MOLECULAR ECOLOGY

Natural selection and the distribution of chromosomal inversion lengths

Journal:	*Molecular Ecology*
Manuscript ID	MEC-21-0527
Manuscript Type:	Original Article
Date Submitted by the Author:	03-May-2021
Complete List of Authors:	Connallon, Tim; Monash University Olito, Colin; Lund University
Keywords:	Adaptation, structural variation, chromosomal rearrangements, inversions, Molecular Evolution, Population Genetics - Theoretical

Natural selection and the distribution of chromosomal inversion lengths

Tim Connallon[1*] and Colin Olito[2]

[1]School of Biological Sciences and Centre for Geometric Biology, Monash University, Clayton, Victoria 3800, Australia

[2]Department of Biology, Section for Evolutionary Ecology, Lund University, Lund 223 62, Sweden

*Corresponding author: tim.connallon@monash.edu

KEYWORDS: adaptation, structural variation, chromosomal rearrangements, inversions

ABSTRACT

Chromosomal inversions contribute substantially to genome evolution, yet the processes governing their fixation remain poorly understood. Theory suggests that a readily measurable property of inversions—their length—can potentially affect their evolutionary fates. Emerging data on the lengths of fixed inversions may therefore provide clues to the evolutionary processes underlying inversion substitutions. However, formal theoretical predictions for the distribution of inversion lengths remain incomplete, making empirical patterns difficult to interpret. We model the relation between inversion length and fixation probability for neutral inversions, underdominant inversions, directly beneficial inversions (*e.g.*, those with favorable breakpoints), and indirectly beneficial inversions that capture locally adapted alleles at migration-selection balance and suppress recombination between them. We also consider how segregating deleterious mutations affect the lengths of inversion substitutions. When neutral, underdominant, or directly beneficial, small inversions contribute the most to genome divergence. In contrast, inversions of intermediate length contribute disproportionately to divergence under the local adaptation scenario. These predictions are robust to effects of deleterious genetic variation, which shift inversion length distributions towards smaller sizes without fundamentally altering their qualitative forms. We also find that the four categories of inversion substitution differ in sensitivity to effective population size (N_e). The lengths of neutral inversion substitutions increase, while underdominant substitutions decrease, with N_e; the lengths of directly and indirectly beneficial inversion substitutions are relatively insensitive to N_e. Our models clarify how inversion length distributions relate to processes of inversion fixation, providing a platform for testing how natural selection shapes the evolution of genome structure.

INTRODUCTION

Since the 1960s, empirical population genetics has overwhelmingly focused on protein and DNA sequence diversity as opposed to variation in genome structure (Charlesworth and Charlesworth 2017). This focus reflects the prevailing sequencing technology of the last half century, which was well-suited for characterizing non-repetitive sequence diversity, yet poorly equipped for identifying structural variants such as chromosomal inversions and gene copy number variants (Wellenreuther et al. 2019). Yet prior to Lewontin and Hubby (1966), most studies of genome evolution fell within the domain of classical cytogenetics (White 1973), which firmly established the evolutionary importance of inversions in recombination suppression (Sturtevant and Beadle 1936; Kirkpatrick 2010), rapid adaptation to environmental changes (Dobzhansky 1970), and the maintenance of genetic polymorphisms and species barriers (Krimbas and Powell 1992; Hoffmann and Rieseberg 2008; Wellenreuther and Bernatchez 2018; Kapun and Flatt 2019; Fuller et al. 2019; Faria et al. 2019).

Recent advances in genome sequencing technology and alignment have made the identification of inversions increasingly feasible (Corbett-Detig et al. 2012; Chakraborty et al. 2018), drawing inversions back into the spotlight of population genetics (Wellenreuther et al. 2019). With increased scope for identifying structural variants, the remaining challenge is to identify the evolutionary processes that have shaped inversion evolution. A long tradition of theoretical research on inversion evolution has identified several possible scenarios for the evolutionary spread of inversions (Kirkpatrick and Barton 2006), including genetic drift of neutral inversions (Kimura and Ohta 1970), fixation of underdominant inversions by drift and selection (*e.g.*, meiosis-disrupting inversions; Lande 1979), direct positive selection on inversions with beneficial breakpoint effects (Krimbas and Powell 1992; Corbett-Detig 2016), and indirect

positive selection on inversions capturing beneficial allele combinations and suppressing recombination between them (Charlesworth and Charlesworth 1973; Kirkpatrick and Barton 2006). Whether and how each scenario contributes to inversion evolution remains an open question.

Fixed inversions, which represent our focus here, play prominent roles in genome divergence and species isolation (Kirkpatrick 2010; Hoffmann and Rieseberg 2008; Cheng and Kirkpatrick 2009), yet the evolutionary mechanisms driving inversion fixation are particularly challenging to identify. Most inversions have been fixed in the distant past, which precludes direct estimation of the forms of selection that drove inversion fixation. Likewise, there are few straightforward approaches for indirectly inferring selection on inversion substitutions. These methodological hurdles stand in contrast to studies of selection on protein sequences, in which long-term patterns of DNA substitution provide some of the most compelling evidence for genome divergence by natural selection (*e.g.*, McDonald and Kreitman 1991; Kern and Hahn 2018). Such hurdles may partially explain why there has been such a strong traditional focus on contemporary inversion polymorphisms—which *are* amenable to fitness assays—in empirical studies of selection on inversions (see Dobzhansky 1970; Lewontin 1974; Krimbas and Powell 1992; Wellenreuther and Bernatchez 2018; Mérot et al. 2020).

One readily measurable property of inversions—their length—may carry information about selection during inversion fixation (Krimbas and Powell 1992), as both the direction and strength of selection on inversions can vary predictably with their lengths (Nei et al. 1967; Van Valen and Levins 1968; Cheng and Kirkpatrick 2019). The length distributions of common inversions (*i.e.*, high-frequency polymorphic and fixed inversions), which have been characterized in several animal taxa (*e.g.*, Van Valen and Levins 1968; Martin 1969; Krimbas and Powell 1992;

Cáceres et al. 1997; Cheng and Kirkpatrick 2019), may therefore provide clues about the processes promoting inversion fixation (Van Valen and Levins 1968; Santos 1986; Cáceres et al. 1997; Cheng and Kirkpatrick 2019). However, theoretical predictions regarding the relation between inversion length and fixation probability remain incomplete, which complicates the interpretation of inversion length data. On the one hand, large inversions could have elevated probabilities of fixation because they are more likely to capture and maintain beneficial allele combinations (Van Valen and Levins 1968; Cheng and Kirkpatrick 2019). Large inversions that happen to be free of deleterious mutations are also expected to fix more readily than small, mutation-free inversions (Nei et al. 1967; Ohta and Kojima 1968; Kimura and Ohta 1970; Cook and Nassar 1972). On the other hand, large inversions are more likely to carry deleterious mutations (Nei et al. 1967; Santos 1986; Connallon et al. 2018) and disrupt meiosis when heterozygous with wild-type chromosomes (Lande 1979; Cheng and Kirkpatrick 2019), which might strengthen purifying selection against large inversions and thereby hinder their spread.

To clarify how the length of an inversion influences its evolutionary fate, we have modeled the fixation probabilities of four categories of inversions: (*i*) neutral inversions; (*ii*) underdominant inversions; (*iii*) inversions that are directly beneficial due to their breakpoints or position effects on gene expression; and (*iv*) inversions that are indirectly favored because they capture beneficial allele combinations and suppress recombination between them. For the last category, we focus on inversions that capture locally adaptive alleles maintained at migration-selection balance, and are driven towards fixation by local positive selection (Kirkpatrick and Barton 2006; Charlesworth and Barton 2018). Several previous models have considered the evolutionary dynamics of one or more of these inversion categories (*i.e.*, Nei et al. 1967; Kimura and Ohta 1970; Lande 1979; Krimbas and Powell 1992; Kirkpatrick and Barton 2006; Charlesworth and Barton 2018), and

some have considered the effects of inversion length on the strength or direction of selection (Van Valen and Levins 1968; Santos 1986; Cheng and Kirkpatrick 2019). Nevertheless, we lack general, quantitative predictions about the relation between inversion length and fixation probability for each category of inversion substitution. Building upon prior theory, we develop explicit, mechanistic models of inversion fixation that depend on the interaction between inversion length, deleterious genetic variation, and direct and indirect selection on the inversion itself. We characterize the relation between inversion length and fixation probability for each scenario of inversion selection, in both the presence and the absence of standing genetic variation for deleterious alleles maintained at mutation-selection balance. We close by reviewing empirical studies of inversion length distributions, reinterpreting them in reference to our models, and outlining future studies that can potentially discriminate between alternative evolutionary scenarios of inversion substitution.

MODELS AND RESULTS

Preliminary comments and overview of the models

Our goal is to predict how different forms of selection on chromosomal inversions affect the distributions of inversions that ultimately approach fixation. We focus on the following four idealized inversion categories, each defined by their fitness effects in the absence of linked deleterious mutations:

1. ***Neutral inversions*** do not affect fitness, per se, though (as we emphasize further below) such inversions can respond indirectly to selection against deleterious mutations within the population (see Nei et al. 1967).

2. ***Underdominant inversions*** reduce fitness by disrupting meiosis when heterozygous with wild-type (non-inverted) chromosomes, but are otherwise as fit as the wild-type karyotype (Lande 1979).
3. ***Directly beneficial inversions*** are intrinsically beneficial because their breakpoints, or position effects on gene expression, increase fitness (see Krimbas and Powell 1992; Corbett-Detig 2016).
4. ***Indirectly beneficial inversions*** are favored because they capture a beneficial combination of alleles and suppress recombination between them. We focus on the influential scenario where an inversion captures locally adaptive alleles maintained at migration-selection balance (Kirkpatrick and Barton 2006; Charlesworth and Barton 2018). Selection tends to drive such inversions towards fixation within the local populations where they are favored (Kirkpatrick and Barton 2006).

models rely on four important simplifying assumptions. First, with the exception of underdominant inversions, we focus on haploid, sexual populations. Our main conclusions apply to diploids provided the genetic variation that mediates selection on inversions is expressed in heterozygotes (see Charlesworth and Barton 2018; Olito and Abbott, 2020), which appears likely given that most mutations are not completely recessive (Charlesworth and Hughes 2000; Charlesworth 2015). Second, we assume that recombination is completely suppressed between inversion and wild-type chromosomes within the chromosomal region that the inversion spans. In reality, genetic exchange can occur via double crossovers (Krimbas and Powell 1992) and gene conversion (Korunes and Noor 2019), though such exchanges are expected to be severely reduced relative to the normal rates of exchange between chromosomes with the same orientations. Third, we assume that the mutation rate to new inversion variants is sufficiently low that the evolutionary

fate of each new inversion is independent of others (*i.e.*, "strong selection, weak mutation"; Gillespie 1991). Our results therefore exclude scenarios involving multiple, simultaneously segregating inversions (Koury 2018) or balanced inversion polymorphisms (*e.g.*, Haldane 1957; Charlesworth 1974; Kirkpatrick and Barton 2006; Berdan et al. 2021). Finally, we assume that deleterious mutations are maintained at mutation-selection balance equilibrium, which requires effectively strong purifying selection against deleterious alleles.

We begin by presenting fixation probabilities and inversion length distributions for the idealized case where deleterious mutation rates per chromosome arm are sufficiently small that they can be neglected. We then extend these baseline models by exploring how deleterious genetic variation maintained at mutation-selection balance affects the inversion dynamics. For reasons of analytical tractability, we focus on an instructive case in which the fitness effects of deleterious mutations are constant across loci and combine multiplicatively. The approach is widely used in multi-locus models of mutation-selection balance (*e.g.*: Haigh 1978; Gordo and Charlesworth 2000; Johnson and Barton 2002; Bachtrog 2008), though it is obviously a simplification of biological reality (Eyre-Walker and Keightley 2007). We relax the assumption of constant fitness effects in the Supplementary Material, and show that modest variation in the fitness effects of deleterious mutations amplifies the consequences of deleterious variation for inversion dynamics that we highlight the main text (see Appendix C of the Supplementary Material).

In the following sections, we present our most important analytical results along with stochastic Wright-Fisher simulations designed to test our approximations and, in some cases, to relax assumptions made for analytical tractability. We confine most of the mathematical and simulation details to the Supplementary Material (see Appendix A-D), and highlight key differences between the analytical and simulation models where they are relevant. Given

considerations of computational time and the large number of simulation replicates that are required to estimate fixation probabilities, our simulations do not explicitly track individual deleterious mutations. They instead incorporate effects of deleterious mutations on inversion dynamics by modelling temporal change in selection coefficients for inversions under deterministic mutation-selection balance (see Appendix B-D in the Supplementary Material).

Inversion substitutions in the absence of deleterious mutational variation

In the absence of deleterious alleles in the population, the fixation probabilities of new inversions depend solely on their direct or indirect fitness effects. Fixation probabilities for neutral, directly beneficial, and underdominant inversions follow from classical population genetics theory (Haldane 1927; Kimura 1962; Lande 1979). Specifically, a newly arising neutral inversion in a haploid population of size N has a fixation probability of $1/N$, independent of its length. Likewise, there is no obvious reason why direct fitness benefits of inversions (*e.g.*, breakpoint and position effects) should correlate with their lengths (Corbett-Detig 2016). An inversion with a direct fitness benefit of s_b (independent of x) is, therefore, expected to fix with probability $\sim 2s_b$ (provided $1/N << s_b << 1$; Haldane 1927).

Fitness costs of underdominant inversions should increase with their sizes, as large inversions are more likely than small ones to experience crossover events that generate unbalanced meiotic products with major (potentially lethal) duplications or deletions. Assuming that homozygotes for inversion and non-inversion (wild-type) chromosomes are equally fit, and the fitness cost to chromosomal heterozygotes is a linear function of inversion length, then the fixation probability of an underdominant inversion of length x is:

$$\Pr(\text{fix}|x) = \frac{\operatorname{erf}\left(\sqrt{N_e s_u x}\right) - \operatorname{erf}\left(\sqrt{N_e s_u x}(1-2q_0)\right)}{2\operatorname{erf}\left(\sqrt{N_e s_u x}\right)} \quad (1)$$

(which follows from eqs. (1-2) in Lande 1979), where erf(·) refers to the error function, q_0 is the initial frequency of the inversion (we assume that $q_0 = 1/2N$ in a diploid population of size N), N_e is the effective population size, and s_u is the fitness cost of an inversion spanning an entire chromosome arm; $N_e s_u x$ represents the effective strength of selection against the minor karyotype.

Previous models suggest that large inversions are more likely than small inversions to capture locally adaptive alleles at multiple loci, which may generate a fixation bias for relatively large inversions (Cheng and Kirkpatrick 2019), though this scenario has not yet been explicitly modelled. As in previous work (Kirkpatrick and Barton 2006; Charlesworth and Barton 2018), we will consider evolution of an island population receiving migrants from a much larger mainland population. We assume that loci contributing strongly to local adaptation are randomly and independently distributed across the genome and sufficiently sparse that individual chromosome arms are unlikely to carry more than two strongly differentiated loci at migration-selection balance (*i.e.*, loci that are not swamped by migration from the mainland). For chromosome arms with two such loci, inversions are favored when they span the pair of loci and capture the locally adaptive allele at each. Charlesworth and Barton (2018) have shown that the establishment probability of such an inversion will be $\Pr(\text{est.}|r, m, S) \approx 2mr(r+S)^{-1}$, where m is the migration rate, $S = s_1 + s_2$ is the sum of the selection coefficients for the locally adapted loci ($m << s_1, s_2$, with additive effects assumed for simplicity), and r is the ancestral rate of recombination between the loci. Assuming that r between a random pair of locally adapted loci is proportional to their physical distance on the chromosome arm, and that double-crossovers between loci are rare in wild-type

individuals, then the probability that an inversion of length x becomes established (and, thus, eventually approaches fixation, locally) will be:

$$\Pr(\text{est.}|x) \approx 2m\left[2\frac{S}{c}\left(\frac{S}{c}+x\right)\ln\left(\frac{S/c}{S/c+x}\right)+2\frac{S}{c}x+x^2\right]+O(m^2) \quad (2)$$

(see Appendix A of the Supplementary Material), where c is the number of crossovers per chromosome arm, per meiosis (*e.g.*, $c = 1$ in many mammal species; Pardo-Manuel de Villena and Sapienza 2001; Dumont and Payseur 2007).

{FIGURE 1}

Fig. 1A compares the relation between inversion length and fixation probability for each of the four scenarios of selection, with simulations confirming that eqs. (1-2) are both extremely accurate. Following intuition, we see that large inversions are more likely to fix in in the local adaptation scenario, and least likely to fix in the underdominance scenario.

What do these relations between inversion length and fixation probability imply about the more empirically-relevant length distribution of inversion substitutions? The length distribution for substitutions in each inversion category depends on the fixation probabilities as well as the length distribution for *new* inversion mutations entering the population. Incorporating both factors, the probability density function for the lengths of fixed inversions is:

$$g(x) = \frac{\Pr(\text{fix}|x)\, f(x)}{\int_0^1 \Pr(\text{fix}|x)\, f(x)dx} \quad (3)$$

(which follows from previous studies, *e.g.*: Van Valen and Levins 1968; Santos 1986), where $f(x)$ represents the probability density function for new inversions; $\Pr(\text{fix}|x)$ is $1/N$ for the neutral case, $2s_b$ for the direct benefit case, and eqs. (1-2) for underdominant and locally adaptive inversion cases, respectively.

Although the length distribution for new inversions is not known, current data on inversion lengths are consistent with at least two models. First, under a "random break-point model", in which the breakpoints of each new inversion are independent and uniformly distributed across the chromosome arm (as roughly consistent with classical mutagenesis data; see Van Valen and Levins 1968), the lengths of new inversions will follow a triangular distribution with probability density function $f(x) = 2(1 - x)$ (see Federer et al. 1967; Van Valen 1968; York et al. 2007). Alternatively, breakpoint locations for each inversion may be closer together than expected if they are independent due to, for example, breakpoint reuse or clustering in chromosomal regions with repetitive sequences (e.g., Pevzner and Tesler 2003; Peng et al. 2006). In this case, the length distribution for new inversions will have many more small inversions than predicted under the random breakpoint model. A truncated exponential distribution for new inversions (hereafter the "exponential model") qualitatively captures such a scenario and has a probability density function of $f(x) = \lambda e^{-\lambda x}\left(1 - e^{-\lambda}\right)^{-1}$, where λ is a positive constant (Cheng and Kirkpatrick 2019).

Fig. 1B compares the length distributions of inversion substitutions for each of the four scenarios of selection, under the random break-point model for new inversions. Small inversions dominate for neutral, underdominant and directly beneficial inversion substitutions, with all three distributions having a mode near zero. In contrast, indirect selection to maintain locally adaptive allele combinations generates an inversion length distribution with an intermediate mode, which reflects the positive relation between inversion length and fixation probability versus the rarity of very large new inversions entering the population. These predictions are general, and emerge under both the random break-point and exponential length distributions for new inversions (see Appendix A of the Supplementary Material).

Effects of deleterious mutations on inversion substitutions

In addition to the fitness effects outlined above, new inversions entering a population will also capture varying numbers of deleterious mutations, which should modify associations between inversion length and fixation probability in two ways. First, large inversions are more likely than small inversions to initially (*i.e.*, when they first arise) capture one or more deleterious mutations, and these will dampen or overturn any fitness advantages associated with the inversion (Connallon et al. 2018). Second, inversions that happen to be initially free of deleterious mutations have a transient advantage over wild-type karyotypes, which will facilitate their spread within the population (Nei et al. 1967). Below, we show that the balance between these countervailing effects of deleterious mutations play out differently for each of the four inversion categories .

In modelling effects of deleterious mutations on inversion fixation probabilities, we assume that deleterious alleles segregate at mutation-selection balance, with no epistasis or linkage disequilibrium between loci, prior to the introduction of each new inversion into the population. Each deleterious mutation an individual carries reduces their fitness by a fixed amount, s_d (we relax this assumption in Appendix C of the Supplementary Material). Effects of deleterious alleles are assumed to combine multiplicatively with other direct or indirect fitness effects (*e.g.*, beneficial or underdominant fitness effects) to determine the overall fitness effects of inversions. In the following, U_d represents the deleterious mutation rate per chromosome arm, and $U_d x$ is the deleterious mutation rate per chromosomal segment of length x. Our simulations incorporate transient effects of selection via deleterious mutations using the approach of Nei et al. (1967), in which the effective selection coefficient for an inversion declines over time due to mutation accumulation subsequent to its origin (see Appendix B-D of the Supplementary Material).

Neutral inversions. We assume in our analytical model that inversions with no direct or indirect beneficial effects on fitness (*i.e.*, intrinsically neutral) can potentially fix by genetic drift if they are initially free of deleterious mutations (we relax this assumption by simulation; see Appendix D of the Supplementary Material). On the one hand, the probability that a new inversion is mutation free is $f_0 = e^{-\frac{U_d x}{s_d}}$ (*i.e.*, due to Poisson distributed deleterious mutation numbers; Haigh 1978; Orr 2000), which declines with inversion size (x) and the average load of mutations per wild-type chromosome arm (U_d/s_d). On the other hand, rare inversions that *are* mutation-free, particularly large ones, will initially increase in frequency because they are transiently favored over wild-type chromosomes. The fixation probability for such an inversion is equal to its frequency at the time the transient benefit erodes, as genetic drift dominates all subsequent dynamics. Following Nei et al. (1967), the deterministic trajectory of an initially mutation-free inversion is given by:

$$q_t = \frac{q_0 \exp\left(U_d x \frac{1 - e^{-s_d t}}{1 - e^{-s_d}}\right)}{1 - q_0 + q_0 \exp\left(U_d x \frac{1 - e^{-s_d t}}{1 - e^{-s_d}}\right)} = \frac{1}{N} \exp\left(U_d x \frac{1 - e^{-s_d t}}{1 - e^{-s_d}}\right) + O(N^{-2}) \quad (4)$$

(see Appendix B of the Supplementary Material), where q_t is the frequency of the inversion t generations since it first arose ($t \geq 0$), and $q_0 = 1/N$ is its initial frequency. The inversion is, therefore, predicted to increase during the first phase to a frequency of $q^* = N^{-1} \exp\left(U_d x \frac{1}{1-e^{-s_d}}\right) + O(N^{-2})$ and thereafter evolve neutrally until it is either fixed or lost from the population. Following the approach of Maynard Smith (1971; see Orr and Unckless 2014), the expected trajectory of an inversion that initially establishes in the population due to this transient benefit will be approximately q_t/Pr(est.), where Pr(est.) is the establishment probability of the inversion due to the transient benefit (see Appendix B-C of the Supplementary Material).

Using the first order approximation from eq. (4), the combined probability that a new inversion is initially mutation free, that it is initially established in the population because it is transiently beneficial, and that it subsequently drifts to fixation, will be:

$$\Pr(\text{fix}|x) \approx f_0 \Pr(\text{est.}|x) \frac{q^*}{\Pr(\text{est.}|x)} \approx \frac{1}{N} \quad (5),$$

with the final approximation appropriate for small s_d. Eq. (5) is valid when N is large, but will otherwise overestimate fixation probabilities of neutral inversions because it overestimates the effect of transient positive selection on the inversion's initial increase. Although the derivation is heuristic and informal, simulations confirm that it is nevertheless a good prediction for the fates of intrinsically neutral inversions in large populations (Fig. 2), where fixation probabilities of neutral inversions are approximately independent of length. This independence arises because the inversion frequency boost from being (initially) mutation free increases with inversion size, whereas the probability of being (initially) mutation free decreases with size. These factors offset one another in large populations. In small populations, the negative effect of the latter outweighs the positive effect of the former, leading to a negative relation between neutral inversion length and fixation probability (Fig. 2; Fig. D1 of the Supplementary Material). The effect amplified when mutations have a distribution of fitness effects (Figs. C1 of the Supplementary Material).

{FIGURE 2}

Underdominant inversions. Analytical results are no longer possible (at least for us) in models that combine underdominant selection with deleterious genetic variation. We therefore carried out simulations of this scenario and contrasted their results with the fixation probabilities predicted by eq. (1), which neglects deleterious mutations (Figs. 3, S1). We assumed that inversions that initially carry deleterious alleles do not fix. The simulations suggest that eq. (1) remains valid when U_d/s_d is small (Figs. 3, S1), and inversion fixation probabilities are roughly

independent of deleterious genetic variation. The analytical prediction breaks down when U_d/s_d and inversion length are both large, with the direction and magnitude of the effect depending on the population size and the effective strength of overdominant selection against inversions. When $2Ns_u$ is small (Fig. 3, top panels; Fig. S1A), deleterious variation dampens fixation probabilities of large inversions (*i.e.*, relative to eq. (2)), particularly in small populations. When $2Ns_u$ is large (Fig. 3, bottom panels; Fig. S1B-C), deleterious genetic variation inflates the fixation probabilities of large inversions. While underdominant selection leads to a negative relation between inversion length and fixation probability, the relation is not necessarily monotonic over the range of x (Fig. 3, bottom panel).

{FIGURE 3}

Directly beneficial inversions. Probabilities of establishment and, ultimately, fixation of directly beneficial inversions can be modelled using time-dependent branching process theory (Peischl and Kirkpatrick 2012; Kirkpatrick and Peischl 2013; Uecker and Hermisson 2011). Consider a directly beneficial inversion that initially carries deleterious mutations at i loci within the inversion. The inversion has a reasonable probability of becoming fixed when its long-term beneficial effect is greater than the cost of carrying the deleterious alleles. Specifically, the inversion can potentially fix when $i < i_{\max}$, where $i_{\max}$ is the nearest integer less than or equal to $-\ln(1 + s_b)/\ln(1 - s_d)$ (*e.g.*, Johnson and Barton 2002). Adapting the general branching process approach developed in Peischl and Kirkpatrick (2012), we can approximate the fixation probability for an inversion that meets this criterion ($i < i_{\max}$) as:

$$\Pr(\text{fix}|x; i) \approx 2s_i\left(1 + \frac{U_d x}{1 - (1 - s_i)e^{-s_d}}\right) \quad (6),$$

where $s_i = (1 + s_b)(1 - s_d)^i - 1$ (see Appendix B of the Supplementary Material). Accounting for the distribution of deleterious mutations captured by new inversions (*i.e.*, Poisson with parameter $U_d x/s_d$), the total probability of fixation becomes:

$$\Pr(\text{fix}|x) \approx \sum_{i=0}^{i_{\max}} 2s_i \left(1 + \frac{U_d x}{1 - (1 - s_i)e^{-s_d}}\right) e^{-U_d x/s_d} \frac{\left(\frac{U_d x}{s_d}\right)^i}{i!} \quad (7),$$

When $s_b \leq s_d$, eq. (7) simplifies to:

$$\Pr(\text{fix}|x) \approx 2s_b \left(1 + \frac{U_d x}{1 - (1 - s_b)e^{-s_d}}\right) e^{-U_d x/s_d} \quad (8).$$

In cases where $s_b > s_d$, eq. (8) represents the probability that an inversion is initially mutation-free and becomes fixed.

{FIGURE 4}

Simulations confirm that eqs. (7) and (8) are good approximations (respectively) for the fixation probabilities of intrinsically beneficial inversions (Fig. 4; also see Figs. S2-S4 in the Supplementary Material). Deleterious mutations can hitchhike to fixation with inversion substitutions when the benefit of the inversion is greater than the fitness cost of each deleterious allele (*e.g.*, when $s_b = 5s_d$, as shown in lower panels of Fig. 4). Despite the assumed independence between inversion length and magnitude of its direct benefit, fixation probabilities of directly beneficial inversions decline with inversion length. This negative relation occurs because the transient benefit of being (relatively) mutation free no longer offsets the probability that the inversion is not mutation free, leading to a net negative effect of deleterious genetic variation on inversion fixation probabilities (the effect can be seen by contrasting f_0, which is an exponential function of $U_d x/s_d$, with eq. (6) which is roughly linear with $U_d x/s_d$).

Indirectly beneficial inversions. For a given length, the strength of indirect selection for an inversion (*i.e.*, due to its suppression of recombination between loci at migration-selection

363 balance) is a random variable rather than a constant, which limits the applicability of eq. (6) to this
364 inversion category. We can, however, use the approach to obtain an approximate lower bound for
365 the fixation probability of locally adaptive inversions. The lower-bound fixation probability is:

$$\Pr(\text{fix}|x) \approx 2x^2 s_I \left(1 + \frac{U_d x}{1-(1-s_I)e^{-s_d}}\right) e^{-U_d x/s_d} \quad (9).$$

366 where $s_I = m\left[1 + \frac{2S}{x^2 c}\left(\frac{S}{c} + x\right)\ln\left(\frac{S/c}{S/c+x}\right) + \frac{2S}{xc}\right]$ (see Appendix B in the Supplementary Material).
367 Eq. (9) represents probability that an inversion is initially mutation-free and ultimately approaches
368 fixation.

369 As in the case for directly beneficial inversions, there is a net negative effect of deleterious
370 genetic variation on inversion fixation probabilities (Fig. 5; also see Figs. S5-S6 in the
371 Supplementary Material). In the absence of deleterious variation, fixation probabilities of locally
372 adaptive inversions increase monotonically with inversion length. With sufficiently large values
373 of U_d/s_d, deleterious variation leads to a non-monotonic relation between inversion length and
374 fixation probability, with inversions of intermediate length having the highest fixation probabilities
375 (Fig. 5, right-hand panels). Simulations of the proportion of inversions reaching high frequencies
376 (*i.e.*, frequencies greater than 0.9) confirm that eq. (9) provides a reasonable baseline (Fig. 5),
377 though in practice it can substantially underestimate the actual proportion of inversions that reach
378 high frequency, particularly when migration is stronger than the fitness costs of deleterious alleles
379 ($m > s_d$), in which deleterious mutations can hitchhike with successful inversions (Fig. 5, bottom
380 panels).

381 {FIGURE 5}

382 ***Length distributions of inversion substitutions.*** Taking into account effects of deleterious
383 genetic variation and other factors affecting selection on inversions, the length distributions for
384 neutral, underdominant and directly beneficial inversion substitutions are, once again, declining

functions of inversion length with modes near zero, whereas the distribution for indirectly selected, locally adaptive inversions has an intermediate mode (Fig. 6). For directly or indirectly beneficial inversions substitutions, deleterious mutations always shift the distributions of inversion lengths towards smaller values (*cf.* top vs. bottom panels in Fig. 6). The magnitude of the shift towards smaller inversion lengths increases with U_d/s_d.

{FIGURE 6}

DISCUSSION

The length distributions of common and fixed inversions often differ markedly from those of rare and experimentally induced inversions (Van Valen and Levins 1968; Olvera et al. 1979; Cáceres et al. 1997; Corbett-Detig 2016), which implies that the fitness effects and establishment probabilities of inversions covary with their sizes. The work presented here extends prior theory for the length distributions of common and fixed inversions. Previously, Van Valen and Levins (1968) modelled how inversion size affects the probability that inversions are favored, with two-locus epistasis assumed to underlie beneficial fitness effects of inversions. Santos (1986) later modelled the probability that inversions are initially free of deleterious mutations (*i.e.*, f_0, above), which is a declining function of inversion size. Most recently, Cheng and Kirkpatrick (2019) modelled inversion fixation probabilities assuming that the fitness of inversion-bearing genotypes is a linear function of inversion size. We have expanded upon these earlier theoretical studies by explicitly modelling the relation between inversion length and fixation probability for several plausible scenarios of selection for inversions, and addressing the consequences of deleterious genetic variation for inversion substitutions under each scenario.

Three specific predictions emerge from our results. First, in the presence of standing genetic variation for deleterious alleles, very large inversions (*e.g.*, spanning the majority of a chromosome arm) are evolutionarily disfavored because of their high likelihoods of capturing deleterious mutations. Consistent with this prediction, cytological data shows a consistent deficit of very large, common inversions in well-studied *Drosophila* species (*e.g.*, relative to predictions of the random break-point model; Krimbas and Powell 1992). Prior studies have explained this deficit in terms of underdominance or the inefficiency of indirect selection within large inversions (*e.g.*, due to higher double-crossover rates; Martin 1969; Krimbas and Loukas 1980; Krimbas and Powell 1992; Cáceres et al. 1999). Our results show that deleterious mutations provide an alternative explanation for the deficit of large inversions (see Santos 1986 for a similar view), particularly in species like *Drosophila*, which have large chromosome arms and show weak underdominance for paracentric inversions (Krimbas and Powell 1992). Selection against large inversions is expected to be pronounced on chromosome arms with high deleterious mutation rates relative to the fitness costs of deleterious alleles (*i.e.*, where U_d/s_d is large). Mutation accumulation experiments in *Drosophila* suggest a deleterious mutation rate of $U_d \sim 0.1$ per chromosome arm and an average heterozygous fitness cost per deleterious mutation of $s_d \sim 0.02$ (Lynch et al. 1999; Haag-Liautard et al. 2007), which corresponds to the middle of the range of parameters that we have used in our simulations (*i.e.*, $U_d/s_d = 0.05$; Figs. 2-5). Insect species with comparable mutation parameters but smaller chromosomes than *Drosophila* (corresponding to smaller U_d/s_d per arm), are more likely to fix inversions that span large fractions of their chromosome arms—a prediction that has yet to be tested.

Our second prediction is that indirect positive selection on inversions—by way of recombination suppression of multi-locus allele combinations—will generate inversion length

distributions that are qualitatively distinct, and therefore empirically distinguishable, from those of other inversion categories. Indirect selection is the only mechanism that can, by itself, generate a fixation bias towards intermediate-length inversion substitutions, and is therefore the simplest explanation for several documented cases of excess of intermediate-length inversions among common and fixed structural variants (*i.e.*, Chironomidae spp.: Van Valen and Levins 1968; Martin 1969; *Anopheles gambiae*: Pombi et al. 2008; various *Drosophila* spp.: Olvera et al. 1979; Cáceres et al. 1997; Cheng and Kirkpatrick 2019, for the X chromosome). Even so, it remains unclear how common such patterns are among animal or plant taxa. Given the promise of obtaining reliable structural variant calls in deep population sequencing datasets (see Introduction), we should soon be able to carry out conclusive tests for indirect selection on inversions by systematically contrasting the length distributions of fixed inversions (which have been filtered by selection) with very rare ones (which have not; see Messer 2009; Corbett-Detig 2016).

Our third prediction is that the fixation of underdominant, neutral, and directly beneficial inversions should generate qualitatively similar length distributions, in spite of the fundamentally different roles of selection and drift in fixing each of these inversion types. All three scenarios predict a distribution of inversion lengths with a mode near zero (Fig. 4), and a strong bias towards small inversions. Such a pattern has been reported in two sequence-based surveys of inversion fixations between *Drosophila* species (York et al. 2007; Cheng and Kirkpatrick 2019, for autosomes), and in multiple cytological studies of common polymorphic inversions (*i.e.*, in Simuliidae spp.: Van Valen and Levins 1968; for "simple" inversions from *D. obscura* and *D. subobscura*: Krimbas and Loukas 1980; Brehm and Krimbas 1991). In light of our models, these patterns could be consistent with underdominant, neutral or directly beneficial mechanisms of inversion evolution (or a combination of them). However, hope remains in future contrasts of

inversion length distributions among species groups with different long-term effective population sizes (N_e), in which: (1) the underdominance mechanism predicts the fixation of *smaller* inversions in lineages with larger N_e (as implied by eq. [1]); (2) the neutral mechanism predicts the fixation of *larger* inversions in lineages with larger N_e (Fig. 1); and (3) mechanisms of direct selection should produce length distributions that are relatively insensitive to population size (as implied by eqs. [6-8]). Although semi-controlled comparisons of inversion dynamics among lineages that vary for N_e must await further data, they represent promising approaches for deciphering the effect of selection on genome evolution (Gossmann et al. 2012; Lanfear et al. 2014; but see Gillespie 2004), and may prove exceptionally useful in studies of natural selection during genome structure evolution. Furthermore, X-linkage and self-fertilization are both expected to increase fixation probabilities for underdominant inversions (Charlesworth et al. 1987; Charlesworth 1992), which should elevate the sizes and rates of inversion substitution on sex chromosomes (as observed in insects; Cheng and Kirkpatrick 2019) and in plants (where underdominant substitutions are common; Hoffmann and Rieseberg 2008).

Several well-characterized polymorphic inversions appear to be ancient, large, and maintained through long-term balancing selection (reviewed in Wellenreuther et al. 2018). Our models of inversion length distributions should apply to inversion substitutions and polymorphic inversions that are destined for fixation. We have not, however, considered how various scenarios of balancing selection might affect the sizes of inversions maintained as stable polymorphisms, and this topic is an obvious area for future attention. Previous theory suggests that multi-locus overdominant selection can both maintain balanced polymorphisms at two or more loci and favour the spread and maintenance of inversions capturing appropriate combinations of partially linked alleles in the system (Kimura 1956; Charlesworth and Charlesworth 1973; Charlesworth 1974).

While such models can favour intermediate-to-large inversion sizes (Van Valen and Levins 1968; Charlesworth and Charlesworth 1973), a formal analysis of the relation between different forms of balancing selection, deleterious genetic variation, and the sizes of balanced polymorphic inversions would be worth future effort. Recent simulation results suggest that the effects of deleterious mutations on inversion evolutionary dynamics are likely to be complex (and interesting) in contexts involving balancing selection, where feedbacks between the relative frequencies of inverted versus ancestral karyotypes and mutation accumulation on each can govern the evolutionary stability of the balanced polymorphism (Berdan et al. 2021).

ACKNOWLEDGEMENTS

We thank Filip Ruzicka, Kay Hodgins, Akane Uesugi, and five anonymous reviewers for many helpful comments on the manuscript. This work was funded by the Australian Research Council, the School of Biological Sciences at Monash University, and the Wenner-Gren foundation.

DATA ACCESSIBILITY

Simulation code can be found at: https://github.com/colin-olito/inversionSize

AUTHOR CONTRIBUTIONS

TC and CO conceived the study and developed and analyzed the models. TC wrote the manuscript with input from CO.

REFERENCES

Bachtrog D. 2008. The temporal dynamics of processes underlying Y chromosome degeneration. Genetics 179:1513-1525.

Berdan EL, Blanckaert A, Butlin RK, Bank C. 2019. Deleterious mutation accumulation and the long-term fate of chromosomal inversions. *Plos Genetics* 17:e1009411.

Brehm A, Krimbas CB. 1991. Inversion polymorphism in *Drosophila obscura. J Heredity* 82:110-117.

Cáceres M, Barbadilla A, Ruiz A. 1997. Inversion length and breakpoint distribution in the *Drosophila buzzatii* species complex: is inversion length a selected trait? *Evolution* 51:1149-1155.

Cáceres M, Barbadilla A, Ruiz A. 1999. Recombination rate predicts inversion size in Diptera. *Genetics* 153:251-259.

Chakraborty M, VanKuren NW, Zhao R, Zhang X, Kalsow S, Emerson JJ. 2018. Hidden genetic variation shapes the structure of functional elements in Drosophila. *Nat genet.* 50:20-25.

Charlesworth B. 1992. Evolutionary rates in partially self-fertilizing species. *Am Nat.* 140:126-148.

Charlesworth B. 2015. Causes of natural variation in fitness: evidence from studies of Drosophila populations. *Proc Natl Acad Sci USA* 112:1662-1669.

Charlesworth B, Barton NH. 2018. The spread of an inversion with migration and selection. *Genetics* 208:377-382.

Charlesworth B, Charlesworth D. 1973. Selection of new inversions in multilocus genetic systems. *Genet Res.* 21:167-183.

Charlesworth B, Charlesworth D. 2017. Population genetics from 1966 to 2016. *Heredity* 118:2-9.

Charlesworth B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex chromosomes and autosomes. *Am Nat.* 130:113-146.

Charlesworth B, Hughes KA. 2000. The maintenance of genetic variation in life-history traits. In Singh RS, Krimbas CB (eds.) *Evolutionary Genetics from Molecules to Morphology*. Cambridge Univ Press: Cambridge, UK, pp 369–392.

Cheng C, Kirkpatrick M. 2019. Inversions are bigger on the X chromosome. *Mol Ecol*. 28:1238-1245.

Connallon T, Olito C, Dutoit L, Papoli H, Ruzicka F, Yong L. 2018. Local adaptation and the evolution of inversions on sex chromosomes and autosomes. *Phil Trans Roy Soc B* 373:20170423.

Cook RD, Nassar RF. 1972. Probability of ultimate survival of a newly occurred inversion in natural populations. *Theor Appl Genet*. 42:368-370.

Corbett-Detig RB. 2016. Selection on inversion breakpoints favors proximity to pairing sensitive sites in *Drosophila melanogaster. Genetics* 204:259-265.

Corbett-Detig RB, Cardeno C, Langley CH. 2012. Sequence-based detection and breakpoint assembly of polymorphic inversions. *Genetics* 192:131-7.

Dobzhansky T. 1970. *Genetics of the Evolutionary Process*. Columbia University Press, New York.

Dumont BL, Payseur BA. 2007. Evolution of the genomic rate of recombination in mammals. Evolution 62:276-294.

Eyre-Walker A, Keightley PD. 2007. The distributon of fitness effects of new mutations. Nat Rev Genet. 8:610-618.

543 Faria R, Johannesson K, Butlin RK, Westram AM. 2019. Evolving Inversions. *Trends Ecol Evol*.
544 34:239-248.
545 Fuller ZL, Koury SA, Phadnis N, Schaeffer SW. 2019. How chromosomal rearrangements shape
546 adaptation and speciation: Case studies in *Drosophila pseudoobscura* and its sibling species
547 *Drosophila persimilis. Mol Ecol.* 28:1283-1301.
548 Gillespie JH. 1991. *The Causes of Molecular Evolution*. Oxford University Press: Oxford, UK.
549 Gillespie JH. 2004. Why k = 4Nus is silly. In Singh RS, Uyenoyama MK (eds.) *The Evolution of*
550 *Population Biology.* Cambridge University Press: Cambridge, UK, pp. 178-192.
551 Gordo I, Charlesworth B. 2000. The degeneration of asexual haploid populations and the speed of
552 Muller's ratchet. Genetics 154:1379-1387.
553 Gossmann TI, Keightley PD, Eyre-Walker A. 2012. The effect of variation in the effective
554 population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biol Evol.*
555 4:658-667.
556 Haag-Liautard C, Dorris M, Maside X, Macaskill S, Halligan DL, Houle D, Charlesworth B,
557 Keightley PD. 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates
558 in Drosophila. *Nature* 445:82-85.
559 Haig J. 1978. The accumulation of deleterious genes in a population—Muller's ratchet. Theor Pop
560 Biol. 14:251-267.
561 Haldane JBS. 1927. The mathematical theory of natural and artificial selection. V. Selection and
562 mutation. Proc Camb Philos Soc. 23:838–844.
563 Haldane JBS. 1957. The conditions for coadaptation in polymorphism for inversions. J Genet.
564 55:218–225.

565 Hoffmann AA, Rieseberg LH. 2008. Revisiting the impact of inversions in evolution: From
566 population genetic markers to drivers of adaptive shifts and speciation? *Annu Rev Ecol Evol*
567 *Syst*. 39:21-42.
568 Johnson T, Barton NH. 2002. The effect of deleterious alleles on adaptation in asexual populations.
569 Genetics 162:395-411.
570 Kapun M, Flatt T. 2019. The adaptive significance of chromosomal inversion polymorphisms in
571 *Drosophila melanogaster*. *Mol Ecol.* 28:1263-1282.
572 Kern AD, Hahn MW. 2018. The neutral theory in light of natural selection. *Mol Biol Evol*.
573 35:1366-1371.
574 Kimura M. 1956. A model of a genetic system which leads to closer linkage by natural selection.
575 *Evolution* 10:278-287.
576 Kimura M. 1962. On the probability of fixation of mutant genes in a population. *Genetics* 47:713-
577 719.
578 Kimura M, Ohta T. 1970. Probability of fixation of a mutant gene in a finite population when
579 selective advantage decreases with time. *Genetics* 65:525-534.
580 Kirkpatrick M. 2010. How and why chromosome inversions evolve. *PLoS Biology* 8:e1000501.
581 Kirkpatrick M, Barton N. 2006. Chromosome inversions, local adaptation, and speciation.
582 *Genetics* 173:419-434.
583 Kirkpatrick M, Peischl S. 2013. Evolutionary rescue by beneficial mutations in environments that
584 change in space and time. *Phil Trans Roy Soc B* 368:20120082.
585 Korunes KL, Noor MAF. 2019. Pervasive gene conversion in chromosomal inversion
586 heterozygotes. *Mol Ecol*. 28:1302-1315.
587 Koury SA. 2018. Meiotic drive and survival probability of newly inverted chromosomes. *bioRxiv*

588 Krimbas CB, Loukas M. 1980. The inversion polymorphism of *Drosophila subobscura*. In Hecht
589 MK, Steere WC, Wallace B (eds.) *Evolutionary Biology*, Vol. 12. Plenum Press: New York,
590 pp. 163-234.
591 Krimbas CB, Powell JR. 1992. *Drosophila Inversion Polymorphism*. CRC Press, London.
592 Lande R. 1979. Effective deme sizes during long-term evolution estimated from rates of
593 chromosomal rearrangement. *Evolution* 33:234–251.
594 Lanfear R, Kokko H, Eyre-Walker A. 2014. Population size and the rate of evolution. *Trends Ecol*
595 *Evol*. 29:33-41.
596 Lewontin RC. 1974. *The Genetic Basis of Evolutionary Change*. Columbia University Press: New
597 York.
598 Lewontin RC, Hubby JL. 1966. A molecular approach to the study of genic heterozygosity in
599 natural populations. II. The amount of variation and degree of heterozygosity in natural
600 populations of *Drosophila pseudoobscura*. *Genetics* 54:595–609.
601 Lynch M, Blanchard J, Houle D, Kibota T, Schultz S, Vassilieva L, Willis J. 1999. Perspective:
602 Spontaneous deleterious mutations. *Evolution* 53:645-663.
603 Martin J. 1969. On the origin of inversion polymorphism. *Am Nat*. 103:267-275.
604 Maynard Smith J. 1971. What use is sex? J Theor Biol. 30:319-335.
605 Mérot C, Llaurens V, Normandeau E, Bernatchez L, Wellenreuther M. 2019. Balancing selection
606 via life-history trade-offs maintains an inversion polymorphism in a seaweed fly. *Nature*
607 *Communications* 11:670.
608 McDonald JH, Kreitman M. 1991. Accelerated protein evolution at the Adh locus in Drosophila.
609 *Nature* 351:652–654.

610 Messer PW. 2009. Measuring the rates of spontaneous mutation from deep and large-scale
611 polymorphism data. *Genetics* 182:1219-1232.
612 Nei M, Kojima K, Schaffer H. 1967. Frequency changes of new inversions under mutation-
613 selection equilibria. *Genetics* 57:741-750.
614 Ohta T, Kojima K. 1968. Survival probabilities of new inversions in large populations. *Biometrics*
615 24:501–516.
616 Olito C, Abbott JK. 2020. The evolution of suppressed recombination between sex chromosomes
617 by chromosomal inversions. biorXiv doi: https://doi.org/10.1101/2020.03.23.003558.
618 Olvera O, Powell JR, de la Rosa ME, Salceda VM, Gaso MI, Guzman J, Anderson WW, Levine
619 L. 1979. Population genetics of Mexican Drosophila VI. Cytogenetic aspects of the inversion
620 polymorphism in *Drosophila pseudoobscura. Evolution* 33:381-395.
621 Orr HA. 2000. The rate of adaptation in asexuals. Genetics 155:961-968.
622 Orr HA, Unckless RL. 2014. The population genetics of evolutionary rescue. PLoS Genet.
623 10:e1004551.
624 Pardo-Manuel de Villena F, Sapienza C. 2001. Recombination is proportional to the number of
625 chromosome arms in mammals. Mammalian Genome 12:318-322.
626 Peischl S, Kirkpatrick M. 2012. Establishment of new mutations in changing environments.
627 *Genetics* 191:895-906.
628 Peng, Q., P. A. Pevzner, and G. Tesler, 2006 The fragile breakage versus random breakage models
629 of chromosome evolution. PLoS Comput. Biol. 2: 100–111.
630 Pevzner, P. A. and G. Tesler, 2003 Human and mouse genomic sequences reveal extensive
631 breakpoint reuse in mammalian evolution. PNAS 100: 7672–7677.

Pombi M, Caputo B, Simard F, Di Deco MA, Coluzzi M, della Torre A, Costantini C, Besansky NJ, Petrarca V. 2008. Chromosomal plasticity and evolutionary potential in the malaria vector *Anopheles gambiae* sensu stricto: insights from three decades of rare paracentric inversions. *BMC Evol Biol.* 8:309.

Santos M. 1986. The role of genic selection in the establishment of inversion polymorphism in *Drosophila subobscura. Genetica* 69:35-45.

Sturtevant AH, Beadle GW. 1936. The relations of inversions in the X chromosome of *Drosophila melanogaster* to crossing over and disjunction. *Genetics* 21:554-604.

Uecker H, Hermisson J. 2011. On the fixation process of a beneficial mutation in a variable environment. *Genetics* 188:915-930.

van Valen L, Levins R. 1968. The origins of inversion polymorphisms. *Am Nat*. 102:5–24.

Wellenreuther M, Bernatchez L. 2018. Eco-evolutionary genomics of chromosomal inversions. *Trends Ecol Evol*. 33:427-440.

Wellenreuther M, Mérot C, Berdan E, Bernatchez L. 2019. Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. *Mol Ecol.* 28:1203-1209.

White MJD. 1973 *Animal Cytology and Evolution*. Cambridge University Press, Cambridge, UK.

York TL, Durrett R, Nielsen R. 2007. Dependence of paracentric inversion rate on tract length. *BMC Bioinformatics* 8:115.

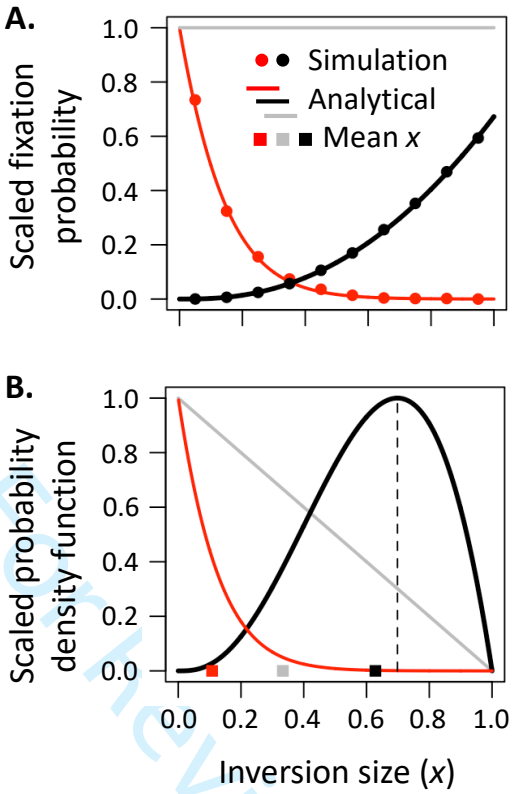


Figure 1. Fixation probabilities and probability density functions for the lengths of inversion substitutions in models neglecting deleterious mutational variation. **Panel A** compares the relation between length and fixation probability for underdominant (red), indirectly beneficial (black), and neutral and directly beneficial inversions (grey). Circles show the proportion of simulation runs that resulted in inversion fixation (in the underdominance model) or inversion frequencies exceeding 90% (in the model of indirect selection via local adaptation). **Panel B** compares the length distributions for the same four types of inversion substitution, with distributions based on eq. (3) and the surrounding text, and assuming a random break-point model of new inversion sizes (*i.e.*, $f(x) = 2(1 - x)$). Squares denote the means for the distributions. The vertical dashed line highlights the mode for the indirect selection model. Each function was scaled relative to its maximum to facilitate comparisons. Results use parameters $m = 0.01$, $s_1 = s_2 = 0.05$, $c = 1$, $N_e s_u = 10$, $N = N_e = 10^4$.

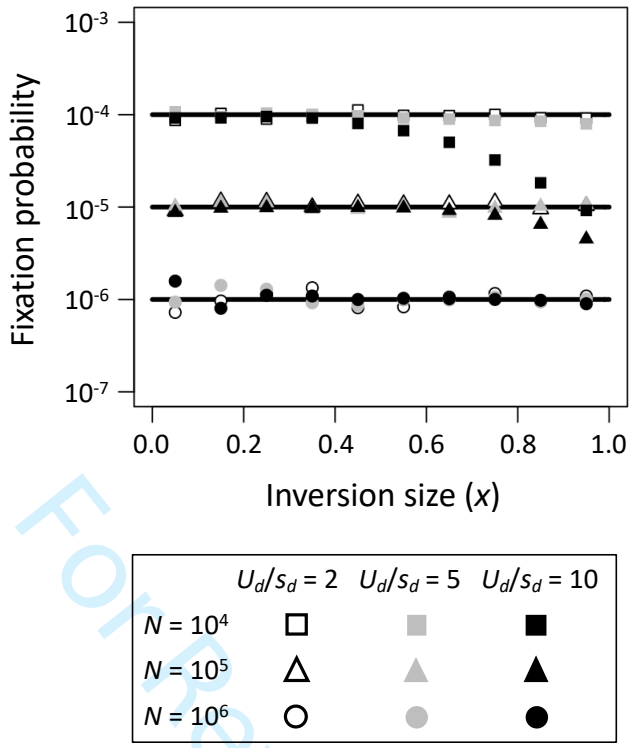
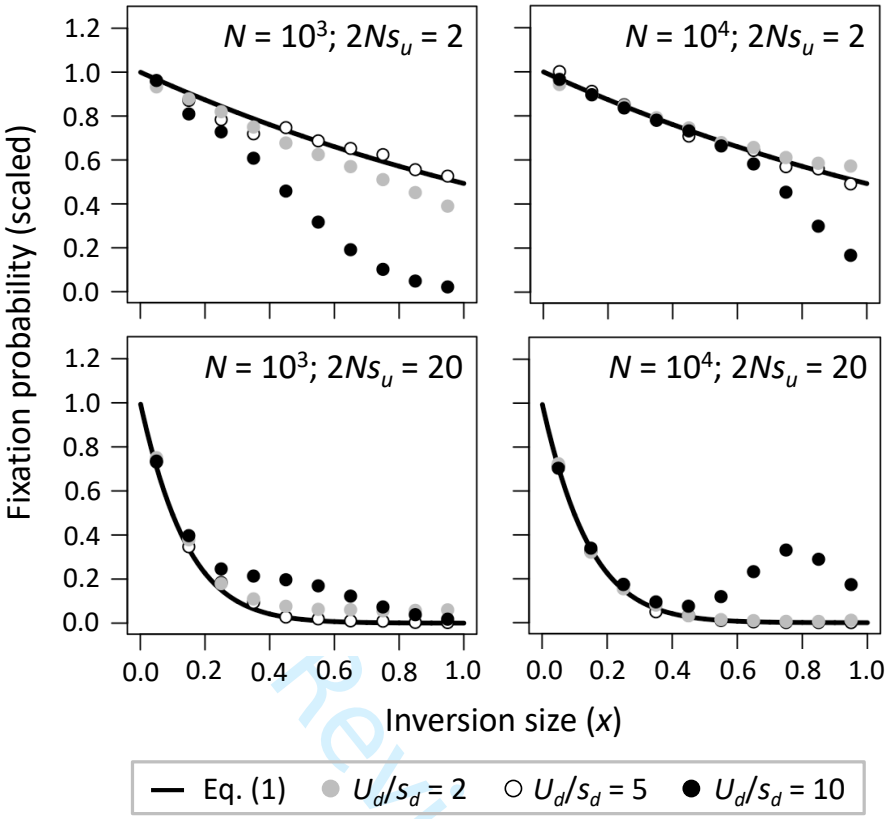


Figure 2. Fixation probabilities for neutral inversions. The solid lines are based on eq. (5), with $N = 10^4$ (top), $N = 10^5$ (middle), and $N = 10^6$ (bottom). Each circle, triangle, or square shows the proportion of $100N$ simulation runs that resulted in inversion fixation (see Appendix D of the Supplementary Material for details of the simulations, and Fig. D1 for additional results). Results use $s_d = 0.01$ with U_d variable. For simulation retults that allow for variation in the fitness effects of deleterious mutations, see Fig. C1 of the Supplementary Material.

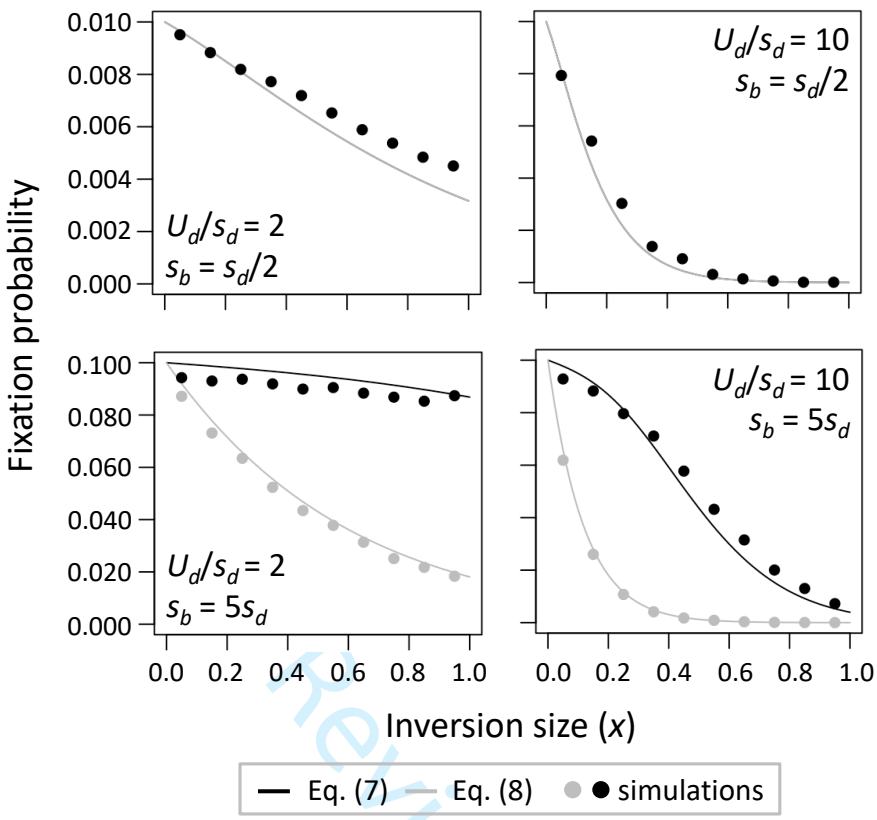


673

674 **Figure 3.** Fixation probabilities for underdominant inversions. Analytical predictions (*i.e.*, eq. (1))
675 ignore effects of deleterious mutations on inversion dynamics. Circles show proportions of
676 simulation runs that result in fixation, with simulations incorporating both underdominant
677 selection and deleterious mutational variation. Each datapoint is based on 10^3N Wright-Fisher
678 simulations (see Appendix D of the Supplementary Material for details of the simulations, and Fig.
679 S1 for additional results).

680

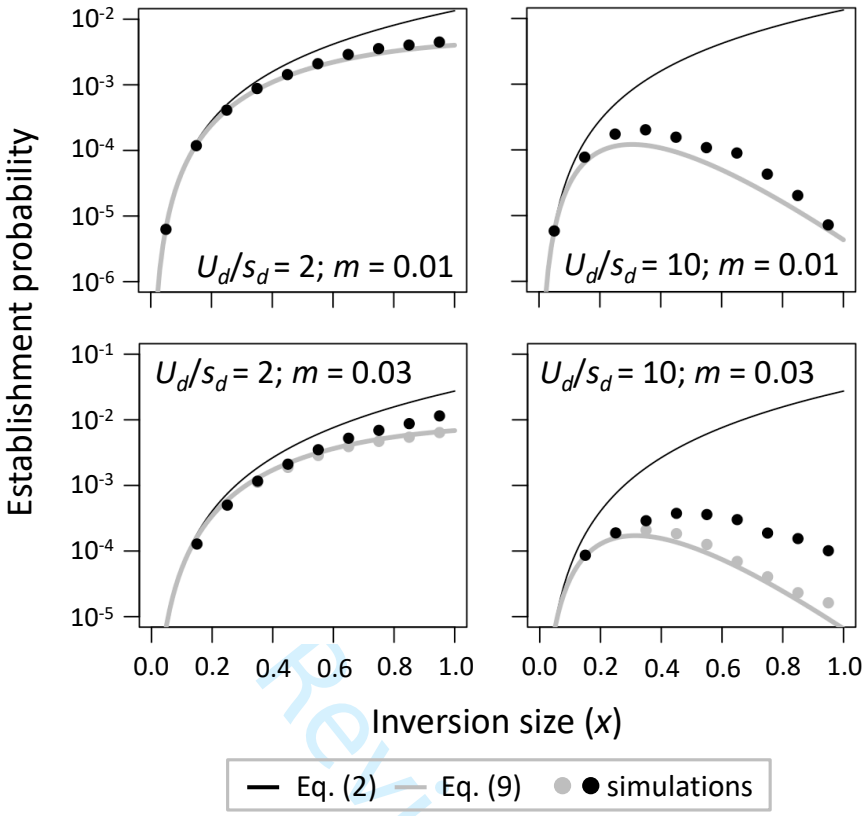
681



682

683 **Figure 4.** Fixation probabilities for directly beneficial inversions. Circles show proportions of
684 simulation runs resulting in fixation (each datapoint based on 10^5 Wright-Fisher simulations; see
685 Appendix D of the Supplementary Material for details; for additional results, see Figs. S2-S4).
686 Black circles show the total probability of inversion fixation, whereas grey circles show the
687 probability that the inversion is initially mutation free and eventually becomes fixed. Results use
688 $s_d = 0.01$, $N = 10^6$, and U_d variable.

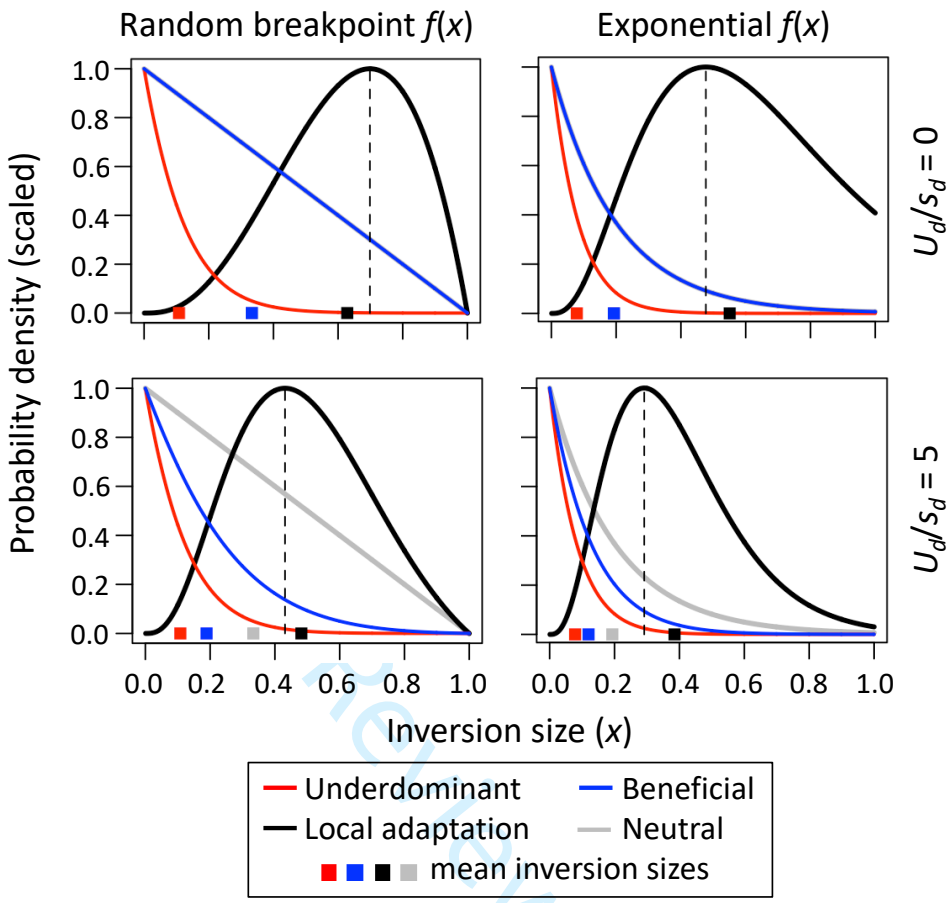
689



690

691 **Figure 5.** Fixation probabilities for indirectly beneficial inversions that capture locally adaptive
692 alleles at migration-selection balance. Black circles show the total probability of inversion
693 establishment, whereas grey circles show the probability that the inversion is initially mutation
694 free and becomes established (each datapoint based on 5×10^5 Wright-Fisher simulations; see
695 Appendix D of the Supplementary Material for details; for additional results, see Figs. S5-S6).
696 Inversions were regarded as established after reaching frequencies above 90%. Results use $s_1 = s_2$
697 $= 0.05$ (top panels), $s_1 = s_2 = 0.15$ (bottom panels), $s_d = 0.01$, $N = 10^6$, and U_d variable.

698



699

700 **Figure 6.** Effects of deleterious mutational variation on probability density functions for the
701 lengths of inversion substitutions. The left-hand panels show probability density functions (*i.e.*,
702 $g(x)$, based on eq. (3)) under the random breakpoint model for new inversion lengths (*i.e.*, $f(x) =$
703 $2(1-x)$). The right-hand panels show probability density functions under the exponential model
704 for new inversion lengths (*i.e.*, $f(x) = \lambda e^{-\lambda x}\left(1-e^{-\lambda}\right)^{-1}$, with $\lambda = 5$ shown). Fixation
705 probabilities are based on eqs. (1), (5), (7), and (9), with parameters $m = s_d = s_b = 0.01$, $s_1 + s_2 =$
706 0.1, $c = 1$, and $N_e s_u = 10$.

707