1b_2a_2$ or $a_1b_1B_2a_2$, see Table 2), which leads to a slightly higher marginal fitness of the $A_1b_1b_2A_2$ haplotype compared to $a_1B_1B_2a_2$. The first maximum is obtained at large intermediate recombination rates; it corresponds to the one observed for codominant DMIs. However, a second one can be observed at lower recombination rate if the population size reaches certain sizes. It results from a subtle balance between drift, recombination and selection, which we explain in detail in the Supplement.

Lastly, the hybrid speciation probability for codominant versus recessive DMIs differs significantly when considering lethal incompatibilities, (see Fig. S6). Hybrid speciation becomes

bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

impossible for codominant DMIs because no viable hybrids can be produced. This is not the case for recessive incompatibilities, as they can partially escape the strong selection against hybrids. In fact, due to the masking effect provided in F1 and F1-like genotypes, we observe an almost indistinguishable pattern for deleterious ($\epsilon = -0.2$) and lethal ($\epsilon = -0.99$) recessive DMIs. Similarly, the time to observed hybrid speciation seems identical between the deleterious and lethal recessive cases.

Figure 5 also illustrates the impact of the population size on the outcome. In general a larger population size results in a higher probability of hybrid speciation. This is especially true when the deterministic outcome corresponds to hybrid speciation (i.e. the "adjacent ABAB" and "crossed AABB" architectures). Derived alleles are less likely to be lost during the reciprocal sorting of the genetic incompatibilities. The main exception to this rule exists when the deterministic outcome is the fixation of one parental haplotype. In that case, an intermediate population size will maximize the likelihood of hybrid speciation, as illustrated in Figure 5 for the "adjacent ABBA" and "Crossed ABBA" architectures (and Figure S7 and S9). This intermediate value corresponds to a balance between a strong drift regime in which the ancestral and "epistasis free" haplotypes are most likely to fix, and the deterministic regime in which the $A_1b_1b_2A_2$ parental haplotype fixes.

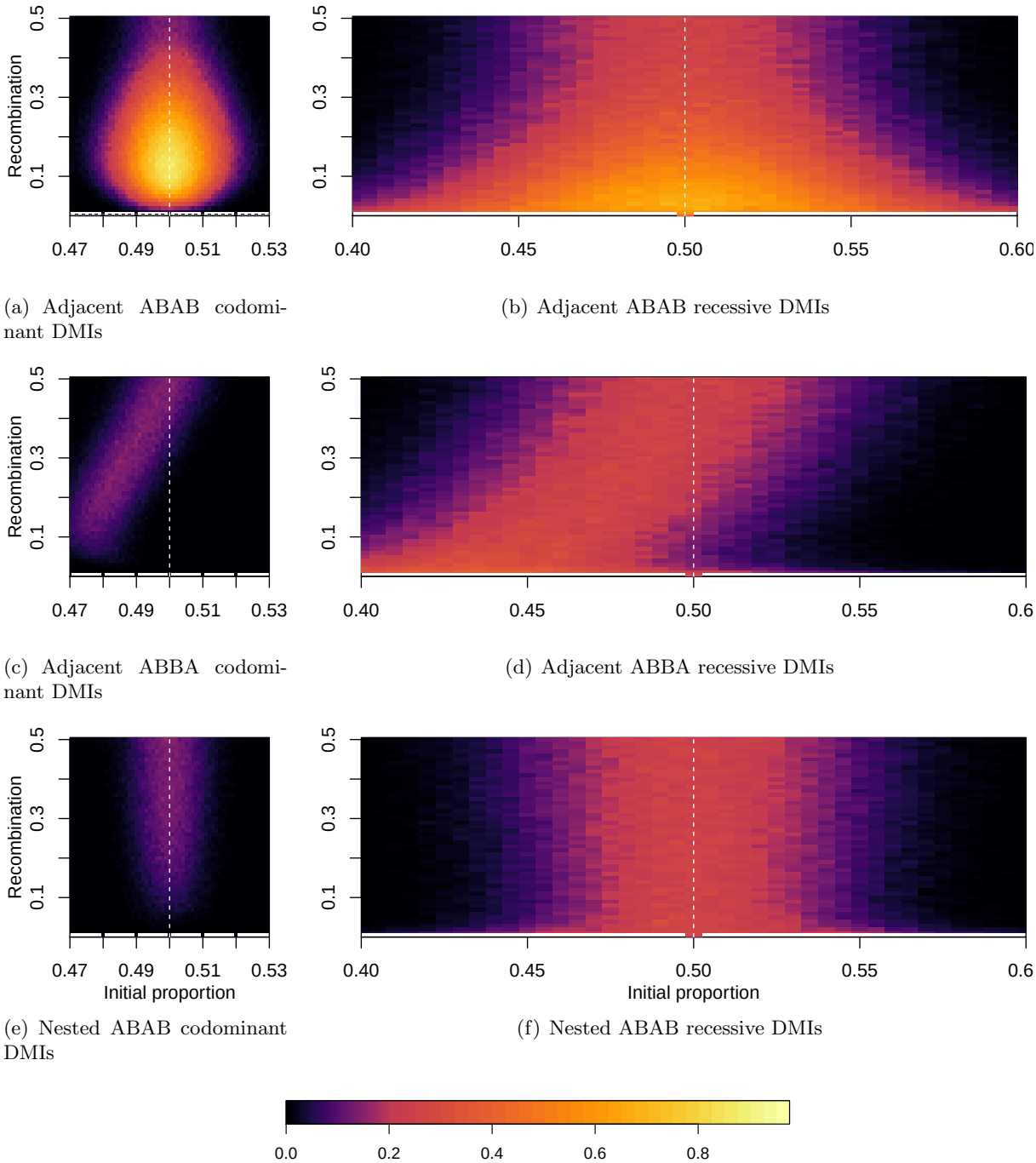
3.2.3 Symmetric contact is not always the best condition for hybrid speciation

Fig. 5 was obtained for $i_p = 0.5$, i.e. when both parental populations contribute equally to the hybrid population. It corresponds the case that is the most frequently investigated Schumer et al. (2015). Fig. 6 illustrates what happens when we release this assumption. From the single-DMI dynamics, one would expect a decrease in the hybrid speciation probability as illustrated in Fig. 5. This is not always true. Depending on the genetic architecture, the probability of hybrid speciation may be higher for asymmetric contributions from the parental populations. This phenomenon is also observed for intermediate recombination rate; thus, only a consideration of dominance scheme, recombination rate, and symmetry together allows for an accurate statement on the hybrid speciation probability (see Fig. 6). Table 2 provides us with an explanation for the observed pattern: for intermediate recombination rate ($r \approx 1/3$), there is on average one recombination event per haplotype. For the two architectures concerned ("adjacent ABBA" and "crossed ABBA"), in this scenario and with perfect symmetry, both alleles A_1 and A_2 have a marginal fitness that is slightly higher than alleles B_1 and B_2 (Fig. S5), which leads to the fixation of the parental haplotype $A_1b_1b_2A_2$ in the deterministic case. Therefore, a lower initial frequency of these alleles at the initial contact balances this selective advantage, which results in higher hybrid speciation probabilities than under symmetry. This behavior was only observed for the two architectures discussed above ("adjacent ABBA" and "crossed ABBA"). Indeed, for all other architectures, the two derived alleles that got a slight indirect selective are A_2 and B_1 for "adjacent ABAB" and "crossed AABB" (which corresponds to the cases of high probabilities of hybrid speciation) or A_2 and B_2 for the two "nested" architectures. In both cases, since the symmetry between the A and B alleles is respected, hybrid speciation is more likely at $i_p = 0.5$.

4 Discussion

We here characterized the purging process of single and multiple DMIs upon formation of an hybrid population. Specifically, we quantified the effects of the genetic architecture and the dominance of the epistatic interactions on the reciprocal sorting of incompatibilities, which has been proposed as a mechanism to induce homoploid hybrid speciation. We found that for the exact same set of loci, their order along the chromosome can increase the likelihood of observing hybrid speciation by more than an order of magnitude. We demonstrate that the main determinant of this pattern is which haplotypes are formed during the F2 breakdown. For the genetic architectures that promote hybrid speciation, there exists a Goldilocks zone in which an intermediate recombination rate maximizes the hybrid speciation probability. In addition, we show that symmetric contact of incompatible loci that are under equal selection pressure does

bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



(a) Adjacent ABAB codominant DMIs

(b) Adjacent ABAB recessive DMIs

(c) Adjacent ABBA codominant DMIs

(d) Adjacent ABBA recessive DMIs

(e) Nested ABAB codominant DMIs

(f) Nested ABAB recessive DMIs

(g) Hybrid speciation probability

Figure 6: Probability of hybrid speciation for both recessive and codominant DMIs as a function of the genetic distance and the initial contribution of both parental species. We represent how different genetic architectures generate different unexpected pattern: for the "adjacent ABAB" architecture, we observe a Goldilock zone for recombination; for the "adjacent ABBA", hybrid speciation is no longer symmetric along the $i_p = 0.5$ axis (the white dashed line).

bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

not always generate the highest probability of hybrid speciation, and that this result cannot be predicted from the study of independent DMIs. Finally, for recessive DMIs in which the F1 generation does not suffer a fitness disadvantage, reciprocal sorting is similarly probable with intermediate and strong epistasis. Conversely, hybrid speciation with lethal codominant DMIs is impossible.

Genetic architecture, genomic islands of divergence and speciation In this manuscript, we quantified the probability of hybrid speciation in a minimal model that considers various genomic parameters. The importance of this question has been recently emphasized by Abbott et al. (2013) "Thus, an important challenge in studies of hybrid speciation is to ask whether there is an 'optimal' genetic distance for homoploid hybrid speciation (Arnold et al., 1999; Gross, 2012)." Although Abbott et al. (2013) were mainly referring to the degree of divergence and thus, to the number of DMIs that have established between two species, we can add an additional important factor to their list: the genetic arrangement of the isolating barriers, and the recombination rate between them. Our results demonstrate that intermediate recombination rates and specific genetic architectures maximize the probability of hybrid speciation.

We can speculate whether the presence of multiple DMIs should increase the probability of hybrid speciation. Based on our results, we believe that too should depend on the nature of the incompatibilities: additional recessive DMIs should increase the hybrid speciation probability while more codominant DMIs should reduce it. Indeed, for codominant DMIs, the fixation of the different A_k (resp. B_k) is correlated (Fig. S3) despite them being located on different chromosomes. Recombination is not sufficient to decrease the initial linkage disequilibrium. Adding additional loci will result in stronger selection against F1 individuals. As we have seen (Fig. S6), for lethal DMIs, hybrid speciation is impossible. Extrapolating from these two observations, we expect that adding more DMIs will create stronger selective pressure against the F1 hybrids, which leads to a stronger correlation in the fixation of the different A_k alleles. We postulate that this effect will outpace the increase in hybrid speciation probability due to having more chances to have at least one pair of reciprocal sorting. On the other hand, since F1 hybrid do not suffer a fitness cost in the case of recessive DMIs, and since the fixation of the different A_k is not correlated, we believe that ia higher number of recessive DMIs should increase the probability of hybrid speciation. Indeed, stronger epistasis seems to not) affect the probability of hybrid speciation in the recessive model (Fig. S6 and S7 Thus, having more than two recessive DMIs should increase the chances that at least two are "reciprocally sorted".

When discussed in the more general context of speciation, our results can be integrated in the discussion about genomic islands of divergence (Via, 2012; Feder et al., 2012). Indeed, during the speciation process, these islands of divergence are formed around the first genes involved in reproductive isolation. These genes will reduce the gene flow locally around them, favoring the accumulation of weakly locally adapted mutations in their vicinity, forming those islands. Since the barrier loci need to be apart from each other to eventually form a genome-wide barrier, the formation of islands of divergence (in our model corresponding to a close distance between the A loci or the B loci) is not necessarily helpful for speciation. Indeed, if the loci are in tight linkage, then the reciprocal sorting of the DMIs is quite unlikely as the sorting at the different A loci will be positively correlated. In addition, as shown by Yeaman (2013), locally adaptive loci tend to be rearranged into clusters. This suggests that optimal configuration for hybrid speciation may be quite rare, especially since the time window in which it can happen may be restricted by this rearrangement mechanism.

The probability of hybrid speciation in nature Our results imply that while specific genetic architectures may indeed induce hybrid speciation with high probability, it remains on average unlikely, which is consistent with putative cases of homoploid hybrid speciation observed in nature Schwarz et al. (2005); Mavárez et al. (2006); Larsen et al. (2010); Hermansen et al. (2011); Kang et al. (2013); Yakimowski and Rieseberg (2014); Lamichhaney et al. (2018). Recently, Runemark et al. (2018) reported that the Italian sparrow hybrid species resulted from

bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

multiple occurrence of hybridization events between the Spanish and House sparrow along the Mediterranean Sea, suggesting that it may not be that rare. The resolution of all hybridization events towards a single mitochondrial origin (i.e. all Italian sparrow possess an House sparrow mitochondrial DNA) suggests that either the mitochondria play an important role in the sorting of the incompatibilities, or that there is an asymmetry in the formation of the different hybrid population, in which Spanish sparrow males mate with House sparrow females. In this manuscript, we only focus on autosomal chromosomes, who are inherited equally from both parents; the potential interplay between organelles and sex chromosomes may add to the complexity of the system; for an analytical comparison between dynamics of a single DMI on autosomal chromosome versus sex chromosome see Höllinger and Hermisson (2017). Nevertheless, we predict a similar pattern for specific genetic architectures ("adjacent ABAB" and "crossed AABB"). In these cases, hybrid speciation can become highly probably and result in the repeatability of a hybrid speciation even. In addition, for both of these architectures the reciprocal sorting of the DMIs is not random but happens for only one of the two possible "hybrid" haplotypes. Finally, we consider a single contact event without any further interactions with both parental populations, which makes our analysis more similar to the colonization of a new environment. Continuous gene flow from one or both parental population should further reduce the probability of hybrid speciation, because migration creates selective pressure against the hybrid haplotype.

The nature of genetic incompatibilities Both theoretical considerations and empirical evidence suggest that most DMIs should be recessive (Orr, 1993; Presgraves, 2003). However, any kind of dominance pattern of the epistatic interactions can in theory exist (Coyne and Orr, 2004). Here, we showed that codominant incompatibilities are resolved much faster than recessive ones. This indicates that we are less likely to observe the former in a polymorphic state whenever a hybridizing population is sampled. Therefore, the frequent observation of recessive incompatibilities (Presgraves, 2003) may not necessarily reflect the true proportion of recessive incompatibilities but rather a sampling bias. Another simplifying assumption that one could consider when studying multiple DMIs is independence of the fate of the non-interacting alleles. Here, we compared the differences between how DMIs on different chromosomes and independent DMIs are resolved. Whereas independent and unlinked DMIs behave similarly in the recessive case, in the codominant case hybrid speciation is less likely than expected from the prediction based on independent loci. That is because despite the absence of physical linkage, recombination does not break linkage disequilibrium fast enough to separate the fate of the different alleles. Thus, our work demonstrates that extrapolation from one to multiple incompatibility pairs is not straightforward. In the case of lethal incompatibilities, only recessive DMIs can be involved in hybrid speciation because the F1 generation in the codominant case does not survive to reproduce. Surprisingly, although a lethal recessive incompatibility generates a stronger genetic barrier to gene flow than a non-lethal recessive DMIs (Bank et al., 2012), the strength of epistasis does not affect the probability of hybrid speciation. Thus, an accumulation of weak recessive DMI pairs could result in optimal conditions for hybrid speciation. The results presented here correspond to the initialization of the reproductive isolation of the hybrid population from its parental sources. In this situation, since the genetic barrier will be weak, the buildup of additional isolating barriers is necessary to fully form a new hybrid species. Note that this later stage is similar to the parapatric model of speciation, therefore the current theory applies (Servedio and Noor, 2003).

Time to hybrid speciation Schumer et al. (2015) argued that hybrid speciation happens quickly, as corroborated by the latest empirical evidence Lamichhaney et al. (2018). We arrive at a more detailed view of the time to hybrid speciation: whereas codominant DMIs are indeed resolved quickly, this not the case for recessive mutations. Nevertheless, both codominant and recessive DMIs lead to similar times to hybrid speciation. This apparent contradiction appears because only fast resolving recessives DMIs result in hybrid speciation. In addition, after resolution of the DMIs, the derived alleles still need to become fixed; this process is usually faster

bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

in recessive populations. Lastly, codominants DMIs resolve much faster when recombination is rare; however this almost never leads to hybrid speciation. Thus, in the codominant case, only the slowest resolving cases, in which linkage disequilibrium is broken, are likely to generate hybrid species. The time to hybrid speciation tends to scale with the size of the hybrid population. Therefore, the short time to reproductive isolation reported in the empirical examples is reflecting the size of the hybrid population and not necessarily an inherent property of hybrid speciation. Lastly, the time to resolution of the DMIs (A_1 or B_1 is lost as well as A_2 or B_2) does not depend on the initial proportion the two parental population contributed. Therefore, even if longitudinal data were available, these would not be informative on the demographic history of the hybrid population.

Population size and selection In our model, we consider populations of constant size. If one relaxes this assumption (i.e. switching from soft selection to hard selection), one would expect hybrid speciation to be more rare for at least two reasons. First, selection against the different derived alleles in the early purging phase is stronger; indeed with soft selection the effect of a mutation is weighted by the mean fitness of the population. Therefore, in maladapted populations, the effect of deleterious mutations is slightly dampened. Second, the expected decrease in population size that is associated with the purging phase increases the impact of drift, which means that reciprocal sorting is less likely even in favorable genetic architectures (as selection is not strong enough to counteract its effect). Lastly, even if the DMIs are resolved in opposite direction, the different derived alleles will be at low frequency when their interacting partner is lost, and therefore far more likely to be lost by drift subsequently. Overall, this implies that hybrid speciation via reciprocal sorting is on average less likely than illustrated here. Furthermore, this kind of contact between two diverged populations (or species) is usually geographically restricted, and therefore happens for small populations. However, this apparent rarity of hybrid speciation can be counteracted by the frequent formation of hybrid populations; this could suggest that the reported cases of homoploid speciation, may simply reflect a geographical distribution conducive to the formation and isolation of hybrid population. The Italian Sparrow seems to fit this scenario remarkably well (Runemark et al., 2018). Lastly, a (Lamichhaney et al., 2018) has recently provided an example, where the number of founding individuals of the new species is $N = 3$. The search for signs of hybrid speciation in very large populations, for example yeast, would be an exciting avenue for hybrid speciation research in the future.

5 Conclusion

The probability of hybrid speciation is subject to continuing debate (Schumer et al., 2015; Nieto Feliner et al., 2017; Schumer et al., 2018). The reciprocal sorting of parental incompatibilities has been proposed as one credible mechanism to achieve hybrid speciation. Our work legitimates the existing disagreement by demonstrating that the hybrid speciation probability via reciprocal parental incompatibility sorting is highly variable and dependent on the genetic architecture and the dominance type of the involved incompatibilities. Specifically, the genetic architecture determines not only whether hybrid speciation is achievable or not, but also whether equal or unequal initial proportions of the parental populations are favorable for hybrid speciation. In addition, we show that across all studied scenarios, intermediate recombination rates maximize the likelihood of reciprocal sorting; i.e., interactions on the same chromosome are favorable for hybrid speciation. Altogether, our work enables the important conclusion that in nature, hybrid speciation via reciprocal sorting of incompatibilities should indeed be rare; at the same time however, it can become almost deterministic (and, thus, repeatable) under optimal genetic and demographic circumstances.

bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

R. Abbott, D. Albach, S. Ansell, J. Arntzen, S. Baird, N. Bierne, J. Boughman, A. Brelsford, C. Buerkle, R. Buggs, et al. Hybridization and speciation. *Journal of Evolutionary Biology*, 26(2):229–246, 2013.

M. Arnold, M. Bulger, J. Burke, A. Hempel, and J. Williams. Natural hybridization: how low can you go and still be important? *Ecology*, 80(2):371–381, 1999.

C. Bank, R. Bürger, and J. Hermisson. The Limits to Parapatric Speciation: Dobzhansky–Muller Incompatibilities in a Continent–Island Model. *Genetics*, 191(3):845–863, 2012.

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

N. Barton and B. Bengtsson. The barrier to genetic exchange between hybridising populations. *Heredity*, 57(3):357–376, 1986.

W. Bateson. Heredity and variation in modern lights. *Darwin and modern science*, pages 85–101, 1909.

C. Buerkle, R. Morris, M. Asmussen, and L. Rieseberg. The likelihood of homoploid hybrid speciation. *Heredity*, 84(4):441–451, 2000.

J. Coyne and H. Orr. *Speciation*. Sinauer Associates Sunderland, MA, 2004.

T. Dobzhansky. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics*, 21(2):113, 1936.

J. Feder, S. Egan, and P. Nosil. The genomics of speciation-with-gene-flow. *Trends in Genetics*, 28(7):342–350, 2012.

S. Gavrilets. Hybrid zones with Dobzhansky-type epistatic selection. *Evolution; international journal of organic evolution*, 51(4):1027–1035, 1997.

B. Gross. Genetic and phenotypic divergence of homoploid hybrid species from parental species. *Heredity*, 108(3):157, 2012.

J. B. S. Haldane. A mathematical theory of natural and artificial selection, part v: selection and mutation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 23, pages 838–844. Cambridge Univ Press, 1927.

P. Hedrick. Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, 22(18):4606–4618, 2013.

J. Hermansen, S. Sæther, T. Elgvin, T. Borge, E. Hjelle, and G. Sætre. Hybrid speciation in sparrows I: Phenotypic intermediacy, genetic admixture and barriers to gene flow. *Molecular Ecology*, 20(18):3812–3822, 2011.

I. Höllinger and J. Hermisson. Bounds to parapatric speciation: A Dobzhansky–Muller incompatibility model involving autosomes, X chromosomes, and mitochondria. *Evolution*, 71(5): 1366–1380, 2017.

J. Kang, M. Schartl, R. Walter, and A. Meyer. Comprehensive phylogenetic analysis of all species of swordtails and platies (Pisces: *Genus Xiphophorus*) uncovers a hybrid origin of a swordtail fish, *Xiphophorus monticolus*, and demonstrates that the sexually selected sword originated in the ancestral lineage of the genus, but was lost again secondarily. *BMC Evolutionary Biology*, 13(1), 2013.

bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

M. Kimura. On the probability of fixation of mutant genes in a population. *Genetics*, 47(6): 713, 1962.

S. Lamichhaney, F. Han, M. Webster, L. Andersson, B. Grant, and P. Grant. Rapid hybrid speciation in Darwin's finches. *Science*, 359(6372):224–228, jan 2018.

P. Larsen, M. Marchan-Rivadeneira, and R. Baker. Natural hybridization generates mammalian lineage with species characteristics. *Proceedings of the National Academy of Sciences*, 107 (25):11447–11452, 2010.

J. Mallet. Hybrid speciation. *Nature*, 446(7133):279–283, 2007.

Je. Mavárez, C. Salazar, E. Bermingham, C. Salcedo, C. Jiggins, and M. Linares. Speciation by hybridization in heliconius butterflies. *Nature*, 441(7095):868–871, 2006.

H. Muller. Isolating mechanisms, evolution and temperature. In *Biol. Symp*, volume 6, pages 71–125, 1942.

G. Nieto Feliner, I. Álvarez, J. Fuertes-Aguilar, M. Heuertz, I. Marques, F. Moharrek, R. Piñeiro, R. Riina, J. Rosselló, P. Soltis, and I. Villa-Machío. Is homoploid hybrid speciation that rare? An empiricist's view. *Heredity*, 118(6):513–516, 2017.

H. Orr. A mathematical model of Haldane's rule. *Evolution*, 47(5):1606–1611, 1993.

H. Orr. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics*, 139(4):1805–1813, 1995.

D. Presgraves. A fine-scale genetic analysis of hybrid incompatibilities in *Drosophila*. *Genetics*, 163(3):955–972, 2003.

L. Rieseberg. Hybrid Origins of Plant Species. *Annual Review of Ecology and Systematics*, 28 (2):359–389, 1997.

A. Runemark, C. Trier, F. Eroukhmanoff, J. Hermansen, M. Matschiner, M. Ravinet, T. Elgvin, and G. Sætre. Variation and constraints in hybrid genome formation. *Nature Ecology & Evolution*, 2018.

M. Schumer, G. Rosenthal, and P. Andolfatto. How common is homoploid hybrid speciation? *Evolution*, 68(6):1553–1560, 2014.

M. Schumer, R. Cui, G. Rosenthal, and P. Andolfatto. Reproductive isolation of hybrid populations driven by genetic incompatibilities. *PLoS Genet*, 11(3):e1005041, 2015.

M. Schumer, G. Rosenthal, and P. Andolfatto. What do we mean when we talk about hybrid speciation? *Heredity*, pages 11–31, 2018.

D. Schwarz, B. Matta, N. Shakir-Botteri, and B. McPheron. Host shift to an invasive plant triggers rapid animal hybrid speciation. *Nature*, 436(7050):546–549, 2005.

O. Seehausen. Conditions when hybridization might predispose populations for adaptive radiation. *Journal of Evolutionary Biology*, 26(2):279–281, 2013.

Ole Seehausen, Gaku Takimoto, Denis Roy, and Jukka Jokela. Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, 17(1):30–44, 2008.

M. Servedio and M. Noor. The role of reinforcement in speciation: Theory and data. *Annual Review of Ecology, Evolution, and Systematics*, 34(Noor 1999):339–364, 2003.

bioRxiv preprint doi: https://doi.org/10.1101/266254; this version posted February 15, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

M. Servedio, J. Hermisson, and G. Doorn. Hybridization may rarely promote speciation. *Journal of evolutionary biology*, 26(2):282–285, 2013.

M. Turelli and H. Orr. Dominance, epistasis and the genetics of postzygotic isolation. *Genetics*, 154(4):1663–79, 2000.

D. Turissini, J. McGirr, S. Patel, J. David, and D. Matute. The rate of evolution of postmating-prezygotic reproductive isolation in Drosophila. *Molecular Biology and Evolution*, 35(2):312–334, 2017.

S. Via. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367 (1587):451–460, 2012.

S. Yakimowski and L. Rieseberg. The role of homoploid hybridization in evolution: A century of studies synthesizing genetics and ecology. *American Journal of Botany*, 101(8):1247–1258, 2014.

S. Yeaman. Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences*, 110(19):E1743–E1751, 2013.