Unboxing mutations: Connecting mutation types with evolutionary consequences

Emma L. Berdan[1*], Alexandre Blanckaert[2], Tanja Slotte[1], Alexander Suh[3,4], Anja M. Westram[5,6], and Inês Fragata[7*]

1. Department of Ecology, Environment and Plant Sciences, Science for Life Laboratory, Stockholm University, Stockholm, Sweden SE-10691
2. Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI 53706
3. School of Biological Sciences – Organisms and the Environment, University of East Anglia, Norwich, United Kingdom, NR47TJ
4. Department of Organismal Biology – Systematic Biology, Uppsala University, Science for Life Laboratory, Uppsala, Sweden SE-75236
5. IST Austria, Am Campus 1, 3400 Klosterneuburg, Austria
6. Faculty of Biosciences and Aquaculture, Nord University, Bodø, Norway, N-8049
7. cE3c – Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

* - Corresponding author, emma.berdan@gmail.com, emmaberdan.weebly.com; irfragata@fc.ul.pt, @ines_fragata

Abstract

A key step in understanding the genetic basis of different evolutionary outcomes (e.g., adaptation) is to determine the roles played by different mutation types. To do this we must simultaneously consider different mutation types in an evolutionary framework. Here we propose a research framework that directly utilizes the most important characteristics of mutations, their population genetic effects, to determine their relative evolutionary significance. We review known population genetic effects of different mutation types and show how these may be connected to different evolutionary outcomes. We provide examples of how to implement this framework and pinpoint areas where more data, theory and synthesis are needed. Linking experimental and theoretical approaches to examine different mutation types simultaneously is a critical step towards understanding their evolutionary significance.

Keywords: mutation, population genetics, distribution of fitness effects, structural variant, mutation rate

---

Introduction

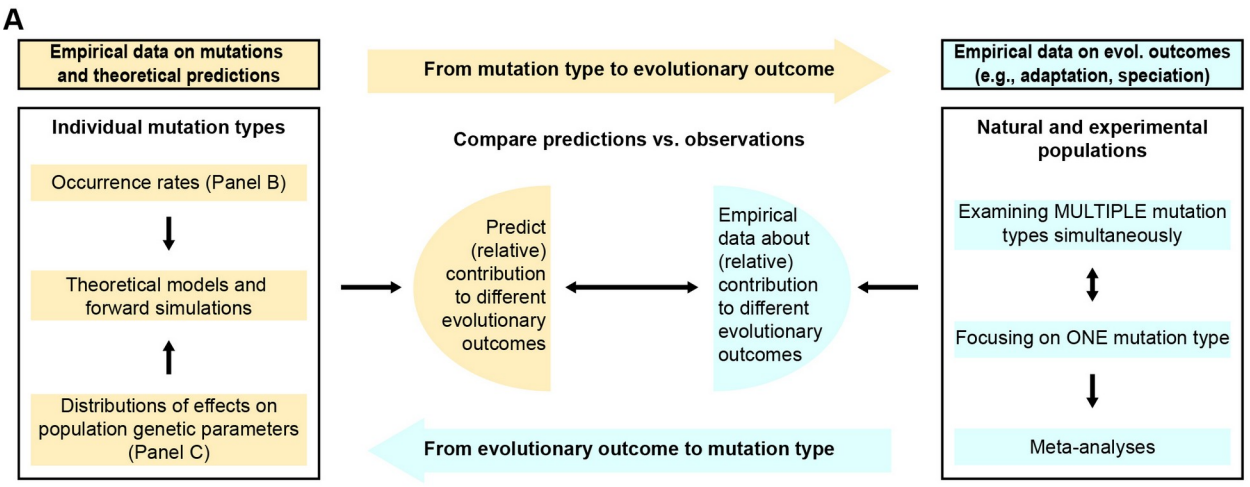
Understanding the genetic underpinnings of adaptation and speciation is a major goal in evolutionary biology. This requires quantifying the number of involved loci and their distribution (e.g., genomic architecture) as well as the type of mutation. There has been a recent focus on types of mutations and detailed discussions of the evolutionary significance of different mutation types abound in the literature (e.g., Choi & Lee, 2020a; Faria, Johannesson, Butlin, & Westram, 2019; Katju & Bergthorsson, 2013). However, most of these reviews, as well as most empirical

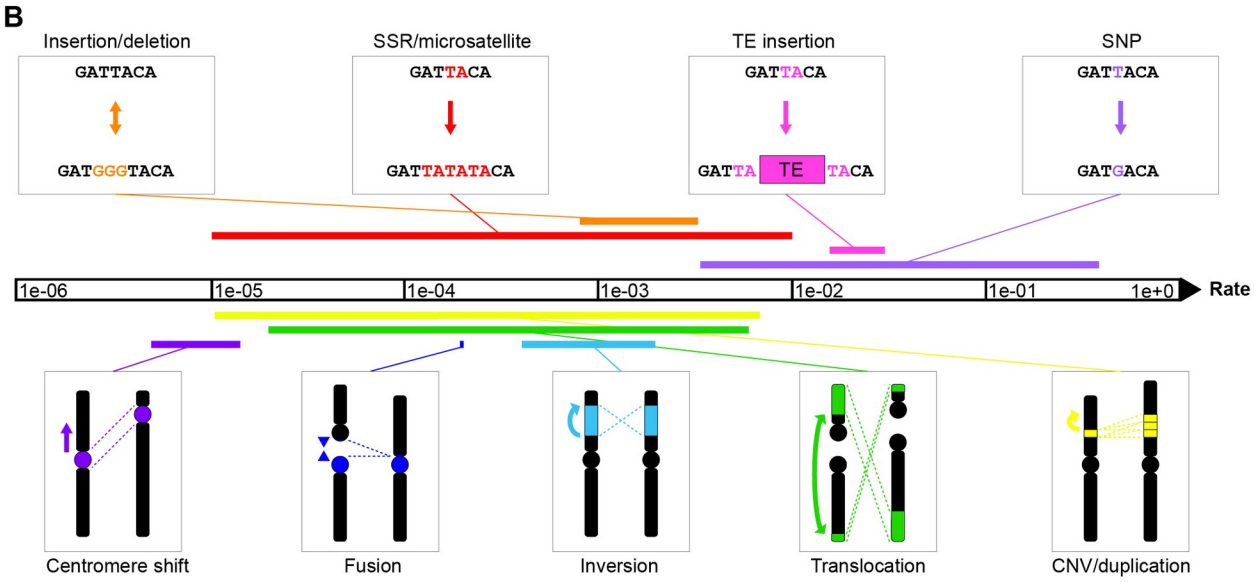
studies, do not examine the full range of mutation types. In order to understand the relative evolutionary significance of different mutation types we must consider them simultaneously in an integrated framework.

From an evolutionary viewpoint, the most important characteristics of a mutation are its occurrence rate and its population genetic effects and how these may influence downstream evolutionary outcomes. Here we propose a research framework that takes advantage of decades of population genetic research (Charlesworth, 2010a; Futuyma, 1986) to directly utilize the population genetic effects of different mutation types to determine their relative evolutionary significance (Figure 1A). The framework combines a forward and a reverse approach. Starting from mutation, the forward approach characterizes how often different types of mutations occur (Figure 1B) and their population genetic effects (Figure 1C) in order to predict their role in different evolutionary outcomes. The reverse approach starts with the evolutionary outcome and, using empirical data, determines the relative contributions of multiple mutation types in a systematic fashion. There is already a wealth of knowledge examining the population genetic effects of different mutations. In contrast, few studies examine multiple types of mutations simultaneously using the reverse approach. Thus, here we concentrate on the forward approach.

To outline the forward approach, we give a non-exhaustive summary of what is known about mutation rates and the population genetic effects of the different mutations, taking advantage of the extensive existing body of work (Choi & Lee, 2020a; Faria et al., 2019; Katju & Bergthorsson, 2013). We stress that quantifying these effects, evaluating differences between mutation types, and determining the distribution of effect sizes will be a critical step forward and

suggest different ways that this can be accomplished. We discuss areas where more empirical data, theory and synthesis are needed for the forward approach, and highlight the importance of applying the reverse approach.

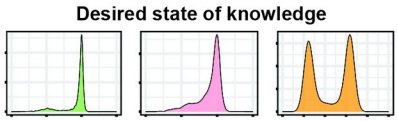




C

Current state of knowledge

Mutation \ Effect	Recombination rate	N_e	Physical distance	Dominance	Distribution of fitness effects
SNP	-	-	-	-	
TE insertion	↓↑-	-	-	-	
CNV/duplication	↓	-	-	↓	
SSR/microsatellite	↑	-	-	-	?
Insertion/deletion	↓	↓	↓↑	↑	?
Translocation	↓	↓	↓↑	-	?
Inversion	↓	↓	↓↑	-	?
Fusion	↓	↓	↑	-	?
Centromere shift	↓↑	↓	-	-	?



Desired state of knowledge

Distribution of effect sizes **for every** population genetic effect **for every** mutation type

Figure 1 - **(A)** Overview schematic illustrating forward and reverse approaches to determine the relative evolutionary significance of different mutation types. Forward approaches start from occurrence rates and population genetic effects of mutations and feed these into theoretical models and simulations to generate

predictions. Reverse approaches start with evolutionary outcomes (ex: adaptation) and examine the roles of different mutations simultaneously or single mutation types at a time, using data obtained either from natural populations or experimental evolution. The results from single mutation studies can be combined in a meta-analysis. The results from the reverse approach can then be compared with the predictions from the forward approach. **(B)** Overview of mutation types and their occurrence rates. Occurrence rates are per genome per generation, come from a wide range of taxa, and are taken from (Beckmann, Estivill, & Antonarakis, 2007; Brumfield, Beerli, Nickerson, & Edwards, 2003; Ducos et al., 2007; Farlow et al., 2015; Feusier et al., 2019; W. Fu, Zhang, Wang, Gu, & Jin, 2010; Gemayel, Vinces, Legendre, & Verstrepen, 2010; Goerner-Potvin & Bourque, 2018; Jarne & Lagoda, 1996; Katju & Bergthorsson, 2013; Maeda, Ohno, Matsunobu, Yoshihara, & Yabe, 1991; Marriage et al., 2009; Marshall, Chueh, Wong, & Choo, 2008; Ossowski et al., 2010; Ramu et al., 2013; Rocchi, Archidiacono, Schempp, Capozzi, & Stanyon, 2012; Schrider, Houle, Lynch, & Hahn, 2013; Sung et al., 2016; Thuillet et al., 2002; Vendrell-Mir et al., 2019; Weng et al., 2019; Yamaguchi & Mukai, 1974). Please note that this is not an exhaustive overview and that actual ranges are likely larger. **(C)** Current and desired state of knowledge regarding the effects of mutation types on population genetic parameters. Current state of knowledge: Arrows indicate an increase (up) or decrease (down) while a dash indicates no effect (known or unknown). For more details of the effects please refer to the *Population Genetic Effects of Mutations: What do we know?* section. Indels and inversions are assumed to be large enough to affect pairing at meiosis while smaller indels and inversions are expected to behave similarly to SNPs. The DFEs are schematics simulated from different statistical distributions and do not reflect real data. Desired state of knowledge: Color and shape of schematic distributions (simulated) do not reflect real data or predictions.

Occurrence Rate

Occurrence rate is a critical parameter when investigating the evolutionary impact of different types of mutations. It can be measured directly, by comparing the number of mutations in gametes, zygotes or offspring (Y.-X. Fu & Huai, 2003) or indirectly, by comparing synonymous (usually presumed neutral) polymorphism data within and between closely related species. Estimates of the occurrence rate vary greatly across taxa and mutation types (Figure 1B). However, different mutations vary with respect to their detectability (for example, **SNPs** are much more likely to be detected with short-read sequencing data than larger insertions or deletions (Ho, Urban, & Mills, 2020)), which can lead to an underestimation of the occurrence rate of different mutations. Thus, one key step in our framework is to 1) develop new methods that allow simultaneous detection of the different mutation types; 2) increase both the number as well as the taxonomic breadth of studies that directly estimate mutation rates and analyse multiple different mutation types within the same taxon.

Population Genetic Effects of Mutations: What do we know?

In the following sections, we discuss the effects of different mutations on population genetic parameters like recombination rate and effective population size. Note that we do not include chromosomal fissions due to lack of information nor whole genome duplications as they have been covered extensively in previous reviews (Fox, Soltis, Soltis, Ashman, & Van de Peer, 2020). Mutations can go beyond DNA sequence changes and affect gene expression through changes in DNA methylation or chromatin state; however, as these are not standard population genetic parameters we discuss them separately in Box 1. To make predictions about evolutionary outcomes, it is important to not just estimate the direction, but also the effect size of the changes in population genetic parameters (see Fig. 1C: current vs. desired state of knowledge). This can be done in two different ways: 1) By directly measuring the effect (e.g., recombination rate) empirically; or 2) By mining information from molecular mechanisms and empirical estimates, and feeding these into theoretical models to derive effects. While we mostly focus on #1 in the main text, we provide an example of #2 (for recombination rate) in Box 2.

Box 1
Beyond DNA sequence alterations

Many mutations affect not only the DNA sequence but also the local state of DNA methylation or histone modifications (e.g., methylation, acetylation), the latter corresponding to the chromatin state of the region. However, the population genetic effects of these changes are often ignored. Changes to the DNA methylation and chromatin state may affect both the regulatory environment of the genes present (potentially altering dominance patterns as well) and the recombination rate. In general, an increase in DNA methylation or

heterochromatinization will often decrease recombination (crossovers are less likely in highly heterochromatic regions (Henderson, 2012)) and gene expression.

DNA methylation levels and chromatin state are two mechanisms that are widely responsible for the regulation of genes, either up-regulating or down-regulating certain regions of the genome (Talbert, Meers, & Henikoff, 2019). Among different mutation types, **transposable elements** (TEs) in particular lead to the alteration of the methylation and chromatin context around themselves. This is because TEs are in a constant arms race with the host, and genomes have evolved multitudes of sequence-specific mechanisms for silencing of new TE insertions via DNA methylation (e.g., CpG dinucleotides) and repressive histone marks (e.g., H3K9me2 and H3K9me3) (Choi & Lee, 2020b; Hollister & Gaut, 2009). These changes in methylation and chromatin state not only affect the mutation itself (and its possible fitness effects) but may spread into adjacent genomic regions, e.g., up to 20 kb away from TE insertions in *Drosophila melanogaster* (Lee & Karpen, 2017) acting as a local DFE modifier. Translocations may also change DNA methylation state; a study on humans found multiple differentially methylated positions with respect to a translocation, 93% of which mapped to the translocation breakpoints (McCartney et al., 2018). Finally, chromosomal fusions have been reported to be associated with larger regions bearing repressive histone marks in mice, potentially leading to a decrease in recombination events (Capilla et al., 2014).

Centromere shifts either happen through a change in chromatin state alone ("neocentromeres"; (Marshall et al., 2008)) or together with the expansion of specific repetitive sequences ("evolutionary new centromeres"; (Rocchi et al., 2012)). Centromeres contain both centromeric chromatin (characterized by the CENP-A histone which is the foundation for the kinetochore) and repressive histone marks (Sullivan & Karpen, 2004). Nevertheless, their effects on population genetics parameters can be expected to be similar to other structural variants. The shift of the centromere along a chromosome will directly reduce the recombination rate in the new centromere-adjacent region (see *Recombination*). A reduction in recombination will increase the rate of accumulation of TEs, spreading DNA methylation and repressive chromatin marks indirectly. This generates a positive feedback loop between the reduction of the local recombination rate, new TE insertions, and change in chromatin state as

previously proposed for regions of low recombination in general (Kent, Uzunović, & Wright, 2017).

More work remains to be done to determine the effect of large structural variants and other mutations on DNA methylation levels and chromatin state. While DNA methylation levels and chromatin state tend to be less permanent than sequence changes, their consequences are far reaching if they also alter the state of the flanking regions. Understanding these effects will help towards building a more unified framework for analyzing the relative role of the various mutation types in evolutionary processes.

Recombination rate and physical distance

The probability that two loci are separated during meiosis is affected by segregation patterns, the physical distance between them (physical linkage), and the per base pair recombination rate. Mutations such as **centromere shifts**, that distort their segregation during female meiosis of heterozygotes (a process known as **centromere drive**), will affect segregation (Malik, 2009). Other mutations such as **fusions** and **translocations** will affect segregation patterns by bringing previously completely (physically) unlinked loci into linkage. Additionally, translocations will also break co-segregation of loci on either side of the breakpoints by moving them to separate chromosomes. Fusions, translocations, **inversions** and large **indels** will all affect physical distance between loci (Smukowski & Noor, 2011).

Mutations may also alter the local recombination rate which can have strong downstream effects on selection and effective population size (see below). Below we summarize the mechanistic ways through which mutations may affect recombination rate and highlight the fact that these

different mechanisms have different population genetic consequences, as further explored and quantified in Box 2.

Recombination in eukaryotes begins with double strand breaks (DSBs) that form during the pairing of the homologs in meiosis and are repaired via two pathways. (1) A crossover event (CO), the outcome of which is visible as a chiasma later in meiosis or (2) the break is repaired as a non-crossover (NCO) event. **Gene conversion** (GC) can occur in both pathways (Hunter, 2015). Direct changes in recombination rate can be due to changes in the pairing process, distribution of recombination events, or pathway taken.

Several mutations affect the alignment and pairing of homologs at the beginning of the recombination process when heterozygous. In inversion heterozygotes, proper synapsis in the inverted region and subsequent crossing over are slightly reduced (Gong, McKim, & Hawley, 2005). A large heterozygous indel will generally form "unpaired DNA loops" preventing COs (Poorman, Moses, Russell, & Cacheiro, 1981). **Copy number variants** (CNVs) can also affect recombination in heterozygotes due to differences in chromosome length, effectively reducing recombination by inhibiting proper pairing (Sjödin & Jakobsson, 2012). Recombination may even be affected outside of the mutated area. For example, COs were suppressed in the regions around large artificial insertions in *C. elegans* (Hammarlund, Davis, Nguyen, Dayton, & Jorgensen, 2005).

Centromeres and their surrounding pericentromeric regions generally reduce recombination by suppressing DSBs and reducing the CO to NCO ratio (Stapley, Feulner, Johnston, Santure, &

Smadja, 2017). As a result, centromere shifts along the chromosome reduce recombination in the new centromere-adjacent region. Conversely, the region of the former inactivated centromere would then be free of centromere-associated recombination reduction. To our knowledge, these aspects of centromere evolution have yet to be appreciated in an evolutionary context.

Mutations may also affect the pathway taken after a DSB (i.e., a CO or an NCO). The presence of fusions changes the rates and distribution of chiasmata (indicative of a CO) in both homo- and heterozygotes in a range of mammals (Dobigny, Britton-Davidian, & Robinson, 2017). For example, in mice (*Mus musculus domesticus*), the number of chiasmata correlates negatively with the number of fusions but the distribution of the chiasmata along the chromosomal arm varies between homozygotes and heterozygotes (Bidau, Giménez, Palmer, & Searle, 2001; Capilla et al., 2014). In inverted regions, DSBs are more likely to be resolved as NCOs; however, the rate of GC is unchanged (Crown, Miller, Sekelsky, & Hawley, 2018; Korunes & Noor, 2019). Conversely, heterozygous translocations do not affect the ratio of COs to NCOs but reduce the rate of GC (Sherizen, Jang, Bhagat, Kato, & McKim, 2005).

Several other mutations can change the recombination landscape on a smaller scale. For example, transposable element (TE) insertions can actively change recombination rates depending on whether or not they attract repressive histone marks (see Box 1; Choi & Lee, 2020b) locally decreasing recombination or contain sequence motifs that turn the region into a recombination hotspot (Kent et al., 2017). Similarly, **simple sequence repeats (SSRs)** can also act as recombination hotspots or recombination repressors (Brandström, Bagshaw, Gemmell, & Ellegren, 2008; Guo, Ling, & Li, 2009; Myers, Bottolo, Freeman, McVean, & Donnelly, 2005).

Recombination may occur normally but lead to the creation of unbalanced gametes in heterozygotes only. In inversions, crossing over in the inverted region leads to gametes with unbalanced chromosomes with potentially large duplications and deletions (Rieseberg, 2001). However, there are several mechanisms that can reduce the creation of unbalanced gametes. When inversions are heterozygous the inverted region can either pair **homosynaptically** or **heterosynaptically** (Torgasheva & Borodin, 2010). COs can only occur in homosynaptically paired regions. Additionally, DSBs can be repaired as NCOs (see above) or the rate of DSBs can be reduced in the inverted region (Fuller, Koury, Phadnis, & Schaeffer, 2019). All of these mechanisms will reduce the recombination rate without the cost of unbalanced gametes. Alternatively, recombination can proceed normally and create balanced products but these products may fail to segregate properly. In translocation heterozygotes, the four involved chromosomes form a quadrivalent structure during meiosis. Segregation from this structure can lead to the creation of aneuploid gametes with a rate of 18% to more than 80% (Morel et al., 2004; Talukdar, 2010). Similarly, nondisjunction rates in fusion heterozygotes may be elevated, ranging from 1.2% to 30% depending on the system (Dobigny et al., 2017).

The mechanism by which recombination is reduced varies greatly between mutation types and this will have strong downstream consequences for the extent of recombination reduction as well as additional population genetic effects such as dominance, which will impact the evolutionary fate of the mutation. Using theoretical tools, we explore the relationship between molecular mechanism and recombination rate in Box 2.

Box 2 - Quantifying the impact of structural variants on recombination

In this box we endeavor to show how the molecular underpinnings of a population genetic effect (recombination) may be incorporated theoretically. To quantify this effect, we derive the probability, $P(x_1, x_2)$, that two loci at position $x_1$and x_2(with $x_1 < x_2$), initially on the same homolog, are separated during meiosis in the presence of various structural variants. We present here only approximations obtained when the rate of double strand break (DSB) is sufficiently small (see Supplement for detailed expressions).

In the absence of a structural variant, the probability that two loci are separated by recombination is given by:

$$P_{rec}(x_1, x_2) \cong \beta_{DSB}\left(\lambda \phi_{GC} + (x_2 - x_1)\phi_{CO}\right)$$

, with ϕ_{GC} and ϕ_{CO} as the probabilities that a DSB leads respectively to gene conversion (GC) and a crossover and λ the length of a GC tract ($x_2 > x_1 + \lambda$). The first term corresponds to one locus being transferred by GC and the second to a crossover between the two focal loci.

Insertion/deletion (Indel):

Recombination only happens in the ancestral (deletion) or derived (insertion) homozygote (its frequency denoted f_{AA}). The two loci are separated with probability:

$$P_{Indel}(x_1, x_2) = f_{AA} P_{rec}(x_1, x_2)$$

Inversion:

Single crossovers occurring within the inversion breakpoints in heterozygotes form gametes with unbalanced chromosomes, leading to inviable zygotes. Therefore, heterozygotes are underdominant and recombination only happens through GC or double crossovers. The probability that two loci in the inverted region are separated is given by (assuming $\beta_{DSB} \ll f_{Aa} - 0.5$):

$$P_{inv}(x) \cong \left(\frac{(x_2 - x_1)(1 - 2f_{Aa})\phi_{CO}}{1 - f_{Aa}} + \lambda \phi_{GC} \right) \beta_{DSB}$$

The first term corresponds to a recombination event happening between the focal loci in the

homozygotes, whose frequency is increased due to underdominant heterozygotes. The second term corresponds to GC and remains unaffected. Double crossovers do not play a significant role under those conditions.

Fusion:

For chromosomal fusions, homologs in heterozygotes may fail to segregate properly (with probability β_{NDJ}), producing unbalanced gametes and reducing the contribution of heterozygotes to the next generation. In addition, the chance of a crossover decreases if at least one fused chromosome is involved. The two loci are separated with probability:

$$P_{fus}(x_1, x_2) \cong \left(S_1(f_{Aa}, f_{AA})(x_2 - x_1)\phi_{CO} + \lambda\,\phi_{GC} \right)\beta_{DSB}$$

The contribution of crossovers is reduced by a factor $S_1(f_{Aa}, f_{AA})$, which depends on the genotypes frequencies and captures both selection against the heterozygote and the reduced crossover probability when at least one fused chromosome is involved.

Translocation:

Similarly, homologs in heterozygotes may fail to segregate properly, producing unbalanced gametes and reducing the contribution of heterozygotes to the next generation. The GC rate in heterozygotes is also reduced. The two loci are separated with probability:

$$P_{trans}(x_1, x_2) \cong \left((x_2 - x_1)\phi_{CO} + S_2(f_{Aa})\lambda\,\phi_{GC} \right)\beta_{DSB}$$

The contribution of gene conversion is reduced by a factor $S_2(f_{Aa})$, which depends on the frequency of the heterozygotes and captures both the effect of selection against, and the reduction of gene conversion within, heterozygotes.

Overall, these results show that the extent of recombination reduction likely differs between mutation types. Although the parameter space remains unexplored, under our assumptions, recombination was most strongly reduced in indels followed by inversions. Surprisingly, recombination in the translocation closely mirrored default recombination rate.

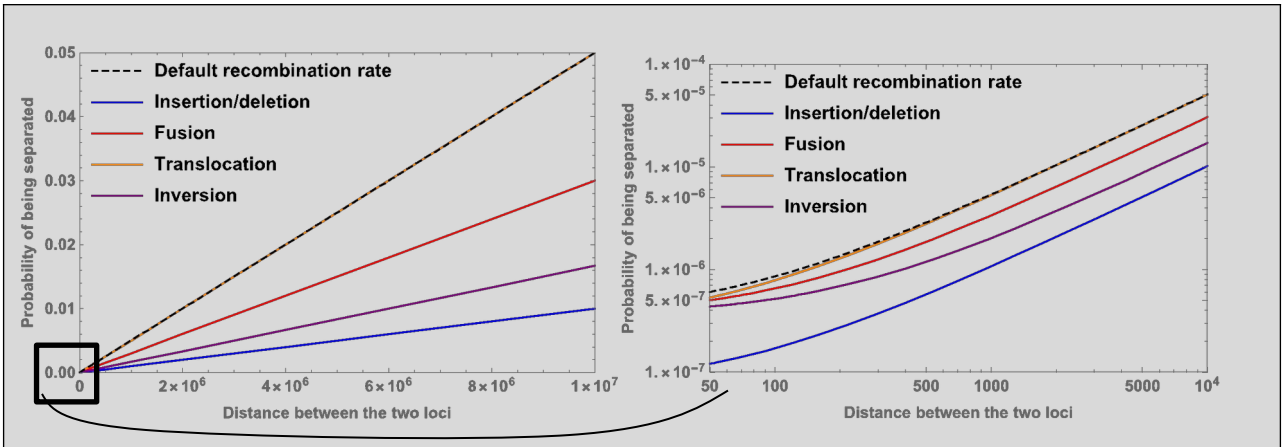


Figure Box 2. Probability that two loci on the same structural variant are separated due to recombination as a function of the distance between the two loci. Parameters: $\beta_{DSB}=10^{-8}; \phi_{CO}=0.5; \phi_{GC}=0.7; \lambda=50; f_{Aa}=0.4; f_{AA}=0.2$. The factors S_1 and S_2 are calculated using the expressions given in the Supplement; here $S_1=-0.6$ and $S_2=-0.8$.

Dominance

Dominance determines the penetrance of a mutation and its visibility to selection and can have complex effects on evolutionary processes. While best examined on a case by case basis, there are a few trends that have been noted between mutation types and dominance. For example, as long as an insertion contains a single-copy gene (i.e. a gene that is not present elsewhere in the genome), alleles that are normally recessive will be expressed in the heterozygous state. CNVs themselves will alter the penetrance of a dominant mutation. For example, duplications of the recessive allele may nullify the dominant mutation or compensate identical alleles with low gene expression level (Beckmann et al., 2007). Other mutations may also have effects that spread outside of the mutated region. For example, many TEs contain regulatory elements for their own mobility, potentially rewiring the regulation of nearby genes, altering dominance patterns (Chuong, Elde, & Feschotte, 2017).

The dominance effects of inversions depend on multiple factors. For example, an inversion might be underdominant if COs in the inversion region lead to unbalanced gametes (see above). On the other hand, recessive deleterious alleles can accumulate within both the standard and the inverted arrangement, generating (associative) overdominance at the level of the inversion because recessive alleles are shielded in inversion heterokaryotypes (Ohta, 1971).

Effective Population Size

Effective population size (N_e) usually reflects the process of drift. N_e can be locally or globally affected by mutations, either by a direct reduction in the number of gene copies or indirectly through the mutation's effects on recombination rate and selection coefficient (Charlesworth, n.d.; Gossmann, Woolfit, & Eyre-Walker, 2011).

Recombination and therefore N_e can be reduced by a variety of mutation types. For example, loci within a large indel will experience this reduction in N_e twice, once due to the reduction in recombination (see Box 2) and once due to lower number of copies of the indel region. The indel as a locus, with two alleles: 'present' and 'absent', will not be affected by either of these processes. Similarly, recombination between different arrangements of a heterozygous inversion is lowered and the arrangements can be viewed as two smaller and partially isolated populations (Berdan, Blanckaert, Butlin, & Bank, 2019; Faria et al., 2019). Translocations and fusions will experience a similar effect.

Changes in fitness due to mutations can also lead to a reduction in N_e. For example, TE insertions are weakly deleterious in many sequence contexts (Choi & Lee, 2020a; Hollister & Gaut, 2009) leading them to be removed by selection along with linked neutral variation (i.e., background

selection). Translocations or chromosome fusions lead to high rates of non-disjunction and subsequent negative selection against heterozygotes reducing their contribution to future generations (Dobigny et al., 2017; Morel et al., 2004; Talukdar, 2010). Conversely, centromere shifts may be under positive selection if they exhibit centromere drive (Malik, 2009).

Selection Coefficient

All of the population genetic effects described above will together regulate the interaction of the mutation with selection and drift, and determine evolutionary outcomes (Box 3). Changes in the interaction with selection and drift are partially quantified in the selection coefficient, a measure of differences in relative fitness, encompassing multiple population genetic effects. The selection coefficient of a mutation depends on a multitude of genomic factors including (1) the genomic context, i.e., whether it alters coding, regulatory, or neutrally evolving regions and (2) whether or not it causes a positional shift; but also on non-genomic factor such as the selective environment (both extrinsic and intrinsic) where the change occurs (Brandström et al., 2008; Crown et al., 2018; Ducos et al., 2007; Flynn et al., 2020; Gemayel et al., 2010; Guo et al., 2009; Hollister & Gaut, 2009; Kayser, Vowles, Kappei, & Amos, 2006; Kent et al., 2017; Korunes & Noor, 2019; Sherizen et al., 2005; Stapley et al., 2017; Weissensteiner et al., 2020). Furthermore, duplicated regions, such as CNVs, have additional effects as they may free up selective constraints and can lead to the emergence of new gene functions (Ohno, 2013).

Selection coefficients can be examined more globally using the **distribution of fitness effects (DFE)**, that summarizes the interaction of the mutation type with drift and selection (Eyre-Walker & Keightley, 2007; Keightley & Eyre-Walker, 2010). Most studies have estimated the DFE of SNPs and have found a bi- or multi-modal distribution, with beneficial mutations being

rare, although the exact shape of distributions vary (Bataillon & Bailey, 2014; Eyre-Walker & Keightley, 2007; Keightley & Eyre-Walker, 2010). However, the DFE of other mutation types may have different properties (but see (Barton & Zeng, 2018)). For example, a study in *E.coli* (Elena, Ekunwe, Hajela, Oden, & Lenski, 1998) showed a long deleterious tail and a high neutral peak for TE insertions. Most CNVs are expected to be found at the extremes of the distribution with either beneficial or largely deleterious effects (Katju & Bergthorsson, 2013). While the DFE allows us to make certain evolutionary predictions it does not quantify critical population genetic effects such as recombination rate or genomic effects such as changes in the regulatory landscape.

Large structural variants alter the efficacy of selection within the mutated region by modifying the recombination rate and local N_e. This aspect can be beneficial, for example inversions may be indirectly selected because they reduce recombination between multiple beneficial alleles located in the same arrangement (e.g., locally adapted alleles under gene flow (Kirkpatrick & Barton, 2006)). However, these changes also alter the evolution of the mutated region in multiple ways, e.g., by reducing the efficacy of purifying selection leading to the accumulation of deleterious alleles. Quantifying the impact of these changes, through a combination of analytical approaches and simulations (e.g., Gilbert, Pouyet, Excoffier, & Peischl, 2020) will be a key step towards linking mutation type with evolutionary significance.

Box 3 - Contribution of different mutation types to speciation

Here, we use speciation as an example to demonstrate the application of our framework.

A) Predicting the contribution of different mutation types

Recombination is a key population genetic effect relevant for speciation. Indeed, in the presence of gene flow, speciation can only progress when associations (linkage disequilibria) between alleles at different loci contributing to population differences are maintained and increase (Smadja & Butlin, 2011), and theoretical work predicts that such associations are facilitated by structural variants that reduce recombination (Kirkpatrick & Barton, 2006). This effect has been mostly studied for inversions, but other recombination-reducing mutations, including fusions and maybe centromere shifts, might also be important.

As discussed in the main text, the contribution of a mutation type to an evolutionary process is determined not only by its population genetic effects but also by its occurrence rate. SNPs are the most commonly occurring types of variants (Fig. 1B). While DFE studies have shown that the majority of SNPs are typically deleterious (Eyre-Walker & Keightley, 2007), SNPs are still likely to make a major contribution to divergence and speciation. TEs have also been hypothesised to be particularly relevant here, as their mutation rates can increase under stress. Increased TE activity in new environments might generate novel diversity, some of which may be adaptive and contribute to population divergence (Stapley, Santure, & Dennis, 2015).

While this Box demonstrates that different mutation types have been predicted to play a role in speciation, their relative importance is less clear. For example, making more detailed predictions about the relative importance of different recombination-reducing mutations requires more empirical data on their effects on recombination, the selection pressures acting on them (e.g., over- vs. underdominance), as well as theoretical models and simulations that directly compare them (Box 2).

B) Empirical data about the relative contribution of different mutation types to speciation with gene flow

There are several well-studied systems where multiple mutation types have been analysed. For example, for stickleback freshwater-marine divergence, causal mutations are known to include SNPs/small indels (Archambeault, Bärtschi, Merminod, & Peichel, 2020) as well as deletions (Chan et al., 2010). Furthermore, differentially adapted populations also differ in the frequency of an inversion (Jones et al., 2012). Flowering plants of the *Mimulus* species complex are another example where multiple mutation types, including SNPs, inversions, translocations and duplications, have been studied (Twyford & Friedman, 2015; Zuellig & Sweigart, 2018).

However, studies looking systematically for the relative contribution of all different mutation types are essentially lacking. Part of the problem is that it is often difficult to pinpoint the exact causal mutation, rather than identifying just a larger genomic region associated with population divergence or underlying divergent traits. As these genomic regions typically contain many small variants, it is difficult to determine whether the causal variant is a SNP, indel, or TE, for example. Additional studies looking at repeated evolution and/or functional studies (i.e. CRISPR/Cas9 modification of a single locus) will be necessary to clarify this. Overall, further work is needed to systematically compare the contribution of different mutation types in the same study system.

Population Genetic Effects of Mutations: Connecting to evolutionary outcomes

To understand the evolutionary significance of a mutation, we must examine its past, current and potential future effects. Many mutation types alter critical population genetic parameters, shifting the evolutionary trajectory of the genomic region. Going forward, a combination of theoretical and empirical tools can be used to: (1) Estimate the population genetic effects of all mutation types simultaneously, (2) Determine how these effects can shape the evolutionary trajectory of the population, (3) Combine 1 and 2 to link these effects to evolutionary outcomes (e.g., speciation, adaptation).

As summarized above, there is already a large body of work describing the molecular underpinnings of a mutation's population genetic effects. However, while we know the general direction of these effects (Figure 1C), we do not know much about their effect size distributions. Empirical studies quantifying the population genetic effects of mutations on a large scale are sorely needed for better characterization of mutation effects and to determine how these effects vary across taxa. The DFE is a good place to start, but we also need to quantify the other population genetic effects of mutations. A critical question is how these effect sizes vary between mutation types. For example, to what extent is recombination reduced in an inversion compared to a fusion? This information can be directly related to evolutionary outcomes, e.g., is a fusion or an inversion a "better" genetic background for maintaining a complex of co-adapted alleles? With enough data it may even be possible to estimate distributions of population genetic effects for different mutation types. Below we discuss both empirical and theoretical ways forward.

Several of the population genetic effects described here are relatively straightforward to measure. For example, the influence of a mutation on recombination can be simply determined using mapping crosses followed by sequencing and bioinformatic detection of gene conversion and crossing over events. The effect of inversions on recombination in *Drosophila* has already been intensively examined using this methodology (Crown et al., 2018; Korunes & Noor, 2019). The application of these methods to other systems and other types of mutations (e.g. fusions) will allow us to determine the distribution of recombination suppression for different types of mutations. Changes in physical distance can be quantified in part by utilizing pre-existing sequence data to examine the distribution of sizes of different types of structural mutations. These types of studies provide a start for building distributions of population genetic effects.

In addition to the empirical approaches described in the last paragraph, we can begin to ask this question using theoretical models. We provide an example in Box 2 showing how molecular genetic information about recombination may be incorporated theoretically. Although exploring the parameter space was beyond our scope, these results illustrate that the extent of recombination reduction likely differs between mutation types. Box 2 represents a starting point for integrating underlying molecular mechanisms with their impacts on population genetic parameters. Complementing analytical approaches with simulations allows for exploration of more complex effects on a larger scale. For example, although a decrease in population size can be easily measured, the corresponding effect on Ne, and therefore on the role of drift, is more complex to quantify.

In order to determine how these population genetic effects link to evolutionary outcomes we must examine how they shift evolutionary trajectories. Theoretical models provide the best avenue for this. There is more than a century of literature developing these methods in population genetics (Box 3, Charlesworth, 2010b). Integrating the feedback loop between the evolution of a structural variant as a locus and the evolution of its allelic content into theoretical models may further our understanding of the link between mutation types and evolutionary outcomes. For example, in the case of an inversion, the resulting reduction in recombination rate generally leads to an accumulation of deleterious alleles in the minor variant, slowly increasing the fitness differential with the major common variant (Berdan et al., 2019). Looking at empirical data can show the result of these shifts in trajectory. Patterns of nucleotide diversity, divergence between arrangements, and the DFE of the alleles within the mutated region can be examined and compared with predictions from simulation studies. In this way, the forward and reverse approaches can be merged (Figure 1A).

Empirical studies also provide critical information about what mutation types have previously been important in evolutionary outcomes. However, most studies do not simultaneously compare different types of mutations. Moving forward will require collecting different types of genomic data sets (e.g. short- and long-read re-sequencing and mapping crosses) from the same population and developing detection pipelines targeted at different mutation types (Mérot, Oomen, Tigano, & Wellenreuther, 2020). Synthesizing information on mutation types and evolutionary outcomes allows us to both explore the relationships between mutation type and the major evolutionary outcomes and to test predictions based on population genetic effects. For example, speciation requires the build-up of linkage disequilibrium between alleles contributing

to reproductive isolation (Box 3, Butlin & Smadja, 2018). Mutations that reduce recombination should aid speciation with gene flow by protecting this nascent linkage disequilibrium. We can thus predict that mutations such as inversions, large indels, TEs, fusions, and centromere shifts might be major drivers of speciation events (Fuller, Leonard, Young, Schaeffer, & Phadnis, 2018). A critical next step would be testing some of these hypotheses in a quantitative rather than review framework, for example using a meta-analysis.

Experimental evolution offers another way to integrate the forward and reverse approaches detailed above. These studies link mutation type with evolutionary outcome in real time (Kawecki et al., 2012), generating results that can be compared with theoretical predictions and empirical results from natural populations. For example, starting populations for experimental evolution studies can incorporate genetic variation for multiple mutation types (e.g. segregating inversions, CNVs, etc). The evolutionary trajectories of these different mutation types can then be followed during the adaptive or divergence process and these can be combined with functional studies to pinpoint adaptive variants. In this way the relative role of different mutations can begin to be dissected. Concomitantly, existing genomic data from previous experimental evolution studies can also be utilized. By using different software programs (e.g., Chen et al., 2016; Kawecki et al., 2012; Liu et al., 2020; Moreno-Cabrera et al., 2020; Shigemizu et al., 2018) to detect different types of mutations, it should be possible to quantify the relative role of different mutations in these different scenarios. Overall, experimental evolution studies can provide a valuable counterpoint to theoretical predictions and data from more traditional population genomic studies. (Moreno-Cabrera et al., 2020)

Concluding remarks and future perspectives

Our framework highlights the fact that each mutation type may affect evolution in several ways and that many different mutation types have similar population genetic effects (Box 2, Figure 1). Analyzing this in a quantitative and comparative way will allow us to explore the evolutionary significance of different mutation types.

Understanding the relative evolutionary significance of different mutations will require viewing their effects in a larger population genetic context. In order to do this we need: (1) Comparable measurements of occurrence rates as well as the population genetic effects of different mutation types; (2) To include these effects in theoretical models and simulations to create predictions about the importance of different mutation types; (3) To empirically estimate the contributions of different mutations to evolutionary outcomes and test the predictions obtained from the theoretical models. Superimposing a more integrated framework on previous and future work will allow us to better understand the relative contributions of different mutation types to key evolutionary outcomes, further illuminating the genetic underpinnings of these processes in a broad sense.

Glossary

Single Nucleotide Polymorphism (SNP) - A single base pair variant in a specific position in the genome.

Copy Number Variant (CNV) - A DNA segment of at least one kb that is present at a variable copy number.

Centromere shift - Repositioning of the centromere along the chromosome.

Centromere drive - Non-Mendelian inheritance of a centromere variant through the asymmetry of female meiosis.

Distribution of Fitness Effects (DFE) - Describes the proportion of new mutations that are beneficial, deleterious or neutral in a specific environment.

Transposable Element (TE) - A selfish genetic element propagating via copy-and-paste or cut-and-paste.

Simple Sequence Repeat (SSR) - Tandem repeats of 2-6 bp motifs.

Hill-Robertson effect - Describes selection having a reduced effect when selected sites are in tight linkage with sites under different selection pressures.

Translocation (**also called balanced or reciprocal translocation)** - Two pieces of non-homologous chromosomes that have broken off and been switched.

Fusion (also called Robertsonian fusion, Robertsonian translocation, and **centric fusion)** - Two acrocentric chromosomes (where the centromere is located near the end of the chromosome) that have experienced breaks at or near the centromere and then fused creating a metacentric chromosome.

Gene Conversion - The replacement of a DNA sequence by the homologous sequence such that the two sequences are then identical.

Homosynaptic pairing - When the two homologs correctly synapse during prophase 1.

Heterosynaptic pairing - When non-homologous (heterologous) synapsis occurs during prophase 1.

Effective population size, N_e - The equivalent population size of a Wright-Fisher population that will generate population genetic statistics closest to the ones of the focal population.

Effective recombination rate (*sensu* Golding (Golding & Strobeck, 1980)**)** - The equivalent recombination rate in a Wright-Fisher population that will generate the same linkage disequilibrium patterns as those found in the focal population.

Indel - Genetic variant that can be either inserted or deleted from the genome

Inversion - A segment of the genome that is rotated 180 degrees.

Acknowledgements

We thank Roger Butlin, Kerstin Johannesson, Valentina Peona, Rike Stelkens, Julie Blommaert, Nick Barton, and João Alpedrinha for helpful comments that improved the manuscript. The authors acknowledge funding from the Swedish Research Council Formas (2017-01597 to AS), the Swedish Research Council Vetenskapsrådet (2016-05139 to AS, 2019-04452 to TS) and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 757451 to TS). ELB was funded by a Carl Tryggers grant awarded to TS. AMW was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 797747. IF was funded by a Junior Researcher contract from FCT (CEECIND/02616/2018)

References

Archambeault, S. L., Bärtschi, L. R., Merminod, A. D., & Peichel, C. L. (2020). Adaptation via pleiotropy and linkage: Association mapping reveals a complex genetic architecture within the stickleback locus. *Evolution Letters*, *4*(4), 282–301. doi: 10.1002/evl3.175

Barton, H. J., & Zeng, K. (2018). New Methods for Inferring the Distribution of Fitness Effects for INDELs and SNPs. *Molecular Biology and Evolution*, *35*(6), 1536–1546. doi: 10.1093/molbev/msy054

Bataillon, T., & Bailey, S. F. (2014). Effects of new mutations on fitness: insights from models and data. *Annals of the New York Academy of Sciences*, Vol. 1320, pp. 76–92. doi: 10.1111/nyas.12460

Beckmann, J. S., Estivill, X., & Antonarakis, S. E. (2007). Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nature Reviews. Genetics*, *8*(8), 639–646. doi: 10.1038/nrg2149

Berdan, E. L., Blanckaert, A., Butlin, R. K., & Bank, C. (in review). *Deleterious mutation accumulation and the long-term fate of chromosomal inversions*. doi: 10.1101/606012

Bidau, C. J., Giménez, M. D., Palmer, C. L., & Searle, J. B. (2001). The effects of Robertsonian fusions on chiasma frequency and distribution in the house mouse (Mus musculus domesticus) from a hybrid zone in northern Scotland. *Heredity*, *87*(Pt 3), 305–313. doi: 10.1046/j.1365-2540.2001.00877.x

Brandström, M., Bagshaw, A. T., Gemmell, N. J., & Ellegren, H. (2008). The relationship between microsatellite polymorphism and recombination hot spots in the human genome. *Molecular Biology and Evolution*, *25*(12), 2579–2587. doi: 10.1093/molbev/msn201

Brumfield, R. T., Beerli, P., Nickerson, D. A., & Edwards, S. V. (2003). The utility of single

nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*, *18*(5), 249–256. doi: 10.1016/S0169-5347(03)00018-1

Butlin, R. K., & Smadja, C. M. (2018). Coupling, Reinforcement, and Speciation. *The American Naturalist*, *191*(2), 155–172. doi: 10.1086/695136

Capilla, L., Medarde, N., Alemany-Schmidt, A., Oliver-Bonet, M., Ventura, J., & Ruiz-Herrera, A. (2014). Genetic recombination variation in wild Robertsonian mice: on the role of chromosomal fusions and Prdm9 allelic background. *Proceedings. Biological Sciences / The Royal Society*, *281*(1786). doi: 10.1098/rspb.2014.0297

Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal, G., Jr, Shapiro, M. D., Brady, S. D., … Kingsley, D. M. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science*, *327*(5963), 302–305. doi: 10.1126/science.1182213

Charlesworth, B. (2010a). *Elements of Evolutionary Genetics*. Roberts Publishers. Retrieved from https://books.google.com/books/about/Elements_of_Evolutionary_Genetics.html?hl=&id=dgNFAQAAIAAJ

Charlesworth, B. (2010b). *Elements of Evolutionary Genetics*. Roberts Publishers. Retrieved from https://books.google.com/books/about/Elements_of_Evolutionary_Genetics.html?hl=&id=dgNFAQAAIAAJ

Charlesworth, B. (n.d.). *Effective population size and patterns of molecular evolution and variation*. doi: 10.1038/nrg2526

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., … Saunders, C. T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* , *32*(8), 1220–1222. doi: 10.1093/bioinformatics/btv710

Choi, J. Y., & Lee, Y. C. G. (2020a). Double-edged sword: The evolutionary consequences of the epigenetic silencing of transposable elements. *PLOS Genetics*, Vol. 16, p. e1008872. doi: 10.1371/journal.pgen.1008872

Choi, J. Y., & Lee, Y. C. G. (2020b). Double-edged sword: The evolutionary consequences of the epigenetic silencing of transposable elements. *PLoS Genetics*, *16*(7), e1008872. doi: 10.1371/journal.pgen.1008872

Chuong, E. B., Elde, N. C., & Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews. Genetics*, *18*(2), 71–86. doi: 10.1038/nrg.2016.139

Crown, K. N., Miller, D. E., Sekelsky, J., & Hawley, R. S. (2018). Local Inversion Heterozygosity Alters Recombination throughout the Genome. *Current Biology: CB*, *28*(18), 2984–2990.e3. doi: 10.1016/j.cub.2018.07.004

Dobigny, G., Britton-Davidian, J., & Robinson, T. J. (2017). Chromosomal polymorphism in mammals: an evolutionary perspective. *Biological Reviews of the Cambridge Philosophical Society*, *92*(1), 1–21. doi: 10.1111/brv.12213

Ducos, A., Berland, H.-M., Bonnet, N., Calgaro, A., Billoux, S., Mary, N., … Pinton, A. (2007). Chromosomal control of pig populations in France: 2002-2006 survey. *Genetics, Selection, Evolution: GSE*, *39*(5), 583–597. doi: 10.1186/1297-9686-39-5-583

Elena, S. F., Ekunwe, L., Hajela, N., Oden, S. A., & Lenski, R. E. (1998). Distribution of fitness effects caused by random insertion mutations in Escherichia coli. *Genetica*, *102-103*(1-6), 349–358. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9720287

Eyre-Walker, A., & Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews. Genetics*, *8*(8), 610–618. doi: 10.1038/nrg2146

Faria, R., Johannesson, K., Butlin, R. K., & Westram, A. M. (2019). Evolving Inversions. *Trends in Ecology & Evolution*, *34*(3), 239–248. doi: 10.1016/j.tree.2018.12.005

Farlow, A., Long, H., Arnoux, S., Sung, W., Doak, T. G., Nordborg, M., & Lynch, M. (2015). The Spontaneous Mutation Rate in the Fission Yeast Schizosaccharomyces pombe. *Genetics*, *201*(2), 737–744. doi: 10.1534/genetics.115.177329

Feusier, J., Scott Watkins, W., Thomas, J., Farrell, A., Witherspoon, D. J., Baird, L., … Jorde, L. B. (2019). Pedigree-based estimation of human mobile element retrotransposition rates. *Genome Research*, Vol. 29, pp. 1567–1577. doi: 10.1101/gr.247965.118

Flynn, J. M., Rossouw, A., Cote-Hammarlof, P., Fragata, I., Mavor, D., Hollins, C., … Bolon, D. N. A. (2020). Comprehensive fitness maps of Hsp90 show widespread environmental dependence. *eLife*, Vol. 9. doi: 10.7554/elife.53810

Fox, D. T., Soltis, D. E., Soltis, P. S., Ashman, T.-L., & Van de Peer, Y. (2020). Polyploidy: A Biological Force From Cells to Ecosystems. *Trends in Cell Biology*, *30*(9), 688–694. doi: 10.1016/j.tcb.2020.06.006

Fuller, Z. L., Koury, S. A., Phadnis, N., & Schaeffer, S. W. (2019). How chromosomal rearrangements shape adaptation and speciation: Case studies in Drosophila pseudoobscura and its sibling species Drosophila persimilis. *Molecular Ecology*, *28*(6), 1283–1301. doi: 10.1111/mec.14923

Fuller, Z. L., Leonard, C. J., Young, R. E., Schaeffer, S. W., & Phadnis, N. (2018). Ancestral polymorphisms explain the role of chromosomal inversions in speciation. *PLoS Genetics*, *14*(7), e1007526. doi: 10.1371/journal.pgen.1007526

Futuyma, D. J. (1986). Reflections on reflections: ecology and evolutionary biology. *Journal of the History of Biology*, *19*(2), 303–312. doi: 10.1007/BF00138881

Fu, W., Zhang, F., Wang, Y., Gu, X., & Jin, L. (2010). Identification of copy number variation hotspots in human populations. *American Journal of Human Genetics*, *87*(4), 494–504. doi: 10.1016/j.ajhg.2010.09.006

Fu, Y.-X., & Huai, H. (2003). Estimating mutation rate: how to count mutations? *Genetics*, *164*(2), 797–805. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12807798

Gemayel, R., Vinces, M. D., Legendre, M., & Verstrepen, K. J. (2010). Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annual Review of Genetics*, *44*, 445–477. doi: 10.1146/annurev-genet-072610-155046

Gilbert, K. J., Pouyet, F., Excoffier, L., & Peischl, S. (2020). Transition from Background Selection to Associative Overdominance Promotes Diversity in Regions of Low Recombination. *Current Biology: CB*, *30*(1), 101–107.e3. doi: 10.1016/j.cub.2019.11.063

Goerner-Potvin, P., & Bourque, G. (2018). Computational tools to unmask transposable elements. *Nature Reviews. Genetics*, *19*(11), 688–704. doi: 10.1038/s41576-018-0050-x

Golding, G. B., & Strobeck, C. (1980). Linkage disequilibrium in a finite population that is partially selfing. *Genetics*, *94*(3), 777–789. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/17249017

Gong, W. J., McKim, K. S., & Hawley, R. S. (2005). All paired up with no place to go: pairing, synapsis, and DSB formation in a balancer heterozygote. *PLoS Genetics*, *1*(5), e67. doi: 10.1371/journal.pgen.0010067

Gossmann, T. I., Woolfit, M., & Eyre-Walker, A. (2011). Quantifying the variation in the effective population size within a genome. *Genetics*, *189*(4), 1389–1402. doi: 10.1534/genetics.111.132654

Guo, W.-J., Ling, J., & Li, P. (2009). Consensus features of microsatellite distribution:

Microsatellite contents are universally correlated with recombination rates and are preferentially depressed by centromeres in multicellular eukaryotic genomes. *Genomics*, Vol. 93, pp. 323–331. doi: 10.1016/j.ygeno.2008.12.009

Hammarlund, M., Davis, M. W., Nguyen, H., Dayton, D., & Jorgensen, E. M. (2005). Heterozygous insertions alter crossover distribution but allow crossover interference in Caenorhabditis elegans. *Genetics*, *171*(3), 1047–1056. doi: 10.1534/genetics.105.044834

Henderson, I. R. (2012). Control of meiotic recombination frequency in plant genomes. *Current Opinion in Plant Biology*, *15*(5), 556–561. doi: 10.1016/j.pbi.2012.09.002

Hollister, J. D., & Gaut, B. S. (2009). Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research*, *19*(8), 1419–1428. doi: 10.1101/gr.091678.109

Ho, S. S., Urban, A. E., & Mills, R. E. (2020). Structural variation in the sequencing era. *Nature Reviews. Genetics*, *21*(3), 171–189. doi: 10.1038/s41576-019-0180-9

Hunter, N. (2015). Meiotic Recombination: The Essence of Heredity. *Cold Spring Harbor Perspectives in Biology*, *7*(12). doi: 10.1101/cshperspect.a016618

Jarne, P., & Lagoda, P. J. (1996). Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution*, *11*(10), 424–429. doi: 10.1016/0169-5347(96)10049-5

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55–61. doi: 10.1038/nature10944

Katju, V., & Bergthorsson, U. (2013). Copy-number changes in evolution: rates, fitness effects and adaptive significance. *Frontiers in Genetics*, *4*, 273. doi: 10.3389/fgene.2013.00273

Kawecki, T. J., Lenski, R. E., Ebert, D., Hollis, B., Olivieri, I., & Whitlock, M. C. (2012).

Experimental evolution. *Trends in Ecology & Evolution*, *27*(10), 547–560. doi: 10.1016/j.tree.2012.06.001

Kayser, M., Vowles, E. J., Kappei, D., & Amos, W. (2006). Microsatellite length differences between humans and chimpanzees at autosomal Loci are not found at equivalent haploid Y chromosomal Loci. *Genetics*, *173*(4), 2179–2186. doi: 10.1534/genetics.106.055632

Keightley, P. D., & Eyre-Walker, A. (2010). What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1544), 1187–1193. doi: 10.1098/rstb.2009.0266

Kent, T. V., Uzunović, J., & Wright, S. I. (2017). Coevolution between transposable elements and recombination. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1736). doi: 10.1098/rstb.2016.0458

Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, *173*(1), 419–434. doi: 10.1534/genetics.105.047985

Korunes, K. L., & Noor, M. A. F. (2019). Pervasive gene conversion in chromosomal inversion heterozygotes. *Molecular Ecology*, *28*(6), 1302–1315. doi: 10.1111/mec.14921

Lee, Y. C. G., & Karpen, G. H. (2017). Pervasive epigenetic effects of Drosophila euchromatic transposable elements impact their evolution. *eLife*, Vol. 6. doi: 10.7554/elife.25762

Liu, Y., Zhang, M., Sun, J., Chang, W., Sun, M., Zhang, S., & Wu, J. (2020). Comparison of multiple algorithms to reliably detect structural variants in pears. *BMC Genomics*, *21*(1), 61. doi: 10.1186/s12864-020-6455-x

Maeda, T., Ohno, M., Matsunobu, A., Yoshihara, K., & Yabe, N. (1991). A cytogenetic survey of 14,835 consecutive liveborns. *Jinrui Idengaku Zasshi. The Japanese Journal of Human*

Genetics, *36*(1), 117–129. doi: 10.1007/BF01876812

Malik, H. S. (2009). The centromere-drive hypothesis: a simple basis for centromere complexity. *Progress in Molecular and Subcellular Biology*, *48*, 33–52. doi: 10.1007/978-3-642-00182-6_2

Marriage, T. N., Hudman, S., Mort, M. E., Orive, M. E., Shaw, R. G., & Kelly, J. K. (2009). Direct estimation of the mutation rate at dinucleotide microsatellite loci in Arabidopsis thaliana (Brassicaceae). *Heredity*, Vol. 103, pp. 310–317. doi: 10.1038/hdy.2009.67

Marshall, O. J., Chueh, A. C., Wong, L. H., & Choo, K. H. A. (2008). Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. *American Journal of Human Genetics*, *82*(2), 261–282. doi: 10.1016/j.ajhg.2007.11.009

McCartney, D. L., Walker, R. M., Morris, S. W., Anderson, S. M., Duff, B. J., Marioni, R. E., … Evans, K. L. (2018). Altered DNA methylation associated with a translocation linked to major mental illness. *NPJ Schizophrenia*, *4*(1), 5. doi: 10.1038/s41537-018-0047-7

Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. *Trends in Ecology & Evolution*, Vol. 35, pp. 561–572. doi: 10.1016/j.tree.2020.03.002

Morel, F., Douet-Guilbert, N., Le Bris, M.-J., Herry, A., Amice, V., Amice, J., & De Braekeleer, M. (2004). Meiotic segregation of translocations during male gametogenesis. *International Journal of Andrology*, *27*(4), 200–212. doi: 10.1111/j.1365-2605.2004.00490.x

Moreno-Cabrera, J. M., Del Valle, J., Castellanos, E., Feliubadaló, L., Pineda, M., Brunet, J., … Gel, B. (2020). Evaluation of CNV detection tools for NGS panel data in genetic diagnostics. *European Journal of Human Genetics: EJHG*, *28*(12), 1645–1655. doi: 10.1038/s41431-020-0675-z

Myers, S., Bottolo, L., Freeman, C., McVean, G., & Donnelly, P. (2005). A fine-scale map of recombination rates and hotspots across the human genome. *Science*, *310*(5746), 321–324. doi: 10.1126/science.1117196

Ohno, S. (2013). *Evolution by Gene Duplication*. Springer Science & Business Media. Retrieved from https://play.google.com/store/books/details?id=5SjqCAAAQBAJ

Ohta, T. (1971). Associative overdominance caused by linked detrimental mutations. *Genetics Research*, *18*(3), 277–286. Retrieved from https://www.cambridge.org/core/journals/genetics-research/article/associative-overdominance-caused-by-linked-detrimental-mutations/B8C9CC6B89C486234CA11A71F2AE741C

Ossowski, S., Schneeberger, K., Lucas-Lledó, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., … Lynch, M. (2010). The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. *Science*, *327*(5961), 92–94. doi: 10.1126/science.1180677

Poorman, P. A., Moses, M. J., Russell, L. B., & Cacheiro, N. L. (1981). Synaptonemal complex analysis of mouse chromosomal rearrangements. I. Cytogenetic observations on a tandem duplication. *Chromosoma*, *81*(4), 507–518. doi: 10.1007/BF00285846

Ramu, A., Noordam, M. J., Schwartz, R. S., Wuster, A., Hurles, M. E., Cartwright, R. A., & Conrad, D. F. (2013). DeNovoGear: de novo indel and point mutation discovery and phasing. *Nature Methods*, *10*(10), 985–987. doi: 10.1038/nmeth.2611

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, Vol. 16, pp. 351–358. doi: 10.1016/s0169-5347(01)02187-5

Rocchi, M., Archidiacono, N., Schempp, W., Capozzi, O., & Stanyon, R. (2012). Centromere repositioning in mammals. *Heredity*, *108*(1), 59–67. doi: 10.1038/hdy.2011.101

Schrider, D. R., Houle, D., Lynch, M., & Hahn, M. W. (2013). Rates and genomic consequences of spontaneous mutational events in Drosophila melanogaster. *Genetics*, *194*(4), 937–954. doi: 10.1534/genetics.113.151670

Sherizen, D., Jang, J. K., Bhagat, R., Kato, N., & McKim, K. S. (2005). Meiotic recombination in Drosophila females depends on chromosome continuity between genetically defined boundaries. *Genetics*, *169*(2), 767–781. doi: 10.1534/genetics.104.035824

Shigemizu, D., Miya, F., Akiyama, S., Okuda, S., Boroevich, K. A., Fujimoto, A., … Tsunoda, T. (2018). IMSindel: An accurate intermediate-size indel detection tool incorporating de novo assembly and gapped global-local alignment with split read analysis. *Scientific Reports*, *8*(1), 5608. doi: 10.1038/s41598-018-23978-z

Sjödin, P., & Jakobsson, M. (2012). Population Genetic Nature of Copy Number Variation. *Methods in Molecular Biology*, pp. 209–223. doi: 10.1007/978-1-61779-507-7_10

Smadja, C. M., & Butlin, R. K. (2011). A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, *20*(24), 5123–5140. doi: 10.1111/j.1365-294X.2011.05350.x

Smukowski, C. S., & Noor, M. A. F. (2011). Recombination rate variation in closely related species. *Heredity*, Vol. 107, pp. 496–508. doi: 10.1038/hdy.2011.44

Stapley, J., Feulner, P. G. D., Johnston, S. E., Santure, A. W., & Smadja, C. M. (2017). Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1736). doi: 10.1098/rstb.2016.0455

Stapley, J., Santure, A. W., & Dennis, S. R. (2015). Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Molecular Ecology*, *24*(9),

2241–2252. doi: 10.1111/mec.13089

Sullivan, B. A., & Karpen, G. H. (2004). Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nature Structural & Molecular Biology*, *11*(11), 1076–1083. doi: 10.1038/nsmb845

Sung, W., Ackerman, M. S., Dillon, M. M., Platt, T. G., Fuqua, C., Cooper, V. S., & Lynch, M. (2016). Evolution of the Insertion-Deletion Mutation Rate Across the Tree of Life. *G3* , *6*(8), 2583–2591. doi: 10.1534/g3.116.030890

Talbert, P. B., Meers, M. P., & Henikoff, S. (2019). Old cogs, new tricks: the evolution of gene expression in a chromatin context. *Nature Reviews. Genetics*, *20*(5), 283–297. doi: 10.1038/s41576-019-0105-7

Talukdar, D. (2010). Reciprocal translocations in grass pea (Lathyrus sativus L.): pattern of transmission, detection of multiple interchanges and their independence. *The Journal of Heredity*, *101*(2), 169–176. doi: 10.1093/jhered/esp106

Thuillet, A.-C., Bru, D., David, J., Roumet, P., Santoni, S., Sourdille, P., & Bataillon, T. (2002). Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, Triticum turgidum (L.) Thell. ssp durum desf. *Molecular Biology and Evolution*, *19*(1), 122–125. doi: 10.1093/oxfordjournals.molbev.a003977

Torgasheva, A. A., & Borodin, P. M. (2010). Synapsis and recombination in inversion heterozygotes. *Biochemical Society Transactions*, *38*(6), 1676–1680. doi: 10.1042/BST0381676

Twyford, A. D., & Friedman, J. (2015). Adaptive divergence in the monkey flower Mimulus guttatus is maintained by a chromosomal inversion. *Evolution; International Journal of Organic Evolution*, *69*(6), 1476–1486. doi: 10.1111/evo.12663

Vendrell-Mir, P., Barteri, F., Merenciano, M., González, J., Casacuberta, J. M., & Castanera, R. (2019). A benchmark of transposon insertion detection tools using real data. *Mobile DNA*, Vol. 10. doi: 10.1186/s13100-019-0197-9

Weissensteiner, M. H., Bunikis, I., Catalán, A., Francoijs, K.-J., Knief, U., Heim, W., … Wolf, J. B. W. (2020). Discovery and population genomics of structural variation in a songbird genus. *Nature Communications*, Vol. 11. doi: 10.1038/s41467-020-17195-4

Weng, M.-L., Becker, C., Hildebrandt, J., Neumann, M., Rutter, M. T., Shaw, R. G., … Fenster, C. B. (2019). Fine-Grained Analysis of Spontaneous Mutation Spectrum and Frequency in Arabidopsis thaliana. *Genetics*, *211*(2), 703–714. doi: 10.1534/genetics.118.301721

Yamaguchi, O., & Mukai, T. (1974). Variation of spontaneous occurrence rates of chromosomal aberrations in the second chromosomes of Drosophila melanogaster. *Genetics*, *78*(4), 1209–1221. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/4218183

Zuellig, M. P., & Sweigart, A. L. (2018). A two-locus hybrid incompatibility is widespread, polymorphic, and active in natural populations of Mimulus *. *Evolution*, Vol. 72, pp. 2394–2405. doi: 10.1111/evo.13596