



**HAL**  
open science

# Towards an improved understanding of molecular evolution: the relative roles of selection, drift, and everything in between

Fanny Pouyet, Kimberbly J Gilbert

► **To cite this version:**

Fanny Pouyet, Kimberbly J Gilbert. Towards an improved understanding of molecular evolution: the relative roles of selection, drift, and everything in between. Peer Community Journal, 2021, 1, pp.e27. 10.24072/pcjournal.16 . hal-03457264

**HAL Id: hal-03457264**

**<https://hal.science/hal-03457264v1>**

Submitted on 30 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Peer Community Journal

Section: Evolutionary Biology

RESEARCH ARTICLE

Published  
2021-11-26

Cite as

Fanny Pouyet and Kimberbly J. Gilbert (2021) *Towards an improved understanding of molecular evolution: the relative roles of selection, drift, and everything in between*, Peer Community Journal, 1: e27.

Correspondence

fanny.pouyet@universite-paris-saclay.fr  
kjgilbert.evolution@gmail.com

Peer-review

Peer reviewed and recommended by PCI Evolutionary Biology, <https://doi.org/10.24072/pci.evolbiol.100103>



This article is licensed under the Creative Commons Attribution 4.0 License.

## Towards an improved understanding of molecular evolution: the relative roles of selection, drift, and everything in between

Fanny Pouyet<sup>1,2,3,4</sup> and Kimberbly J. Gilbert<sup>5,1,6,4</sup>

Volume 1 (2021), article e27

<https://doi.org/10.24072/pcjournal.16>

### Abstract

A major goal of molecular evolutionary biology is to identify loci or regions of the genome under selection versus those evolving in a neutral manner. Correct identification allows accurate inference of the evolutionary process and thus comprehension of historical and contemporary processes driving phenotypic change and adaptation. A fundamental difficulty lies in distinguishing sites targeted by selection from both sites linked to these targets and sites fully independent of selection. These three categories of sites necessitate attention in light of the debate over the relative importance of selection versus neutrality and the neutral theory. Modern genomic insights have proved that complex processes such as linkage, demography, and biased gene conversion complicate our understanding of the role of neutral versus selective processes in evolution. In this perspective, we first highlight the importance of the genomic and (a)biotic context of new mutations to identify the targets of natural selection. We then present mechanisms that may constrain the evolution of genomes and bias the inference of selection. We discuss these mechanisms within the two critical levels that they occur: the population level and the molecular level. We highlight that they should be taken into account to correctly distinguish sites across the genome subject to selective or non-selective forces and stress that a major current field-wide goal is to quantify the absolute importance of these mechanisms.

<sup>1</sup>Institute of Ecology and Evolution, University of Bern, Switzerland, <sup>2</sup>Interdisciplinary laboratory of Digital Sciences, Université Paris-Saclay, France, <sup>3</sup>Sorbonne Université, CNRS, IBPS, Laboratory of Computational and Quantitative Biology, France, <sup>4</sup>Swiss Institute of Bioinformatics, Lausanne, Switzerland, <sup>5</sup>Department of Computational Biology, University of Lausanne (UNIL), Switzerland, <sup>6</sup>Institute of Plant Sciences, University of Bern, Altenbergrain 21, 3013 Bern, Switzerland

Peer Community Journal is a member of the  
Centre Mersenne for Open Scientific Publishing  
<http://www.centre-mersenne.org/>



CENTRE  
MERSENNE

## Contents

1	Introduction .....	2
2	Genomic and environmental context .....	3
3	Population-level processes .....	6
	3.1 Population size .....	6
	3.2 Population structure .....	6
	3.3 Spatial and temporal variation .....	7
4	Molecular-level processes .....	7
	4.1 Linked selection .....	7
	4.2 Hypermutability of CpG sites and mutation rate variation .....	8
	4.3 Segregation distortion .....	9
	4.4 Mobility of genetic elements .....	9
	4.5 Gene conversion .....	10
5	Conclusion .....	11
6	Conflict of interest disclosure .....	12
7	Acknowledgements .....	12
	References .....	12

## 1. Introduction

Understanding the relative importance of evolutionary forces in driving adaptive change has been a longstanding goal of evolutionary biology. In today's genomic era, accurately and precisely addressing this question has become more feasible than ever before. Genomic data has allowed, for example, quantification of introgression rates between populations or species (e.g. Ellstrand et al., 2013) and accurate estimation of mutation rates within species or across the genome (Besenbacher et al., 2019; Ellegren et al., 2003; Hodgkinson and Eyre-Walker, 2011; Zhu et al., 2014). Yet the interplay between neutral evolution and selective forces has remained a difficult problem to address. Since the advent of population genetics as a field, debate over the relative importance of these processes has arisen, been resolved, and re-arisen (e.g. Gillespie, 1995; Kimura et al., 1968; Kreitman, 1996; Ohta and Kimura, 1971). Most recently, 50 years since the advent of the neutral theory, this debate has been rekindled in light of emerging genomic data (Jensen et al., 2018; Kern and Matthew W Hahn, 2018). In an era of limited genetic tools and data, the neutral theory aimed to explain the greater than expected genetic diversity observed based on the actions of natural selection alone. Kern and Hahn (2018) have most recently argued that modern genomic data allows us to reject the applicability of neutral theory for understanding molecular evolution, while Jensen et al. (2018) have replied that this is not the case. A major dividing view on this point is whether a large proportion of the genome is affected by adaptive natural selection (directly or indirectly), and there is ample space for additional data across a wider range of species to contribute towards these investigations and our understanding of molecular evolution.

Natural selection functions in a diversity of modes. Negative selection – also termed purifying selection – acts to reduce the frequency of deleterious mutations (*i.e.* mutations that reduce an individual's fitness, with selection coefficient  $s < 0$ ) while positive selection favors the fixation of beneficial mutations ( $s > 0$ ). Both these modes of selection reduce diversity by favoring or disfavoring specific alleles, but selection can also maintain genetic diversity when there is a selective advantage to being in the heterozygous state, *e.g.* balancing selection. There is also the case of sexual selection that we do not address here as its many and complex cases merit a review of their own.

Genetic drift is the neutral corollary to natural selection, where allele frequencies change due to random chance and sampling effects. Sites that are truly neutral are defined as those with  $s = 0$ . If we define the effective size of a population as the size of an ideal population that

experiences the same amount of genetic drift as the observed population (Wright, 1931), then the appropriate measure of the strength of selection is ( $Ns$ ). This is key because genetic drift can act across a range of weakly selected sites when  $N$  is sufficiently large, while at a smaller  $N$ , these sites may behave in a neutral manner.

Inference methods making use of empirical genomic data often require data solely originating from neutral processes, e.g. to infer demography with SFS-based methods such as FastSimCoal (Excoffier et al., 2013), or to infer the distribution of fitness effects (DFE) which requires contrasting SFS from neutral versus selected sites (Keightley and Eyre-Walker, 2007; Tataru and Bataillon, 2019). *A priori* misidentification of selected versus neutral sites may strongly bias resultant inferences, having a major impact on downstream interpretations. Approaches that search for signatures of selection and identify causal variants for adaptation or other phenotypic change mainly rely on identifying outlier regions of genomic differentiation (e.g. GWAS,  $F_{ST}$  outlier tests, or environment-allele correlations; Beaumont and Nichols, 1996; Foll et al., 2008; Luu et al., 2017; Whitlock and Lotterhos, 2015), sometimes incorporating the signature across a stretch of the genome (e.g. Schrider and Kern, 2016).

Yet, a departure from genetic drift alone is not sufficient to merit a conclusion of selection. Even supervised machine learning methods that use summary statistics to infer a history of selective sweeps, such as S/HiC (Schrider and Kern, 2016), are sensitive to confounding factors such as complex demographic history to accurately identify the variants under selection. In this review we highlight both the importance of considering the genomic, biotic, and abiotic context in which new mutations occur and the major evolutionary processes that can change allele frequencies, creating a major confounding factor for evolutionary inference of natural selection.

## 2. Genomic and environmental context

The context in which mutations occur plays an important role in the actions of selection versus drift. This context encapsulates both the genomic environment as well as the biotic and abiotic environment of an organism containing those genes, therefore becoming relevant at both the molecular and population levels of interaction. When interactions between loci occurring together in the genome create non-additive phenotypic changes (Cordell, 2002; Fisher, 1918; G Martin et al., 2007) this can greatly complicate the inference of selection. In the presence of epistatic interactions, an allele at a given site may only be beneficial to an organism when its genomic environment contains another mutant allele at a different site in the genome, so that when together these alleles generate a phenotypic change. The concept of epistasis is tightly linked to the relative fitness effects of alleles, where a specific allele at one locus might change the sign of the selection coefficient at another locus. This phenomenon may create strong patterns of linkage disequilibrium and interfere with detection of selection since the selective effect is dependent on the combinations of alleles across loci. Such an example of epistasis is provided in the Segregation distortion paragraph.

In some cases, interference between selected sites of differential fitness effects can alter the strength of selection on a genomic region. More positively (or negatively) selected sites with physical linkage between each other can behave as a larger multi-site locus with an amplified selection coefficient representative of all the selected alleles in the region. In other cases, the strength of selection may be reduced when sites have competing impacts on fitness, termed Hill-Robertson interference (Felsenstein, 1974). Amplification of the strength of selection by tightly-linked, jointly selected sites may simplify detection of selection but complicate the identification of precise sites under selection. Conversely, Hill-Robertson interference may complicate identifying both the presence of selection and the sites it targets. All of these effects depend on many parameters, for instance, with a single population undergoing partial self-fertilization, selective interference on deleterious alleles tends to reduce mean fitness and increase inbreeding depression. This effect is stronger when deleterious alleles are more recessive but only weakly dependent on the strength of selection against deleterious alleles and the recombination rate (Roze, 2015). Selective interference thus affects the relative impact of adaptive and non-adaptive processes in the genome.

Equally complicating is the individual-level scenario where fitness is dependent on the local community of organisms and whatever traits they exhibit (e.g. frequency-dependent selection) or on the abiotic local conditions exerted by variable environments on the phenotype (spatially or temporally varying selection). For example, in scale-eating cichlids, frequency-dependent selection can drive the handedness of individual phenotypes, where it is advantageous to be the rarer morph (Hori, 1993). Interestingly, competition between alleles at the genomic level can also lead to frequency-dependent selection, with the most famous example being the amino acid polymorphism at position 6 of  $\beta$ -globin in Africa, which is associated with resistance to malaria (see Taylor et al., 2012, for review and meta-analyses). Spatially varying selection results from environmental variation across geographic space (e.g. Gagnaire et al., 2012). When populations exist across variable environments, distinct combination of alleles may arise within subpopulations locally adapted to their environments. Thus, the strength of selection on non-neutral sites can vary over time and space as an organism's environment changes. For the purposes of evolutionary inference and understanding the action of selection, experimental design is key so that empirical analyses can be conducted where the time point is equivalent and the environment is equivalent (or otherwise controlled for) to as much of an extent as possible (e.g. Gorter et al., 2018). Spatially and temporally varying selection may be best accounted for by this direct approach or by studying natural clines of allele frequencies (Endler, 1977; Machado et al., 2016). However, for sites in the genome that are not targets of selection, nor purely under genetic drift, remaining processes can change allele frequencies in ways indicative of selection and are thus essential to bear in mind for evolutionary inference.

In the remainder of this perspective, we consider evolutionary processes at both the population and molecular levels that have the potential to bias the inference of selection. Without consideration of such processes, inaccurate conclusions may be drawn about the role of selection in the evolutionary process. An extreme example of such mis-inference is Evans et al. (2005) and Mekel-Bobrov et al. (2005), where the result that selection drove brain size in humans was shown to be equivalently explained by neutral demographic processes (Currat et al., 2006). Scenarios involved at the molecular level can also interfere with correct inference of selection, including transmission bias (e.g. meiotic driver genes; Bravo Núñez et al., 2018, for a review), biased gene conversion (e.g. human accelerated genomic regions, HAR; N. Galtier and Duret, 2007; Pollard et al., 2006), or inference of the DFE (e.g. in flycatchers; Bolivar et al., 2018).

Though these are few examples, we argue that these population and molecular processes are prevalent enough to act as major determinants of genomic diversity and are particularly easy to confound with selection. We also argue that even if some cases of demographic processes may be versions of genetic drift, in that they are due to sampling process, the signatures that they leave in populations differs from that of single, ideal populations undergoing drift, emphasizing their importance in distinguishing selection from neutrality. We define these as non-adaptive processes since they are neither the direct action of selection nor are they purely subject to drift (Table 1). This non-adaptive category should help to improve studies of both neutral or selective processes impacting the genome by considering the relevant sites impacted by either process.

	Acts on neutral variants?	Non-adaptive?	Explanation
Selection Negative Positive Balancing Sexual			Adaptive processes increase fitness
Genetic drift	yes	yes	Neutral and non-adaptive as independent of fitness; depending on $N$ , may impact site(s) with $s \neq 0$
<b>Population-level processes</b>			
Bottleneck, demographic expansion	yes	yes	Lead to rapid genetic drift, can mimic positive or negative selection
Population structure	yes	yes	Can mimic local adaptation, responsible for many GWAS false-positives
Spatial expansion (gene surfing)	yes	yes	Strong genetic drift over space, can mimic selection
<b>Molecular-level processes</b>			
Background selection	yes	yes	Impact neutral diversity, can change fitness, and act on selected sites depending on the relative selection coefficients of nearby sites
Linked selection	yes	yes	
Selective sweeps	yes	yes	
Associative overdominance	yes	yes	
CpG hyper-mutability	yes	yes	Methylated Cytosines deaminate to Thymines, increasing the frequency of T within the genome
Segregation distortion	yes	yes	Meiotic gene drivers manipulate their transmission even though they can be detrimental
Gene conversion	yes	yes	Increases the frequency of certain variants deterministically; regardless of fitness

**Table 1** – Evolutionary processes discussed in this paper that impact the fate of genetic variants. We focus on SNPs in this table and omit a category on the mobility of genetic elements, but transposable elements are also impacted by molecular-level genomic linkage processes.

### 3. Population-level processes

#### 3.1. Population size.

A major difficulty in distinguishing neutral sites from those impacted by selection is demographic history. The term demography incorporates several factors, but at its core is defined as change in population size,  $N$ . Population size combined with the selection coefficient determines the effective strength of selection on existing genomic diversity and therefore has major effects on the evolutionary process. Changes in population size result during population bottlenecks, population expansions (within one locale), spatial population expansions over geographic space (e.g. range expansion), or migration among populations (a more complicated case where a larger gene pool becomes relevant).

While selection or drift may act on specific variants or regions of the genome, demographic change affects the whole genome equally. Population bottlenecks have long been known to impact genetic diversity and change the efficiency of selection acting on alleles with  $s \neq 0$  within the population. In such a case, stronger drift impacts sites of both  $s = 0$  and  $s$  close to zero (nearly neutral mutations), with an increasing range of  $s$  values as population size decreases and selection becomes less efficient. These alleles are driven to more rapid fixation or more rapid loss, a pattern of allele frequency change which can mimic that of positive selection and selective sweeps (see the Linked selection paragraph).

Inferring past population size bottlenecks is a rich field with many methods to do so from genomic data (Heled and Drummond, 2008; Li and Durbin, 2011; Liu and Fu, 2015; Terhorst et al., 2017). Importantly, these methods rely on the use of neutral variants to obtain a proper inference (Gattepaille et al., 2013), and are perhaps particularly important in conservation genetics to identify species at risk due to a recent bottleneck (rather than, e.g. an incorrect inference suggesting low diversity is due to a selective sweep; Peery et al., 2012). The bias that results from a demographic history of range expansion after a bottleneck has been particularly notable for humans having expanded out of Africa and more troublesome for distinguishing demography from selection. Studies attempting to find signatures of selection in humans may suffer from biased inferences due to these neutral historic processes (Amos and Bryant, 2011; Heller et al., 2013; AR Martin et al., 2017). Such a demographic history is particularly intriguing as it combines not only impactful changes in population size, but movement over geographic space which includes complications of population structure and spatially varying selection.

#### 3.2. Population structure.

Spatial population genetic structure and migration among subpopulations also plays an important role in the inference of selection. Population substructure can mimic a signal of local adaptation, where some populations which happen to exist in different environments possess different genetic signatures, leading methods to identify these differentiated loci as targets of selection. For instance, several recent studies have encountered this difficulty where signals previously thought to be selective were instead due to the lack of accounting for genetic structure among populations (Berg et al., 2019; Tian et al., 2008). Additionally, the process of migration into populations or admixture among species can create an influx of novel genetic material. Even if fully neutral, the presence of such heterozygosity in the population leaves a signal indicative of either adaptive processes (e.g. balancing selection) or non-adaptive processes (e.g. secondary contact or gene flow among structured populations) (Matthew William Hahn, 2018).

It is again of vital importance for studies inferring selection that population structure be identified and accounted for. Many such approaches exist and vary depending on the form in which structure presents itself: isolation by distance versus more distinct populations over space, with varying degrees of migration occurring across the landscape. For example, isolation-with-migration models aim to infer the amount of migration between isolated populations leading to the level of polymorphism observed (Hey, 2010). Isolation by distance can also be difficult in the face of inferring selection due to the correlation of allele frequency changes over space

with environmental changes. Fortunately, much work has been done to correct for this population structure when inferring selection over landscapes (Caye et al., 2019; de Villemereuil and Gaggiotti, 2015; Gautier, 2015; Günther and Coop, 2013)

### 3.3. Spatial and temporal variation.

Spread or growth of populations across geographic space also introduces the complexity of changing environmental conditions (or analogously temporally changing environments may have similar impacts). Many populations and species are known to have undergone or are expected to undergo this demographic change: from post-glacial recolonizations, to species invasions, to shifting species ranges in response to climate change (Davis and Shaw, 2001; Thomas, 2010). During spatial expansions, not only does population size change with repeated bottlenecks of founder individuals, but these populations colonize new geographic space, resulting in a process termed gene surfing. Gene surfing is a unique genetic process that can leave genomic signatures similar to those of selection, yet are due entirely to demographic processes. Sequential founder events reduce the effective population size in colonizing populations and thus the efficiency of selection, thereby allowing alleles that might otherwise be subject to strong selection to surf to high frequency at the expanding wave front of a population (Edmonds et al., 2004; Klopstein et al., 2005). Because surfing can lead to the increase or even fixation of a given allele (be it neutral or not), it is easily mistaken for the product of selective forces. Yet unlike selection, surfing can also cause deleterious variants to increase and result in severe fitness loss at expanding fronts, termed expansion load (Peischl et al., 2013). This demographic process alters the actions of natural selection and genetic drift within the genome and has potentially large effects on population fitness, emphasizing its importance as a non-adaptive force in evolution.

## 4. Molecular-level processes

### 4.1. Linked selection.

The fact that recombination breaks apart combinations of alleles at an increasing probability with greater distance along the genome results in many sites being physically linked and evolving in a non-independent manner. The background where a new mutation occurs therefore influences that variant's probability of fixation, as any more strongly-selected target sites nearby will influence that linked site, as first pointed out by Fisher (1930) and Muller (1932) (Figure 1). A extended review on this topic was written by Gordo and Charlesworth (2001).

Neutral sites in a background with one or more sites under negative selection will have a lower probability of fixation than unlinked neutral sites. This is due to background selection (BGS), where negative selection against a variant reduces the frequency of nearby neutral variants (B. Charlesworth et al., 1993). Sites subject to BGS fall in the category of non-adaptive evolution because these variants are not directly selected against nor are they evolving neutrally since selection indirectly impacts them. These linked sites evolve in a non-neutral fashion, so even if phenotypically and adaptively they confer no change in phenotype, they must not be considered neutral for inferential purposes, a point which is widely recognized in the field.

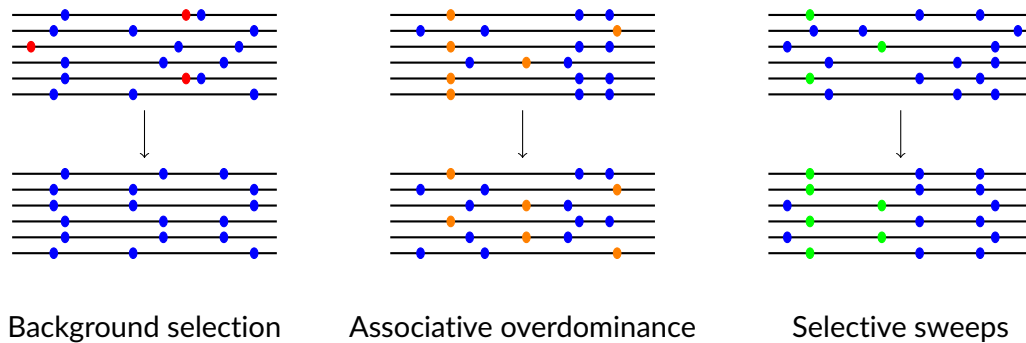
Similarly, the occurrence of a mutation conferring a fitness benefit can also result in the reduction of genetic diversity through a selective sweep (Figure 1). When selection increases the frequency of a beneficial allele in the population, nearby neutral variants likewise increase in frequency, hitchhiking along to fixation with the beneficial variant. Whether selection acts on a single novel variant (hard sweeps; Messer and Neher, 2012) or on standing genetic variation (soft sweeps; Hermisson and Pleuni S. Pennings, 2005; Pleuni S Pennings and Hermisson, 2006a,b) can influence the extent of the impact on allele frequency change for linked neutral sites. Several population level processes may even be contributors to instances where standing genetic variation results in a soft sweep, for example if existing diversity shifts to become beneficial, perhaps due to environmental change.

Finally, genetic linkage can also lead to an increase in genetic diversity when neutral sites fall near a partially deleterious recessive allele or near an allele under balancing selection. In the presence of partially recessive deleterious alleles, this increase in diversity is termed associative



overdominance (AOD; Becher et al., 2020; Gilbert et al., 2020; Ohta and Kimura, 1970; Zhao and Brian Charlesworth, 2016), and is limited to regions of low recombination (Figure 1). In contrast, sites linked to those under classical balancing selection should increase in diversity across regions spanning the range of possible recombination rates.

Most methods to detect selection on linked polymorphisms are based on the site frequency spectrum (SFS), such as Tajima's D which depends on the pairwise nucleotide diversity and the number of segregating sites (Tajima, 1989) and other methods which are based on the haplotype structure (Hudson and Kaplan, 1995; Sabeti, Reich, et al., 2002; Sabeti, Varilly, et al., 2007; Voight et al., 2006). Methods based on haplotype structure are most effective to detect recent episodes of hitchhiking (Garud et al., 2015).



**Figure 1** – Genetic linkage can drastically change the frequency of neutral alleles in a population, falling into three categories depending on the manner of selection towards a focal site. For each form of selection shown, the top row shows the haplotypes of a non-recombining region in the initial population. The bottom row shows the resultant haplotypes after an episode of selection. For illustrative purpose, each haplotype contains 3 derived alleles (circles): neutral ones are in blue, beneficial in green, and deleterious alleles are in red or orange to indicate dominant or recessive, respectively. For background selection (left), neutral diversity is reduced due to negative selection on nearby linked deleterious alleles and homozygosity increases at the population level. Associative overdominance (center) prevents combinations of homozygous neutral alleles from accumulating in the population in regions of low recombination. At the beginning, each haplotype contains one deleterious recessive allele (orange circle) and loci carrying such alleles are genetically linked as the region is non-recombining. Selection favors combinations of heterozygous deleterious alleles as they are recessive. Neutral heterozygosity is favored at the population level and diversity increases. Selective sweeps (right) reduce diversity and increase allele frequency through the hitchhiking of neutral variants that are linked to beneficial mutations under positive selection. Figure inspired from Alves et al. (2012).

#### 4.2. Hypermutable of CpG sites and mutation rate variation.

CpG sites in which a cytosine and a guanine appear consecutively, can experience high levels of mutational pressure. Cytosines at CpG sites are one of the preferential targets of methylation in vertebrates and some other species. Methylated cytosines spontaneously deaminate to thymines, leading to an increase in the frequency of TpG sites within the genome and a relative deficit of CpG (reviewed in Hodgkinson and Eyre-Walker, 2011), potentially leaving a signature indicative of selection.

A recent paper by Laurin-Lemay et al. (2018) found that a large proportion of mammalian codon usage, such as the preferential usage of the GCC Alanine codon compared to its synonymous GCG in humans, can be explained by the hypermutability of CpG sites, even though this is often unaccounted for in codon substitution models. The authors advocate for evaluating the impact of such model violations on statistical tests in phylogenetic analyses. Interestingly, CpG hypermutability is also an underappreciated process in the field of population genomics where

the favored strategy has been to filter out hypermutable sites before performing evolutionary inferences (Pouyet, Aeschbacher, et al., 2018). The hypermutability of these sites and the subsequent bias to certain alleles has been shown to shift site frequency spectra in ways that might interfere with population genetic inferences that are based on the SFS (Harpak et al., 2016).

### 4.3. Segregation distortion.

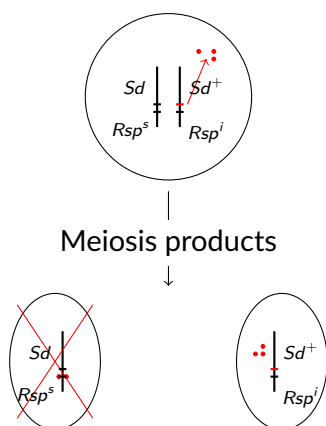
Meiotic gene drivers manipulate the transmission process during meiosis to their own advantage, leading to their over-representation in gametes despite the lack of advantage to the carrier. As segregation distortion encompasses a wide variety of mechanisms, we provide three such examples. The first one is the segregation distorter gene complex (Figure 2; Larracuenta and Presgraves, 2012), present at low frequency in all natural populations of *Drosophila melanogaster*. This complex involves the *Sd* locus and its target responder locus (*Rsp*). There is variable transmission advantage between the different *Sd* alleles, and one of these alleles recently swept to fixation in Africa causing strong linkage disequilibrium and loss of genetic diversity (Presgraves et al., 2009). A second example is *Wolbachia*, where a maternally inherited bacteria of arthropods manipulates host reproduction. *Wolbachia* is known to be a selfish element favoring its own propagation through, for instance, the inability of infected males to successfully reproduce with uninfected females (Turelli and Hoffmann, 1991). At the level of the host, the presence of *Wolbachia* is also associated to segregation distortion: the maternal inheritance induces genetic linkage on host mitochondria and mitochondria in infected females are over-represented in the next generation. This effect reduces the effective population size and the efficacy of selection in mitochondria and could drive fixation of mitochondrial haplotypes (Cariou et al., 2017; GD Hurst and Jiggins, 2005). A third example comes from centromere evolution, where asymmetry at female meiosis causes only one of the four products of meiosis to become the oocyte nucleus and can lead to a kind of segregation distortion (Henikoff et al., 2001). Centromeres have a central role in preventing aneuploidy by facilitating the assembly of several components required for chromosome separation (see Tanaka et al., 2013, for a review). This centromere drive model includes proteins such as Cid and highly repetitive satellite sequences that bind to microtubules during meiosis I (Henikoff et al., 2001). Centromeres which preferentially transmit to the oocyte nucleus can rapidly drive to fixation even with a slight advantage at each meiosis (Henikoff et al., 2001).

Meiotic drivers are predicted to be evolutionarily labile by favoring their fixation in the population even though they are detrimental for their carriers (Lindholm et al., 2016). Identifying loci under selection in these cases is anything but straightforward, as the signal of meiotic drive might be easily confounded with selective sweeps and positive selection (Presgraves et al., 2009).

### 4.4. Mobility of genetic elements.

Detecting the signature of selection is often restricted to SNPs even though transposable elements (TE) and other structural genomic changes such as inversions are likely also prevalent biases for adaptive evolution. Until further studies improve our understanding of mobile genetic elements and our ability to identify them, it is certainly possible that these regions of the genome may play a role in changing diversity and biasing or interrupting our ability to infer selection or neutral demographic parameters, e.g. in inverted regions that can no longer recombine, deleterious variation can become masked or beneficial variation may be maintained in tight linkage.

TEs are widespread across the tree of life and in some species can represent a major fraction of the genome (de Koning et al., 2011; Wicker et al., 2007). TEs are associated with the creation of new mutations and changes in recombination patterns (e.g. Bartolome et al., 2002). McClintock (1950) first discovered that mobile elements were associated with phenotypic changes in maize. However, these sorts of structural genomic changes are often disregarded in favor of SNPs when inferring sources of adaptive evolution mainly because of methodological limitations (Villanueva-Canas et al., 2017). First, TE families are difficult to identify as they are repetitive elements spread throughout the genome with limited descriptive features such as target site duplications or terminal repeats and transposases (Xiong et al., 2014). They are also commonly associated with genetic load as they can lead to diseases if inserted into genes (Chen et al.,



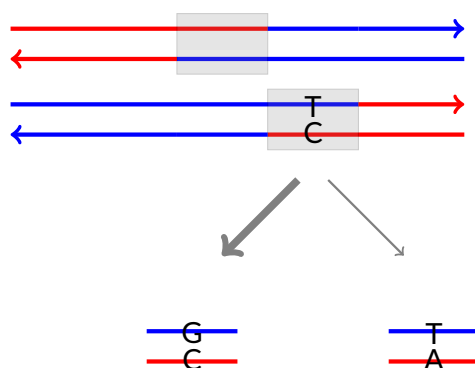
**Figure 2** – Figure simplified from Bravo Núñez et al. (2018). An example of segregation distortion of meiotic gene drivers with the “killer-target” strategy is shown for the SD gene complex in *Drosophila melanogaster*. This complex involves the *Sd* and *Rsp* loci. The “target” locus *Rsp* harbours two alleles: the *Rsp<sup>s</sup>* and *Rsp<sup>i</sup>* that are respectively sensitive and insensitive to the meiotic driver. *Sd* produces a “killer” element (red dots, a protein in this example) which interferes with *Rsp<sup>s</sup>* and kills the meiotic products that inherit *Rsp<sup>s</sup>*.

2005). For instance, the Alu family of insertions are associated to haemophilia or breast cancer in humans (Batzner and Deininger, 2002). Second, because repetitive elements are enriched for TEs, it is difficult to assemble these genomic regions (Bourgeois and Boissinot, 2019). For the past decade, new techniques have emerged to infer selection on TE insertions and are divided into two main classes: SFS-based or haplotype-based methods (Villanueva-Canas et al., 2017). It is clear that as our understanding of TE dynamics improves, such knowledge may greatly contribute to our understanding of selection acting at the genomic level and modes of adaptive evolution.

#### 4.5. Gene conversion.

Meiotic recombination reshuffles the genetic material of parents to produce a new set of genetic material in offspring. During recombination, homologous gene conversion can result from the conversion of an acceptor locus at heterozygous sites in donor sequences. Biased gene conversion (BGC), occurring at locations where recombination breaks the DNA strand, makes the probability of transmitting one of the two alleles larger than the probability of losing it. BGC is comprised of two main mechanisms: double-strand-break-driven (dBGC; Myers et al., 2010; Paigen and Petkov, 2018) and GC-driven (gBGC; Duret and N. Galtier, 2009; Lesecque et al., 2013), each of which have different mechanistic origins and consequences. To our knowledge dBGC is not expected to be confounded with selection and will not be discussed further herein. On the contrary, gBGC is often responsible for false positives in inferences of selection (Nicolas Galtier et al., 2009; Ratnakumar et al., 2010).

gBGC is a transmission bias in favor of G/C over A/T alleles when a mismatch is repaired after a meiotic recombination event. This leads to the increase of GC content in regions of high recombination over evolutionary time (Duret and N. Galtier, 2009) (Figure 3). Evidence for gBGC has been shown in many organisms (Duret and N. Galtier, 2009; Lassalle et al., 2015; Lesecque et al., 2013; Webster et al., 2006) and has strong consequences on genomic architecture, ranging from global GC-enrichment of genomes to variation in codon usage between genes (Pouyet, Mouchiroud, et al., 2017), and can be confounded with translational selection (Gingold et al., 2014). When G/C alleles are beneficial as compared to A/T, gBGC amplifies the speed of fixation of a beneficial allele in the population, while in other cases when G/C alleles are slightly deleterious, gBGC counteracts the effect of selection (if the transmission bias parameter,  $b$ , is stronger than the selection parameter,  $s$ ). At the population level, gBGC shifts the site frequency spectrum towards the left for strong-to-weak polymorphisms (respectively to the right for weak-to-strong) mimicking the effect of natural selection in high recombination rate regions (Glémin et al., 2015;



**Figure 3** – Figure adapted from Glémin et al. (2014). GC-biased gene conversion (gBGC) occurs after the formation of hetero-duplexes during a meiotic recombination event. Heteroduplexes can result from crossing-over (shown here) as well as from non-crossover events. In the presence of gBGC, mismatches are repaired more than half of the time in favor of guanine and cytosine rather than adenine and thymine. This bias towards enrichment of guanine and cytosine is of few percent in yeast or humans, leading to the accumulation of GC content within genomes over time.

Lachance and Tishkoff, 2014; Pouyet, Aeschbacher, et al., 2018). One way to consider the consequences of gBGC is to contrast genetic diversity of a region between weak-to-strong and strong-to-weak mutations. Additionally, gBGC should not leave the same signature as linked selection because gBGC acts solely on its targets sites and does not affect the surrounding diversity. There is no inherent selective bias driving the genomic changes resulting from gene conversion, nor can a biased process by definition be considered neutral. Instead, this process is best considered as non-adaptive evolution, where fitness is not impacted, but there is clearly a non-neutral change in allele frequencies over time. As illustrated by Hurst (2019), segregation distortion and BGC are rarely considered simultaneously even though they act similarly to increase their transmission to the detriment of other alleles. Both BGC and segregation distortion along with structural genomic changes are common in nature, and including these processes in studies of adaptive evolution will allow us to properly identify targets of positive selection (Villanueva-Canas et al., 2017).

## 5. Conclusion

In the modern genomic era, emerging data will hopefully allow us to build a complete picture of the relevant genomic, demographic, and environmental scenarios where different evolutionary processes are expected to dominate changes in molecular diversity over time. Identifying a variant as subject to natural selection is difficult since the selective environment of an allele is a combination of its own innate properties impacting the genome, along with epistatic effects due to its genomic environment, as well as its demographic situation (how efficient selection is in the population where this individual exists), and lastly its (a)biotic environment (e.g. stressful environments for the organism harboring this variant, or the frequency of conspecific phenotypes). Understanding the processes that may bias our inferences of sites under selection is paramount to better understanding the evolutionary forces leading to genomic change. To a large extent these biases result from the difficulty or inability to distinguish non-adaptive sites from sites under direct selection. As discussed, this is largely due to the demographic processes in a population's past that make otherwise neutral sites mimic selection and fall into the category of non-adaptive, as well as due to the molecular processes that change neutral allele frequencies in biased ways and can even counteract the effect of selection on selected sites (gBGC, for instance). There is an implied distinction worth explicitly stating: even if the majority of sites within the genome may be neutral in terms of their selection coefficient, it is very likely the case that the majority of the genome evolves due to the impact of selective forces, even if that targets few specific sites, due to the degree of linkage within the genome. This still does not, however, discount the fact

that non-adaptive evolutionary processes have an important impact on genomic change. Incorporating all of this information in future studies is a tall task, particularly since empirical study of biology is further complicated by changing environments and demographics that are not always apparent to observers, nor always sufficiently sampled. We hope that this perspective has highlighted the importance of recognizing and distinguishing the complex interactions of selective, non-adaptive, and neutral processes acting within and among genomes and serves to move the field of evolutionary genomics forward in understanding the drivers of molecular diversity.

## 6. Conflict of interest disclosure

The authors of this article declare that they have no financial conflict of interest with the content of this article.

## 7. Acknowledgements

We would like to thank Benoit Nabholz, an anonymous reviewer, and Guillaume Achaz for critical feedback, as well as Tyler Kent, Stephan Peischl, and Marie Cariou for their input on earlier versions, all of which greatly improved this manuscript.

A previous version of this article has been peer-reviewed and recommended by Peer Community In Evol Biol (<https://doi.org/10.24072/pci.evolbiol.100103>).

## References

- Alves I et al. (July 2012). *Genomic Data Reveal a Complex Making of Humans*. *PLoS genetics* **8**, e1002837. <https://doi.org/10.1371/journal.pgen.1002837>.
- Amos W, Bryant C (2011). *Using human demographic history to infer natural selection reveals contrasting patterns on different families of immune genes*. *Proceedings of the Royal Society B: Biological Sciences* **278**, 1587–1594. <https://doi.org/10.1098/rspb.2010.2056>.
- Bartolome C et al. (2002). *On the Abundance and Distribution of Transposable Elements in the Genome of Drosophila melanogaster*. *Molecular Biology and Evolution* **19**, 926–937. <https://doi.org/10.1093/oxfordjournals.molbev.a004150>.
- Batzer MA, Deininger PL (2002). *Alu repeats and human genomic diversity*. *Nature Reviews Genetics* **3**, 370–379. <https://doi.org/10.1038/nrg798>.
- Beaumont MA, Nichols RA (1996). *Evaluating loci for use in the genetic analysis of population structure*. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **263**, 1619–1626. <https://doi.org/10.1098/rspb.1996.0237>.
- Becher H et al. (2020). *Patterns of genetic variability in genomic regions with low rates of recombination*. *Current Biology* **30**, 94–100. <https://doi.org/10.1016/j.cub.2019.10.047>.
- Berg JJ et al. (2019). *Reduced signal for polygenic adaptation of height in UK Biobank*. *Elife* **8**, e39725. <https://doi.org/10.7554/eLife.39725>.
- Besenbacher S et al. (2019). *Direct estimation of mutations in great apes reconciles phylogenetic dating*. *Nature Ecology and Evolution* **3**, 286–292. <https://doi.org/10.1038/s41559-018-0778-x>.
- Bolivar P et al. (2018). *Biased Inference of Selection Due to GC-Biased Gene Conversion and the Rate of Protein Evolution in Flycatchers When Accounting for It*. *Molecular Biology and Evolution* **35**, 2475–2486. <https://doi.org/10.1093/molbev/msy149>.
- Bourgeois Y, Boissinot S (2019). *On the Population Dynamics of Junk: A Review on the Population Genomics of Transposable Elements*. *Genes* **10**, 419. <https://doi.org/10.3390/genes10060419>.
- Bravo Núñez MA et al. (2018). *Genetic Villains: Killer Meiotic Drivers*. *Trends in Genetics* **34**, 424–433. <https://doi.org/10.1016/j.tig.2018.02.003>.
- Cariou M et al. (2017). *The global impact of Wolbachia on mitochondrial diversity and evolution*. *J Evol Biol*. **30**, 2204–2210. <https://doi.org/10.1111/jeb.13186>.
- Caye K et al. (2019). *LFMM 2: fast and accurate inference of gene-environment associations in genome-wide studies*. *Molecular biology and evolution* **36**, 852–860. <https://doi.org/10.1093/molbev/msz008>.

- Charlesworth B et al. (1993). *The effect of deleterious mutations on neutral molecular variation*. *Genetics* **134**, 1289–1303. <https://doi.org/10.1093/genetics/134.4.1289>.
- Chen JM et al. (2005). *Meta-Analysis of gross insertions causing human genetic disease: Novel mutational mechanisms and the role of replication slippage*. *Human mutation* **25**, 207–221. <https://doi.org/10.1002/humu.20150>.
- Cordell HJ (2002). *Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans*. *Human Molecular Genetics* **11**, 2463–2468. <https://doi.org/10.1093/hmg/11.20.2463>.
- Currat M et al. (2006). *Comment on "Ongoing Adaptive Evolution of ASPM, a Brain Size Determinant in Homo sapiens" and "Microcephalin, a Gene Regulating Brain Size, Continues to Evolve Adaptively in Humans"*. *Science* **313**, 172. <https://doi.org/10.1126/science.1122712>.
- Davis MB, Shaw RG (2001). *Range shifts and adaptive responses to Quaternary climate change*. *Science* **292**, 673–679. <https://doi.org/10.1126/science.292.5517.673>.
- de Koning APJ et al. (2011). *Repetitive Elements May Comprise Over Two-Thirds of the Human Genome*. *PLOS Genetics* **7**, 1–12. <https://doi.org/10.1371/journal.pgen.1002384>.
- de Villemereuil P, Gaggiotti OE (2015). *A new FST-based method to uncover local adaptation using environmental variables*. *Methods in Ecology and Evolution* **6**, 1248–1258. <https://doi.org/10.1111/2041-210X.12418>.
- Duret L, Galtier N (2009). *Biased gene conversion and the evolution of mammalian genomic landscapes*. *Annual Review of Genomics and Human Genetics* **10**, 285–311. <https://doi.org/10.1146/annurev-genom-082908-150001>.
- Edmonds CA et al. (2004). *Mutations arising in the wave front of an expanding population*. *Proceedings of the National Academy of Sciences* **101**, 975–979. <https://doi.org/10.1073/pnas.0308064100>.
- Ellegren H et al. (2003). *Mutation rate variation in the mammalian genome*. *Current opinion in genetics & development* **13**, 562–568. <https://doi.org/10.1016/j.gde.2003.10.008>.
- Ellstrand NC et al. (2013). *Introgression of crop alleles into wild or weedy populations*. *Annual Review of Ecology, Evolution, and Systematics* **44**, 325–345. <https://doi.org/10.1146/annurev-ecolsys-110512-135840>.
- Endler J (1977). *Geographic Variation, Speciation and Clines*. Monographs in population biology.
- Evans PD et al. (2005). *Microcephalin, a Gene Regulating Brain Size, Continues to Evolve Adaptively in Humans*. *Science* **309**, 1717–1720. <https://doi.org/10.1126/science.1113722>.
- Excoffier L et al. (2013). *Robust demographic inference from genomic and SNP data*. *PLoS Genetics* **9**, e1003905. <https://doi.org/10.1371/journal.pgen.1003905>.
- Felsenstein J (1974). *The evolutionary advantage of recombination*. *Genetics* **78**, 737–756. <https://doi.org/10.1093/genetics/78.2.737>.
- Fisher RA (1918). *The Correlation Between Relatives on the Supposition of Mendelian Inheritance*. *Trans. R. Soc. Edin.*, 399–433. <https://doi.org/doi:10.1017/S0080456800012163>.
- Fisher RA (1930). *The Genetical Theory of Natural Selections*. Oxford: The Clarendon Press. <https://doi.org/10.1016/j.gde.2014.09.005>.
- Foll M et al. (2008). *An approximate Bayesian computation approach to overcome biases that arise when using amplified fragment length polymorphism markers to study population structure*. *Genetics* **179**, 927–939. <https://doi.org/10.1534/genetics.107.084541>.
- Gagnaire PA et al. (2012). *The Genetic Consequences of Spatially Varying Selection in the Panmictic American Eel (Anguilla rostrata)*. *Genetics* **190**, 725–736. <https://doi.org/10.1534/genetics.111.134825>.
- Galtier N, Duret L (2007). *Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution*. *Trends in Genetics* **23**, 273–277. <https://doi.org/10.1016/j.tig.2007.03.011>.
- Galtier N et al. (2009). *GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates*. *Trends in Genetics* **25**, 1–5. <https://doi.org/10.1016/j.tig.2008.10.011>.

- Garud NR et al. (2015). *Recent Selective Sweeps in North American Drosophila melanogaster Show Signatures of Soft Sweeps*. *PLOS Genetics* **11**, 1–32. <https://doi.org/10.1371/journal.pgen.1005004>.
- Gattepaille LM et al. (2013). *Inferring population size changes with sequence and SNP data: lessons from human bottlenecks*. *Heredity* **110**, 409–419. <https://doi.org/10.1038/hdy.2012.120>.
- Gautier M (2015). *Genome-wide scan for adaptive divergence and association with population-specific covariates*. *Genetics* **201**, 1555–1579. <https://doi.org/10.1534/genetics.115.181453>.
- Gilbert KJ et al. (2020). *Transition from background selection to associative overdominance promotes diversity in regions of low recombination*. *Current Biology* **30**, 101–107. <https://doi.org/10.1101/748004>.
- Gillespie JH (1995). *On Ohta's hypothesis: most amino acid substitutions are deleterious*. *Journal of molecular evolution* **40**, 64–69. <https://doi.org/10.1007/BF00166596>.
- Gingold H et al. (2014). *A dual program for translation regulation in cellular proliferation and differentiation*. *Cell* **158**, 1281–92. <https://doi.org/10.1016/j.cell.2014.08.011>.
- Glémin S et al. (2015). *Quantification of GC-biased gene conversion in the human genome*. *Genome research* **25**, 1215–1228. <https://doi.org/10.1101/gr.185488.114>.
- Glémin S et al. (2014). *GC content evolution in coding regions of angiosperm genomes: a unifying hypothesis*. *Trends in Genetics* **30**, 263–270. <https://doi.org/10.1016/j.tig.2014.05.002>.
- Gordo I, Charlesworth B (2001). *Genetic linkage and molecular evolution*. *Current Biology* **11**, 684–686. [https://doi.org/10.1016/S0960-9822\(01\)00408-0](https://doi.org/10.1016/S0960-9822(01)00408-0).
- Gorter FA et al. (2018). *Local Fitness Landscapes Predict Yeast Evolutionary Dynamics in Directionally Changing Environments*. *Genetics* **208**, 307–322. <https://doi.org/10.1534/genetics.117.300519>.
- Günther T, Coop G (2013). *Robust identification of local adaptation from allele frequencies*. *Genetics* **195**, 205–220. <https://doi.org/10.1534/genetics.113.152462>.
- Hahn MW (2018). *Molecular population genetics*. Sinauer Associates/Oxford University Press.
- Harpak A et al. (Dec. 2016). *Mutation Rate Variation is a Primary Determinant of the Distribution of Allele Frequencies in Humans*. *PLOS Genetics* **12**, 1–22. <https://doi.org/10.1371/journal.pgen.1006489>.
- Heled J, Drummond AJ (2008). *Bayesian inference of population size history from multiple loci*. *BMC Evolutionary Biology* **8**, 289. <https://doi.org/10.1186/1471-2148-8-289>.
- Heller R et al. (2013). *The confounding effect of population structure on Bayesian skyline plot inferences of demographic history*. *PloS one* **8**. <https://doi.org/10.1371/journal.pone.0062992>.
- Henikoff S et al. (2001). *The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA*. *Science* **293**, 1098–1102. <https://doi.org/10.1126/science.1062939>.
- Hermisson J, Pennings PS (2005). *Soft Sweeps*. *Genetics* **169**, 2335–2352. <https://doi.org/10.1534/genetics.104.036947>.
- Hey J (2010). *Isolation with migration models for more than two populations*. *Molecular biology and evolution* **27**, 905–920. <https://doi.org/10.1093/molbev/msp296>.
- Hodgkinson A, Eyre-Walker A (2011). *Variation in the mutation rate across mammalian genomes*. *Nature Reviews Genetics* **12**, 756. <https://doi.org/10.1038/nrg3098>.
- Hori M (1993). *Frequency-dependent natural selection in the handedness of scale-eating cichlid fish*. *Science* **260**, 216–219. <https://doi.org/10.1126/science.260.5105.216>.
- Hudson RR, Kaplan NL (1995). *Deleterious background selection with recombination*. *Genetics* **141**, 1605–1617. <https://doi.org/10.1093/genetics/141.4.1605>.
- Hurst GD, Jiggins FM (2005). *Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts*. *Proceedings of the Royal Society B: Biological Sciences* **272**, 1525–1534. <https://doi.org/10.1098/rspb.2005.3056>.
- Hurst LD (2019). *A century of bias in genetics and evolution*. *Heredity* **123**, 33–43. <https://doi.org/10.1038/s41437-019-0194-2>.
- Jensen JD et al. (2018). *The importance of the neutral theory in 1968 and 50 years on: a response to Kern & Hahn 2018*. *Evolution* **73**, 111–114. <https://doi.org/10.1111/evo.13650>.

- Keightley PD, Eyre-Walker A (2007). Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* **177**, 2251–2261. <https://doi.org/10.1038/nrg2146>.
- Kern AD, Hahn MW (2018). The neutral theory in light of natural selection. *Molecular biology and evolution* **35**, 1366–1371. <https://doi.org/10.1093/molbev/msy092>.
- Kimura M et al. (1968). Evolutionary rate at the molecular level. *Nature* **217**, 624–626. <https://doi.org/10.1038/217624a0>.
- Klopfstein S et al. (2005). The fate of mutations surfing on the wave of a range expansion. *Molecular biology and evolution* **23**, 482–490. <https://doi.org/10.1093/molbev/msj057>.
- Kreitman M (1996). The neutral theory is dead. Long live the neutral theory. *Bioessays* **18**, 678–683. <https://doi.org/10.1002/bies.950180812>.
- Lachance J, Tishkoff SA (2014). Biased gene conversion skews allele frequencies in human populations, increasing the disease burden of recessive alleles. *American journal of human genetics* **95**, 408–420. <https://doi.org/10.1016/j.ajhg.2014.09.008>.
- Larracuent AM, Presgraves DC (2012). The Selfish Segregation Distorter Gene Complex of *Drosophila melanogaster*. *Genetics* **192**, 33–53. <https://doi.org/10.1534/genetics.112.141390>.
- Lassalle F et al. (2015). GC-Content Evolution in Bacterial Genomes: The Biased Gene Conversion Hypothesis Expands. *PLoS Genet.* **11**, e1004941. <https://doi.org/10.1371/journal.pgen.1004941>.
- Laurin-Lemay S et al. (2018). Multiple Factors Confounding Phylogenetic Detection of Selection on Codon Usage. *Molecular Biology and Evolution* **35**, 1463–1472. <https://doi.org/10.1093/molbev/msy047>.
- Lesecque Y et al. (2013). GC-biased gene conversion in yeast is specifically associated with crossovers: molecular mechanisms and evolutionary significance. *Mol. Biol. Evol.* **30**, 1409–1419. <https://doi.org/10.1093/molbev/mst056>.
- Li H, Durbin R (2011). Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496. <https://doi.org/10.1038/nature10231>.
- Lindholm AK et al. (2016). The Ecology and Evolutionary Dynamics of Meiotic Drive. *Trends in Ecology and Evolution* **31**, 315–326. <https://doi.org/10.1016/j.tree.2016.02.001>.
- Liu X, Fu YX (2015). Exploring population size changes using SNP frequency spectra. *Nature genetics* **47**, 555–559. <https://doi.org/10.1038/ng.3254>.
- Luu K et al. (2017). *pcadapt*: an R package to perform genome scans for selection based on principal component analysis. *Molecular ecology resources* **17**, 67–77. <https://doi.org/10.1111/1755-0998.12592>.
- Machado HE et al. (2016). Comparative population genomics of latitudinal variation in *Drosophila simulans* and *Drosophila melanogaster*. *Molecular Ecology* **25**, 723–740. <https://doi.org/10.1111/mec.13446>.
- Martin AR et al. (2017). Human demographic history impacts genetic risk prediction across diverse populations. *The American Journal of Human Genetics* **100**, 635–649. <https://doi.org/10.1016/j.ajhg.2017.03.004>.
- Martin G et al. (2007). Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nature genetics* **39**, 555–560. <https://doi.org/10.1038/ng1998>.
- McClintock B (1950). The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences* **36**, 344–355. <https://doi.org/10.1073/pnas.36.6.344>.
- Mekel-Bobrov N et al. (2005). Ongoing Adaptive Evolution of ASPM, a Brain Size Determinant in *Homo sapiens*. *Science* **309**, 1720–1722. <https://doi.org/10.1126/science.1122822>.
- Messer P, Neher RA (2012). Estimating the strength of selective sweeps from deep population diversity data. *Genetics* **191**, 595–605. <https://doi.org/10.1534/genetics.112.138461>.
- Muller HJ (1932). Some genetic aspects of sex. *The American Naturalist* **66**, 118–138. <https://doi.org/10.1086/280418>.
- Myers S et al. (2010). Drive Against Hotspot Motifs in Primates Implicates the PRDM9 Gene in Meiotic Recombination. *Science* **327**, 876–879. <https://doi.org/10.1126/science.1182363>.



- Ohta T, Kimura M (1970). Development of associative overdominance through linkage disequilibrium in finite populations. *Genetical Research* **16**, 165–177. <https://doi.org/10.1017/S0016672300002391>.
- Ohta T, Kimura M (1971). On the constancy of the evolutionary rate of cistrons. *Journal of Molecular Evolution* **1**, 18–25. <https://doi.org/10.1007/BF01659391>.
- Paigen K, Petkov PM (2018). PRDM9 and Its Role in Genetic Recombination. *Trends in Genetics* **34**, 291–300. <https://doi.org/10.1016/j.tig.2017.12.017>.
- Peery MZ et al. (2012). Reliability of genetic bottleneck tests for detecting recent population declines. *Molecular ecology* **21**, 3403–3418. <https://doi.org/10.1111/j.1365-294X.2012.05635.x>.
- Peischl S et al. (2013). On the accumulation of deleterious mutations during range expansions. *Molecular ecology* **22**, 5972–5982. <https://doi.org/10.1111/mec.12524>.
- Pennings PS, Hermisson J (2006a). Soft Sweeps III: The Signature of Positive Selection from Recurrent Mutation. *PLOS Genetics* **2**, 1–15. <https://doi.org/10.1371/journal.pgen.0020186>.
- Pennings PS, Hermisson J (2006b). Soft Sweeps II: Molecular Population Genetics of Adaptation from Recurrent Mutation or Migration. *Molecular Biology and Evolution* **23**, 1076–1084. <https://doi.org/10.1093/molbev/msj117>.
- Pollard K et al. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167–172. <https://doi.org/10.1038/nature05113>.
- Pouyet F, Aeschbacher S, et al. (2018). Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. *Elife* **7**, e36317. <https://doi.org/10.7554/eLife.36317>.
- Pouyet F, Mouchiroud D, et al. (2017). Recombination, meiotic expression and human codon usage. *Elife* **6**, e27344. <https://doi.org/10.7554/eLife.27344>.
- Presgraves DC et al. (May 2009). Large-scale selective sweep among segregation distorter chromosomes in African populations of *Drosophila melanogaster*. *PLOS Genetics* **5**, 1–13. <https://doi.org/10.1371/journal.pgen.1000463>.
- Ratnakumar A et al. (2010). Detecting positive selection within genomes: the problem of biased gene conversion. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **365**, 2571–2580. <https://doi.org/10.1098/rstb.2010.0007>.
- Roze D (2015). Effects of Interference Between Selected Loci on the Mutation Load, Inbreeding Depression, and Heterosis. *Genetics* **201**, 745–757. <https://doi.org/10.1534/genetics.115.178533>.
- Sabeti PC, Reich DE, et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837. <https://doi.org/10.1038/nature01140>.
- Sabeti PC, Varilly P, et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918. <https://doi.org/10.1038/nature06250>.
- Schrider DR, Kern AD (2016). S/HIC: Robust Identification of Soft and Hard Sweeps Using Machine Learning. *PLOS Genetics* **12**, e1005928. <https://doi.org/10.1371/journal.pgen.1005928>.
- Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595. <https://doi.org/10.1093/genetics/123.3.585>.
- Tanaka TU et al. (2013). Three wise centromere functions: see no error, hear no break, speak no delay. *EMBO reports* **14**, 1073–1083. <https://doi.org/10.1038/embor.2013.181>.
- Tataru P, Bataillon T (2019). polyDFEv2. 0: testing for invariance of the distribution of fitness effects within and across species. *Bioinformatics* **35**, 2868–2869. <https://doi.org/10.1093/bioinformatics/bty1060>.
- Taylor SM et al. (2012). Haemoglobinopathies and the clinical epidemiology of malaria: a systematic review and meta-analysis. *The Lancet Infectious Diseases* **12**, 457–468. [https://doi.org/10.1016/S1473-3099\(12\)70055-5](https://doi.org/10.1016/S1473-3099(12)70055-5).
- Terhorst J et al. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature genetics* **49**, 303. <https://doi.org/10.1038/ng.3748>.
- Thomas CD (2010). Climate, climate change and range boundaries. *Diversity and Distributions* **16**, 488–495. <https://doi.org/10.1111/j.1472-4642.2010.00642.x>.

- Tian C et al. (2008). Accounting for ancestry: population substructure and genome-wide association studies. *Human molecular genetics* **17**, 143–150. <https://doi.org/10.1093/hmg/ddn268>.
- Turelli M, Hoffmann AA (1991). Rapid spread of an inherited incompatibility factor in California *Drosophila*. *Nature* **353**, 440–442. <https://doi.org/10.1038/353440a0>.
- Villanueva-Canas JL et al. (2017). Beyond SNPs: how to detect selection on transposable element insertions. *Methods in Ecology and Evolution* **8**, 728–737. <https://doi.org/10.1111/2041-210X.12781>.
- Voight BF et al. (2006). A map of recent positive selection in the human genome. *PLoS biology* **4**, e72. <https://doi.org/10.1371/journal.pbio.0040154>.
- Webster MT et al. (2006). Strong Regional Biases in Nucleotide Substitution in the Chicken Genome. *Molecular Biology and Evolution* **23**, 1203–1216. <https://doi.org/10.1093/molbev/msk008>.
- Whitlock MC, Lotterhos KE (2015). Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of  $F_{ST}$ . *The American Naturalist* **186**, S24–S36. <https://doi.org/10.1086/682949>.
- Wicker T et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8**, 973–982. <https://doi.org/10.1038/nrg2165>.
- Wright S (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159. <https://doi.org/10.1093/genetics/16.2.97>.
- Xiong W et al. (2014). HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proceedings of the National Academy of Sciences* **111**, 10263–10268. <https://doi.org/10.1073/pnas.1410068111>.
- Zhao L, Charlesworth B (2016). Resolving the conflict between associative overdominance and background selection. *Genetics* **203**, 1315–1334. <https://doi.org/10.1534/genetics.116.188912>.
- Zhu YO et al. (2014). Precise estimates of mutation rate and spectrum in yeast. *Proceedings of the National Academy of Sciences* **111**, E2310–E2318. <https://doi.org/10.1073/pnas.1323011111>.