Disentangling Types of Linked Selection Using Patterns of Nucleotide Variation in *Drosophila pseudoobscura*

Katharine L. Korunes, Kieran Samuk, and Mohamed A. F. Noor

Abstract A central goal in evolutionary genetics is to understand the forces that drive patterns of nucleotide variation within species. Recombination rate and nucleotide diversity are positively correlated across loci within many species. Selective sweeps and background selection both may contribute to this pattern by reducing nucleotide variation disproportionately in regions of low recombination. While there is unambiguous documentation of some forms of selective sweeps, it is less clear how much background selection and soft selective sweeps contribute. Here, we discuss the study of these different types of linked selection, paying particular attention to the large body of work devoted to linked selection in *Drosophila*. We discuss patterns of nucleotide variation in the classic model system *Drosophila pseudoobscura*, and we leverage this system to test for the effects of linked selection in the near absence of hard, complete selective sweeps. Our case study of variation in sets of loci shared between *Drosophila pseudoobscura* and *D. miranda* suggests that selection at linked sites may reduce nucleotide variation even in the absence of hard selective sweeps. We discuss how this work relates to recent approaches and challenges in the study of how selection at linked sites influences patterns of nucleotide variation.

Keywords Background selection · Linked selection · Recombination · Selective sweeps

1 Introduction

One of the most striking and widespread patterns in genomic data is the positive correlation between local recombination rates and nucleotide diversity (e.g., Nachman 2001; Tenaillon et al. 2001; Takahashi et al. 2004; Begun et al. 2007;

K. L. Korunes (✉) · K. Samuk · M. A. F. Noor
Biology Department, Duke University, Durham, NC, USA
e-mail: katharine.korunes@duke.edu

Julian R. Dupuis and Om P. Rajora (eds.), *Population Genomics: Insects*,
Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2021_87,

Cutter and Choi 2010; Burri et al. 2015). Since this pattern was observed in *Drosophila* 30 years ago (Aguade et al. 1989; Begun and Aquadro 1992; W. Stephan and Langley 1989), many studies have attempted to elucidate the evolutionary forces that drive this association. The parsimonious explanation of a neutral process, such as mutagenic effects of recombination, driving this association was undermined by the observation that sequence divergence between *Drosophila melanogaster* and its sister species, *D. simulans,* is not associated with recombination rates (Begun and Aquadro 1992; but see Halldorsson et al. 2019 regarding the mutagenic effects of recombination in humans). Instead, evidence supports the hypothesis that much of this pattern can be explained by natural selection eliminating variation in regions of low recombination (Begun and Aquadro 1992; Begun et al. 2007; McGaugh et al. 2012). Strong linkage in regions of low recombination allows positive and/or negative selection at a given locus to reduce nucleotide diversity at surrounding, linked loci. On a per-base pair level, this results in a greater reduction in diversity due to linked selection in regions of low vs. high recombination.

Hard selective sweeps result from positive selection when a single haplotype containing an advantageous mutation replaces multiple ancestral haplotypes, and alleles linked to the advantageous mutation become fixed along with the mutation as it spreads (Fig. 1a; Maynard Smith and Haigh 1974; Kaplan et al. 1989). However, selection at linked sites may also reduce nucleotide variation in the absence of hard selective sweeps. Linked selection can also occur via partial sweeps, soft sweeps, and background selection. Partial sweeps are selective sweeps that are incomplete, where a mutation has increased in frequency but has not reached fixation. Soft sweeps occur when multiple copies of an allele (on different backgrounds) contribute to a selective sweep, though these copies may or may not be identical by descent (Hermisson and Pennings 2005). Hermisson and Pennings (2005) introduced a model in which standing genetic variation is the basis of a soft sweep, such as when an allele segregating in the population becomes advantageous after a shift in selective pressures, such as a change in the environment (Fig. 1b). A soft sweep can also occur when an adaptive allele repeatedly enters the population (Pennings and Hermisson 2006). For example, migration or mutation might repeatedly introduce an advantageous allele into a population (Fig. 1c). Like hard sweeps, partial and soft sweeps may reduce diversity at nearby neutral sites as a result of a beneficial mutation increasing in frequency due to positive selection.

Negative selection can also reduce linked neutral variation through the action of background selection. Under background selection, multiple deleterious mutations arise over time, and selection eliminates the deleterious mutations along with closely linked neutral variation (Fig. 1d) (Charlesworth et al. 1993). Neutral mutations can only drift to high frequencies in a population if they exist on haplotypes that remain free of (or have the fewest/weakest) deleterious mutations.

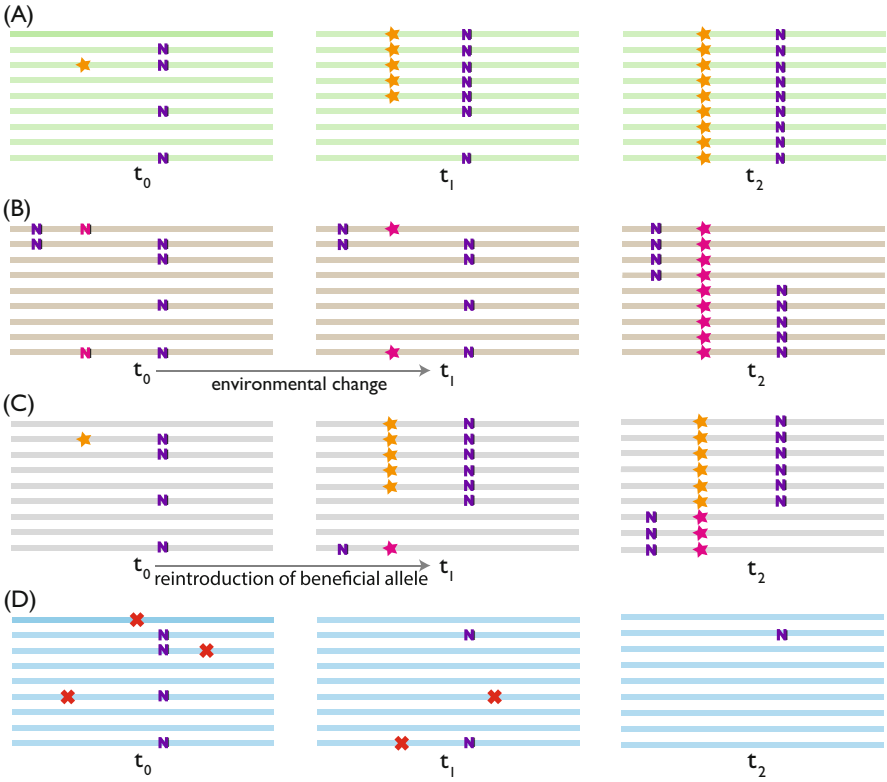

Fig. 1 Models of how selection influences linked variation. The illustrated population has a neutral polymorphism segregating at time t_0 at a site indicated by "N." Under the hard sweeps model (**a**), a novel advantageous mutation (gold star) arises at t_0, increases in frequency to yield a partial sweep at t_1, and has swept to fixation by t_2. (**b**, **c**) represent soft sweeps. In (**b**), a previously neutral mutation becomes advantageous (magenta star) due to an environmental change between t_0 and t_1, so the beneficial allele copies contributing to the sweep are identical by descent but segregating on different backgrounds due to recombination. In (**c**), the contributing allele enters the population recurrently via mutation or migration. Under background selection (**d**), haplotypes with deleterious mutations (red) are eliminated, which reduces linked neutral variation

1.1 Inferring the Relative Contributions of Types of Linked Selection

There is a long-standing debate over the relative contributions of different types of linked selection to shaping patterns of genetic diversity (e.g., Stephan 2010; Jensen 2014; Renzette et al. 2016). Because recombination reduces the strength of linkage, different modes of linked selection all reduce nucleotide variation disproportionately in regions of low recombination. However, the independent effects of different types of linked selection are difficult to disentangle. Many studies document regions of

low nucleotide diversity centered on adaptive alleles that were previously rare or absent in the population (e.g., de Groot et al. 2002; Bersaglieri et al. 2004; Colosimo et al. 2005; Hoekstra et al. 2006; see Stephan 2016 for review). While such studies provide unambiguous evidence of hard selective sweeps, it is less clear how much background selection and soft selective sweeps contribute to patterns of diversity in the genome. Few studies seek empirical signatures of linked selection in the absence of hard sweeps, and these signatures are often more controversial. For example, some studies argue that soft sweeps might be a very prevalent mode of selection (Messer and Petrov 2013; Schrider and Kern 2017), while others argue that this is the result of demographic and neutral processes mimicking the signatures of soft sweeps (Harris et al. 2018; Jensen 2014).

Despite the evidence for many individual selective sweeps, the genome-wide prevalence of sweeps is still ambiguous. There is evidence for widespread selective sweeps in well-studied species such as *Drosophila simulans,* where most neutral sites across the genome appear to have been affected by positive selection at linked sites, specifically by repeated hard sweeps over evolutionary time ("recurrent selective sweeps") (Sella et al. 2009). However, such cases of hard sweeps may not be representative of typical adaptation in all organisms, and other processes may be more prevalent as drivers of reduced neutral variation in regions of low recombination. Analyses of patterns of diversity surrounding conserved regions in 179 human genomes suggest that hard sweeps are not the primary determinant of genetic diversity in modern human populations (Alves et al. 2012; Hernandez et al. 2011). Notably, an apparent lack of hard sweeps does not necessarily imply an absence of positive selection, as adaptation may frequently involve polygenic adaptation or soft sweeps (e.g., Pritchard et al. 2010). Similarly, recurrent selective sweeps appear to be rare in plants, and background selection may dominate as a driver of patterns of genetic diversity in many plant populations (reviewed in Slotte 2014).

1.2 The Dynamics of Background Selection

Other recent work focuses on the role of the strength of selection and underscores the importance of accounting for weakly deleterious mutations. In particular, background selection caused by weakly deleterious alleles has been highlighted as a key factor shaping patterns of genetic variation (Lohmueller et al. 2011). Though the effects of background selection are in part dictated by the strength of selection, this occurs very differently than it does under selective sweeps. Unlike selective sweeps, which are most pronounced in the case of strong selection and dominant mutations, the strongest reduction of linked diversity due to background selection occurs when selection is relatively weak and mutations are recessive (Charlesworth 2012a; Charlesworth et al. 1993; Cutter and Payseur 2013). Strongly deleterious alleles are kept at low frequencies in the population, while weakly deleterious alleles often persist longer in the population. This implies that if there is strong linkage between neutral sites and sites under selection, then weakly deleterious alleles can be

recombined onto a greater number of haplotypes compared to strongly deleterious alleles before being eliminated (Nordborg et al. 1996). This observation means that we must use caution when inferring background selection based on constraint per se (i.e. as a proxy for the strength of negative selection). It is tempting to assume that genes that are more conserved among taxa, or under greater constraint, would experience greater background selection. However, as discussed, this is not necessarily true if highly constrained genes experience stronger selection.

In addition to the strength of selection, the genomic context of deleterious alleles, including other nearby sites under selection and the strength of linkage, affects background selection. Recent work to help distinguish types of linked selection includes simulations modeled after both humans and *Drosophila* showing that the spatial arrangement of constrained sites can dictate whether the valley of diversity caused by background selection mirrors signatures caused by positive selection (Schrider 2020). Similarly, any factors that increase linkage affect background selection. A striking example is the apparent role of background selection in shaping patterns of variation of the *Drosophila* X chromosome, as illustrated by a recent investigation of silent site diversity in *D. melanogaster* (Charlesworth 2012b), where mean silent site diversity in the ancestral population appears to be approximately the same for X-linked loci and autosomal loci (Andolfatto 2001; Hutter et al. 2007). The similarity of neutral diversity for X-linked and autosomal loci deviates from the null expectation based on the ratio of the effective population sizes (N_e) of the X chromosome and each autosome. This ratio of effective population sizes is expected to be ¾, given an equal sex ratio and a random distribution of the number of progeny of each sex (Wright 1931). According to the results from Charlesworth (2012a), background selection can explain the equality of N_e between the X chromosomes and the autosomes, if background selection is more effective on the autosomes. This explanation fits with the fact that there is no meiotic recombination in male *Drosophila,* so the X chromosome (of which 2/3 are present in females) has a greater effective recombination rate compared to the autosomes (of which half are in females). The observed X:autosome silent site diversity ratio was closer to the expectation in *D. pseudoobscura,* and recent data reveals a lower than expected ratio in *D. simulans* (Jackson et al. 2017). The evolutionary mechanisms underlying X:autosome differences remain poorly understood, but it is possible that the ratios in *D. pseudoobscura* and *D. simulans* are due to higher recombination rates in these species (Charlesworth 2012b; Jackson et al. 2017).

These observations of X:autosome neutral diversity illustrate the intricacies of how background selection operates under varying degrees of linkage. Relatedly, our ability to infer the effects of background selection is often complicated by demographic history (Torres et al. 2020), which often involves changes in N_e. The interplay between N_e and background selection can also be seen in asexual and selfing species, where linkage is often strong. Evidence suggests that background selection plays a prominent role in decreasing genetic diversity under uniparental inheritance (Agrawal and Hartfield 2016; Roze 2016). Similarly, background selection has been implicated as a driver of low observed variability on non-recombining chromosomes such as the dot chromosome of *D. melanogaster,* the neo-Y of

D. miranda (Kaiser and Charlesworth 2009), and recently, the Y chromosome in humans (Wilson Sayres et al. 2014). On non-recombining chromosomes or in low recombination genomic regions, interference among selected loci plays a major role in the action of background selection. Hill–Robertson interference (Felsenstein 1974) influences the interaction between recombination and selection, and interference among the sites that cause background selection can result in a larger role of drift in shaping allele frequencies (reviewed in Charlesworth and Campos 2014).

1.3 Recent Progress in Genome-Wide Scans for Selection

Genome-wide scans for positive selection often rely on the hard sweeps model, with neutrality as a null hypothesis (Vitti et al. 2013). However, as seen in a recent implementation of the SweepFinder2 framework, some emerging methods for genome-wide scans for sweeps account for background selection (DeGiorgio et al. 2016; Huber et al. 2016). The controversy over the role of selective sweeps necessitates a baseline prediction of genetic diversity patterns under background selection alone. The standard methodology for distinguishing sweeps and background selection typically relies on detecting shifts in site frequency spectra, analyzed through statistics such as Tajima's D (Tajima 1989), Fu and Li's D (Fu and Li 1993), or Fay and Wu's H (Fay and Wu 2000). This approach often yields inconclusive results, partially because these statistics typically assume a standard neutral model and may neglect the effects of demography or population structure.

In recent years, the rapid accumulation of genomic data has inspired new possible approaches to this problem. McVicker et al. (2009) inferred a map of background selection effects in hominid genome evolution using polymorphism data from five primate species, including humans. McVicker et al. (2009) estimated the effects of purifying selection on neutral diversity as a function of selection coefficients, recombination rates, and deleterious mutation rates. At the megabase scale, their map of background selection effects explains a large proportion of observed diversity patterns across the genome, though the associated estimate of the deleterious mutation rate from the data is substantially higher than other estimates from the literature (Kondrashov 2003; Nachman and Crowell 2000).

Similarly, predictions from background selection models alone may explain a large proportion of observed variation in *Drosophila melanogaster*. According to one estimate, background selection may explain as much as 70% of the observed variance in nucleotide diversity across individual *D. melanogaster* (Comeron 2014). Evidence continues to indicate that genome-wide patterns of variation are broadly consistent with expectations of background selection in many species (Charlesworth 2012b; Charlesworth and Campos 2014; Comeron 2014; Hernandez et al. 2011; Lohmueller et al. 2011; McVicker et al. 2009; Pouyet et al. 2018; Renzette et al. 2016). Recent work even suggests that background selection may affect neutral divergence between distantly related species (Phung et al. 2016).

The recent attempts to infer genome-wide effects of background selection bolster the argument that background selection *could* be a major predictor of genome-wide diversity. While studies such as McVicker et al. (2009) and Comeron (2014) underscore that a background selection model can explain a large proportion of observed diversity patterns across polymorphism data (in primates and *D. melanogaster,* respectively), a common limitation of this type of analysis is the lack of a relative assessment of the explanatory power of a sweeps-only model or a joint model (background selection and sweeps together). Recent work aims to simultaneously estimate the effects of both positive and negative selection in genomic data (Elyashiv et al. 2016). These emerging modeling approaches have advanced our understanding of how different selective forces contribute to lower diversity in regions of low recombination. However, there is essentially no direct empirical evidence of the operation of background selection, and no standard methodology for detecting it. Notably, Elyashiv et al. (2016) and McVicker et al. (2009) both estimated a higher deleterious mutation rate than direct empirical studies have reported. Discrepancies such as these underscore the need to continue improving models of linked selection and to inform them with empirical observations.

2 A Case Study from *D. pseudoobscura*: A Test for the Effects of Linked Selection in the Near Absence of Recent Hard Selective Sweeps

Differentiating the effects of different types of linked selection is critical to understanding the extent to which linked selection influences genome-wide patterns of variation. Further, identifying the evolutionary forces acting locally on specific regions of the genome is central to delineating which portions may be evolving neutrally or close to neutrally. There is growing interest in identifying such neutrally evolving sequences to remove biases in demographic inference (Pouyet et al. 2018; Schrider et al. 2016). To leverage empirical evidence towards differentiating the effects of types of linked selection, we present a new potential approach for assessing the effects of selection at linked sites not attributable to recent hard, complete sweeps on within-species patterns of genetic diversity. We test this method using the model system *Drosophila pseudoobscura* and its close relative, *D. miranda.* Previous work shows that *D. pseudoobscura* exhibits a positive association between recombination rate and nucleotide diversity (Kulathinal et al. 2008; McGaugh et al. 2012). Background selection, selective sweeps, or a combination of these forces may explain this association. We attempt to distinguish these forces by comparing the *D. pseudoobscura* genome to closely related genomes. *D. miranda* serves as an appropriate point of comparison, because it is a close relative but (unlike *D. persimilis*) does not hybridize with *D. pseudoobscura*. An additional species, *D. lowei*, provides an outgroup for polarizing differences as ancestral or derived when examining substitutions between *D. pseudoobscura* and *D. miranda*.

We examine the effects of linked selection in the absence of hard sweeps by looking at variation in genes lacking fixed differences between *Drosophila pseudoobscura* and *D. miranda*. This approach greatly reduces the probability that a hard sweep has occurred recently in the subset of genes examined. Here, we present our assessments of nucleotide variation in this set of loci, and we use the findings and caveats (elaborated in *Case Study: Results* and *Discussion*) of this case study to stimulate discussion on potential future directions in differentiating the effects of different types of linked selection. This examination of variation within *Drosophila pseudoobscura* suggests that removing the effect of recent, hard sweeps does not dramatically alter the strength of the relationship between neutral variation and recombination rates, supporting the idea that selection at linked sites may reduce nucleotide variation even in the absence of hard selective sweeps.

2.1 Case Study: Methods

2.1.1 Ortholog Identification

We examined variation between *D. pseudoobscura* and *D. miranda* using whole-genome sequence data from 29 *D. pseudoobscura* and 11 *D. miranda* strains (McGaugh et al. 2012; Samuk et al. 2020; data accessions provided in Table 1). *D. miranda* is not independently annotated, but it diverged from *D. pseudoobscura* within the past ~2 million years (Wang and Hey 1996), facilitating alignment of the *D. miranda* genomes to the *D. pseudoobscura* genome assembly. One *D. lowei* genome was included for use as an outgroup. We aligned the sequences from these 41 genomes to the most recent *D. pseudoobscura* reference assembly (Dpse_3.04: GCA_000001765.2) using BWA-0.7.5a (Li and Durbin 2009). We used GATK v3.8 to call SNPs (McKenna et al. 2010; Van der Auwera et al. 2013). Prior to downstream analyses, we filtered based on GATK's hard filtering recommendations to exclude sites with QualByDepth (QD) < 2.0, FisherStrand Bias (FS) > 60, and StrandOddsRatio (SOR) > 3.0, MQ < 40, MQRankSum < −12.5, ReadPosRankSum < −8. For each of *D. pseudoobscura's* 16,959 annotated genes published by Flybase (http://flybase.org, Full Annotation Release 3.04), we used the annotated gene span to identify the variant data at that locus for each of the 41 genomes. Scripts used to analyze the data are available at https://github.com/kkorunes/Dpseudoobscura_LinkedSelection).

2.1.2 Recombination Map

Recombination rates were obtained from two independent studies of crossover rates within *D. pseudoobscura* (McGaugh et al. 2012; Samuk et al. 2020). LD-based recombination maps are also available for these species (Smukowski Heil et al.

Table 1 Genomic data

Strain	NCBI accession
Drosophila lowei	
Lab3Lowei	SRX091467: SRR330416, SRR330418
Drosophila miranda	
MA28	SRX950183: SRR1873751
MAO101-4	SRX950187: SRR1873752
MAO3-3	SRX950188: SRR1873753
MAO3-4	SRX950189: SRR1873754
MAO3-5	SRX950190: SRR1873755
MAO3-6	SRX950211: SRR2042916
ML14	SRX965452: SRR1925723
ML16	SRX965455: SRR1925728
ML6f	SRX965460: SRR1925734
SP138	SRX965461: SRR1925735
SP235	SRX965462: SRR1925736
Drosophila pseudoobscura	
AFC12	SRX091462: SRR330321, SRR330322, SRR330323, SRR330324
FS18	SRX091310: SRR330103, SRR330104, SRR330105, SRR330106
MAT32	SRX091461: SRR330317, SRR330318, SRR330319, SRR330320
MATTL	SRX091324: SRR330129, SRR330130, SRR330131, SRR330197
MSH9	SRX091465: SRR330329, SRR330330, SRR330331, SRR330333
MSH24	SRX091463: SRR330325, SRR330326, SRR330327, SRR330328
PP1134	SRX091323: SRR330125, SRR330126, SRR330127, SRR330128
PP1137	SRX091311: SRR330107
A24	SRX7842600: SRR11230564
M15	SRX7842599: SRR11230565
M20	SRX7842598: SRR11230566
A30	SRX7842597: SRR11230567
A57	SRX7842596: SRR11230568
A49	SRX7842595: SRR11230569
A48	SRX7842594: SRR11230570
A12	SRX7842593: SRR11230571
M27	SRX7842591: SRR11230573
Flag14	SRX7842590: SRR11230574
A19	SRX7842589: SRR11230575
M17	SRX7842588: SRR11230576
A60	SRX7842587: SRR11230577
M14	SRX7842586: SRR11230578
MV2-25	SRX7842585: SRR11230579
A6	SRX7842584: SRR11230580
M6	SRX7842583: SRR11230581
M13	SRX7842582: SRR11230582
A56	SRX7842581: SRR11230583
A14	SRX7842580: SRR11230584
A47	SRX7842579: SRR11230585

2015) and are highly correlated with the empirical recombination maps generated by McGaugh et al. (2012). Our use of the empirical recombination maps avoids issues of circularity from measuring sequence diversity using LD-based recombination maps. The recombination dataset from McGaugh et al. 2012 consists of two fine-scale recombination maps for two separate pairs of inbred lines of *D. pseudoobscura* (McGaugh et al. 2012). We averaged these two nearly identical replicate maps from McGaugh et al. (2012), which together cover much of chromosome 2 and the right arm of the X chromosome (chromosome XR). The recombination dataset generated by Samuk et al. (2020) covers much of chromosome 2, chromosome 4, and both arms of the X chromosome (XL and XR). For each chromosome, we used the recombination map with the greatest marker density, which led to the selection of chromosome 2 from McGaugh et al. (2012) and chromosomes 4, XL, and XR from Samuk et al. (2020). Together, the recombination datasets used amount to approximately 65% (~99 mb) of the ~153 mb *D. pseudoobscura* genome assembly. For each of the 231 genomic intervals denoted by flanking markers (averaging 431 kb apart), the maps provide a recombination rate reported in Kosambi centiMorgans per megabase (cM/mb). We used the coordinates of the markers to determine which interval contains each of the genes under examination.

2.1.3 Analysis of Fixed Differences and Neutral Variation

For each gene, we compared the predicted gene span (including untranslated regions (UTRs), coding sequences, and introns) across all 41 individuals to identify fixed differences between *D. pseudoobscura* and *D. miranda*. To reduce possible effects of selective sweeps in or near the gene, we included the introns and UTRs since evidence suggests that selection regularly acts upon these regions (e.g., Kohn et al. 2004; Andolfatto 2005). To assess the effect of extending the screen to include nearby regions beyond the UTRs, we repeated the analyses including 50 bp upstream and 50 bp downstream of each gene span in our screen for fixed differences. Evidence suggests that LD decays rapidly in *Drosophila* (Langley et al. 2000), and screening the gene span itself would exclude many sweeps of advantageous mutations near the coding region. However, some (e.g., strongly selected) sweeps that occurred outside of the UTR boundaries could still affect nucleotide diversity within the gene. We also tested larger region sizes to assess whether extending our filter further upstream and downstream would affect the results. We found that extending the upstream/downstream regions does not change the directions of the association between recombination rate and nucleotide diversity, but the power to detect a relationship is reduced (Table 2). Thus, we primarily focus on results from screening the genes themselves.

To increase the number of genes we can examine while maintaining the criterion of "absence of recent selective sweeps," we compared the loci to the outgroup species, *D. lowei,* and we used a simple parsimony argument to determine which species likely possesses the ancestral allele for each fixed difference. The lineage with the ancestral allele at a given locus was considered to be the lineage less likely

Table 2 The effect of upstream/downstream region size on measures of neutral diversity in association with recombination rate

Loci[a]	Correlation (Pearson's R) between fourfold degenerate site π and local recombination rate
All genes	$n = 6{,}640$, $R = 0.19$, $p < 2.2 \times 10^{-16}$
Autosomal	$n = 3{,}723$, $R = 0.21$, $p < 2.2 \times 10^{-16}$
X chromosome	$n = 2{,}917$, $R = 0.08$, $p = 4.32 \times 10^{-6}$
No sweeps (within the gene)	$n = 151$, $R = 0.25$, $p = 0.0017$
Autosomal	$n = 107$, $R = 0.15$, $p = 0.1114$
X chromosome	$n = 44$, $R = 0.41$, $p = 0.0059$
No sweeps (gene + 50 bp buffer)	$n = 121$, $R = 0.27$, $p = 0.0028$
Autosomal	$n = 85$, $R = 0.17$, $p = 0.1127$
X chromosome	$n = 36$, $R = 0.42$, $p = 0.0099$
No sweeps (gene + 100 bp buffer)	$n = 77$, $R = 0.21$, $p = 0.0602$
Autosomal	$n = 49$, $R = 0.19$, $p = 0.2029$
X chromosome	$n = 28$, $R = 0.24$, $p = 0.2090$

[a]Limited to genes within the recombination map. "No sweeps" indicates that these loci had no derived fixed differences in *D. pseudoobscura*

to have been affected by recent sweeps, whereas the lineage with the derived allele may have experienced a sweep at that locus. For example, if a gene has several differences between *D. pseudoobscura* and *D. miranda,* but all differences are derived in *D. miranda*, then that *D. pseudoobscura* locus is presumably free of the effects of recent hard sweeps. Following this argument, only when *all* fixed differences were derived in *D. miranda* did we analyze neutral variation at that locus within *D. pseudoobscura.* Using this strategy to identify hard sweeps (Fig. 2), we have obtained a set of *D. pseudoobscura* loci where we can examine the effects of linked selection in the relative absence of hard selective sweeps.

To assess neutral variation, we computed fourfold degenerate site π (custom scripts provided in accompanying Github repository linked above), since synonymous sites represent a category of sites that evolve under relatively minimal constraint (Andolfatto 2005). For genes that have multiple predicted isoforms, we analyzed the longest isoform entry for each gene. R version 3.6.1 was used to perform statistical tests, including linear regressions and analysis of covariance (ANCOVA) to assess differences in the slope and y-intercepts of the regression lines of local recombination rate and neutral diversity between the sets of loci (R Core Team, 2019, "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/).

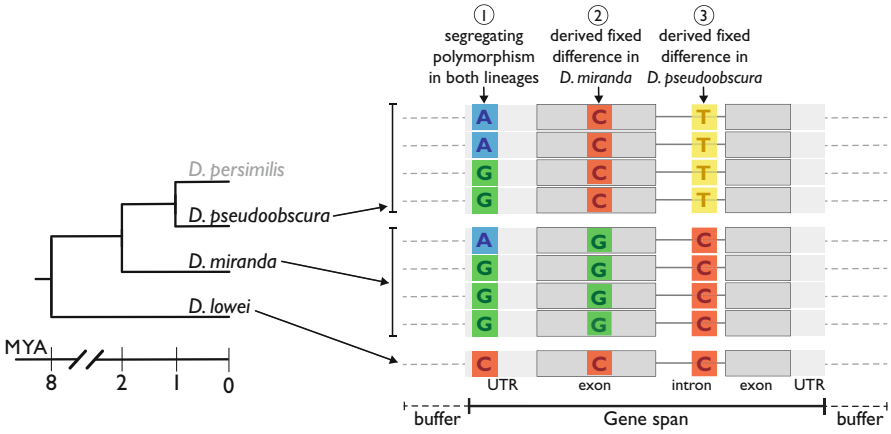

Fig. 2 Schematic of the logic underlying sweep exclusion. A cladogram (left) shows the phylogenetic context of the analyzed species. For the examined orthologs shared between *D. pseudoobscura* and *D. miranda*, we compared each locus across the 41 genomes described above. As an example (right), we depict 4 *D. pseudoobscura,* 4 *D. miranda*, and the outgroup *D. lowei* sequence to illustrate our strategy for excluding hard sweeps in predicted gene spans. This example gene has a within-species polymorphism (site 1) and two fixed differences between *D. pseudoobscura* and *D. miranda* (sites 2 and 3). By comparing to the outgroup *D. lowei,* we find that *D. pseudoobscura's* fixed "C" at site 2 matches the likely ancestral state. In contrast, site 3 has a fixed, derived "T" in *D. pseudoobscura.* This fixed difference at site 3 represents a possible recent sweep in *D. pseudoobscura* and would disqualify this whole gene from the set of *D. pseudoobscura* loci unlikely to have been affected by recent sweeps.

2.2 *Case Study: Results*

Of the examined orthologs shared between *D. pseudoobscura* and *D. miranda,* 6,640 loci are contained within the recombination map. These 6,640 loci exhibit a positive association between recombination rate and neutral diversity (fourfold degenerate site π) within *D. pseudoobscura* (Table 2; Fig. 3a), in agreement with previous studies within *D. pseudoobscura* (Kulathinal et al. 2008; McGaugh et al. 2012). This significant positive relationship exists for both autosomal and X chromosome loci (Fig. 3b, c).

Next, we examined how screening out fixed differences affects the positive relationship between recombination rate and neutral diversity. The effects of hard selective sweeps on within-species nucleotide variation can be minimized by screening *D. pseudoobscura* loci for derived fixed differences in their coding regions, introns, and untranslated regions (Fig. 2). We identified 151 *D. pseudoobscura* loci with no derived fixed differences within the gene's coding sequence, introns, or untranslated regions. This set of loci is unlikely to have experienced recent hard sweeps within the gene span, since we have reduced the possibility that a novel mutation arose and spread to fixation within *D. pseudoobscura* after the divergence of *D. pseudoobscura* and *D. miranda* from their common ancestor. These 151 loci

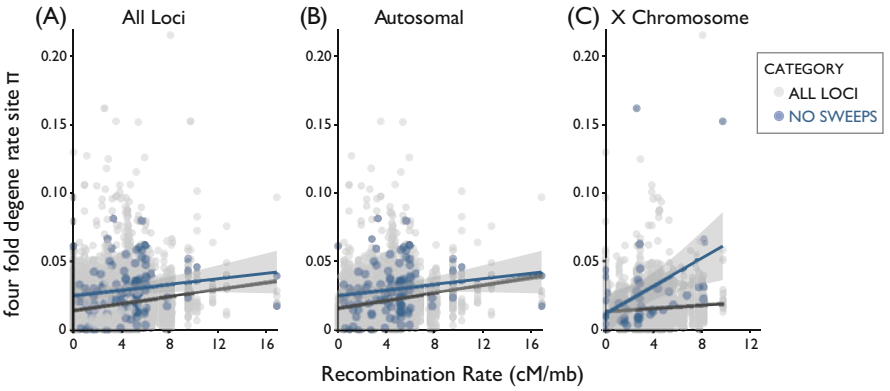

Fig. 3 Neutral diversity (fourfold degenerate site π) regressed onto local recombination rate. This relationship is shown for all loci (gray) and for the subset of loci with no fixed differences (*D. pseudoobscura* vs. *D. miranda*) derived in *D. pseudoobscura* (dark blue), shown for (**a**) the full set of loci, (**b**) autosomal loci, and (**c**) loci on the X chromosome

exhibit a positive association between recombination rate and neutral diversity (fourfold degenerate site π) within *D. pseudoobscura* (Table 2; Pearson's $R = 0.25$, $p = 0.0017$). The set of 151 genes is a relatively small proportion of the total number of loci, but this proportion should not be interpreted as a percentage of loci not affected by sweeps. This approach does not yield a binary between loci affected by sweeps vs. not affected by sweeps. Instead, the set of 151 loci represents a subset with reduced possible effects of selective sweeps in or near the gene. Some loci may have acquired fixed differences by processes other than hard sweeps, and screening for fixed differences may not capture all hard sweeps, as discussed further below.

We next discuss how the positive association between recombination rate and neutral diversity observed in the set of genes without derived fixed differences might differ from that observed in the full set of loci. Because hard sweeps reduce neutral variation, a selective sweeps model predicts less neutral variation compared to predicted nucleotide variation in the absence of hard sweeps. Removing effects of hard sweeps may also result in a weaker correlation between neutral variation and recombination. If background selection and/or soft sweeps have detectable effects on neutral variation, we expect that the correlation between neutral variation and recombination rate in the absence of hard sweeps will be positive but weaker than the correlation in the presence of hard sweeps.

When we examine the correlation between neutral diversity within *D. pseudoobscura* and recombination rate at loci less likely to have experienced hard sweeps (Fig. 3, "no sweeps"), the association does not appear especially different from the full set of loci. For autosomal loci, the slope of the regression line of fourfold degenerate site π against local recombination rate did not differ significantly between the full set of loci and the subset of loci without hard sweeps

(Fig. 3b; ANCOVA, $F = 0.514$, df $= 1$, $p = 0.47$ for the interaction of recombination rate by locus category). However, the y-intercept was higher for the set of loci without hard sweeps compared to the set of all autosomal loci (ANCOVA, $F = 27.05$, df $= 1$, $p = 2.08 \times 10^{-7}$, for locus category effect). The higher y-intercept indicates an overall higher level of neutral variation in the absence of potential recent hard sweeps. In contrast, the subset of X chromosome loci exhibited significantly different slopes (Fig. 3c; ANCOVA, $F = 31.43$, df $= 1$, $p = 2.25 \times 10^{-8}$ for the interaction of recombination rate x locus category) between all X chromosome loci and the subset of loci without hard sweeps). Given the limited sample size of X chromosome loci, we approach X vs. autosome comparisons with caution in this case study, but we emphasize that this is an important area for future work.

Finally, we considered whether these differences observed between the loci less likely to have experienced hard sweeps and the full set of loci might occur by chance. We compared the observed outcomes for each chromosome to the range of possible outcomes for any random subset of loci. To compare the results to the distribution of possible outcomes, we took random subsamples of the full set of genes (10,000 sets of loci randomly sampled with replacement), where the number of sampled genes was equal to the number of loci in the original data set without hard sweeps (107 autosomal and 44 X chromosome loci). By randomly resampling from the full set of genes from each chromosome, we find that the differences observed between the full set and the set without the effects of sweeps could result from sampling. Bootstrapping and examining the difference between correlation coefficients tell us about the probability that the observed *difference* in correlation coefficients might be observed by chance. Relative to an approach of comparing observed correlations directly, this approach of sampling the difference in correlation coefficients is more robust to wide confidence intervals around correlation coefficients.

Using the correlation between fourfold degenerate site π and recombination rate in the set of loci with no hard sweeps, we compared this value to the correlation observed in each random sample. Figure 4 shows the distribution of differences between observed Pearson's R in each random sample and observed Pearson's R in the set of loci with no hard sweeps. The means of both the autosomal and X chromosome distributions in Fig. 4 differ from zero (autosomal t-test: $t = 49.007$, df $= 9{,}999$, $p < 2.2 \times 10^{-16}$; X chromosome t-test: $t = -134.4$, df $= 9{,}999$, $p < 2.2 \times 10^{-16}$), suggesting that the difference in observed correlation in the set of loci without hard sweeps is unlikely to have occurred by chance. However, the distributions in Fig. 4 include zero, suggesting that the association between neutral variation and recombination in the set of loci without sweeps falls within the distribution of randomly selected sets of loci (Fig. 4), Together, these observations suggest that removing the effects of hard sweeps does not drastically change the relationship between neutral diversity and recombination rate, though it may alter the strength of this relationship.

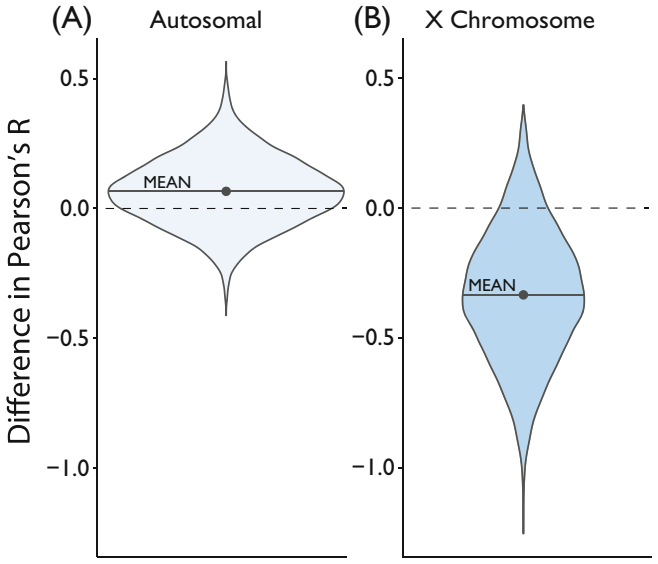

Fig. 4 Comparison of the set of loci with no hard sweeps to the sampling distribution of correlation coefficients between fourfold degenerate site π and recombination rate. Using the correlation between fourfold degenerate site π and recombination rate in the set of loci with no hard sweeps ("no sweeps R"), we compared this value to the correlation observed in random samples. Here, we show the distribution of difference between observed Pearson's *R* in each random sample and the "no sweeps R" for autosomal (**a**) and X chromosome (**b**) loci. Each distribution above includes 10,000 sets of loci that were randomly sampled with replacement from the full set of autosomal or X chromosome loci. The randomly sampled sets contained the same number of genes as the set of loci with no hard sweeps for that genomic region (107 autosomal loci and 44 X chromosome loci)

3 Discussion

The case study presented here aims to empirically examine the effects of linked selection via background selection and soft sweeps on within-species patterns of genetic diversity using a strategy that reduces the effects of recent hard sweeps. We demonstrate how this strategy can be applied to sister taxa to identify which loci have potentially experienced recent hard sweeps. This method can be applied to any pair of diverging taxa where sequence data is available, and the taxa are sufficiently similar at the nucleotide level to allow for direct comparison of coding sequences. That said, we next explore the important limitations of this approach and use these limitations to underscore areas where future work is needed.

First, the approach used in our case study is perhaps conservative in that it excludes all genes with fixed differences, even though many of these loci may be affected by background selection or soft sweeps, and several may in fact have never experienced a hard sweep. In fact, because background selection decreases the effective population size (N_e) at a locus and increases the relative effect of drift,

fixed differences are likely to arise under the influence of background selection. Lineage sorting—or the sorting of ancestral polymorphism by chance—may also yield fixed differences that did not involve sweeps. Additionally, we cannot exclude that ancient hard sweeps, predating the split of the two species, may have long-lasting effects on neutral diversity that still contribute to the association between recombination and diversity. Under an infinite alleles model, recovery of heterozygosity after a sweep occurs in a far shorter time frame than the divergence time between *D. pseudoobscura* and *D. miranda*, but it is unclear how this translates to recovery of π across the span of a gene. Another important note about this particular case study is that *D. miranda* does not hybridize with *D. pseudoobscura*. The strategy for excluding sweeps would need to be re-evaluated in species pairs with recent gene flow, which could allow new mutations in either lineage to be exchanged with the other, reducing the possibility that a sweep would be detectable based on a fixed difference. Further, we cannot definitively exclude effects from hard sweeps that occurred beyond the untranslated regions of the genes studied. Though LD decays rapidly in *Drosophila*, even in regions of low recombination (Langley et al. 2000), it is difficult to exclude the possibility that LD within a gene may have been affected by a strong or recent sweep in a very nearby region.

Obtaining empirical evidence of linked selection beyond hard sweeps remains challenging. As reviewed above (see Sect. 1.3), genome-wide patterns of variation in many taxa are broadly consistent with expectations of background selection being a major driver of genome-wide patterns of diversity (e.g., McVicker et al. 2009; Hernandez et al. 2011; Lohmueller et al. 2011; Charlesworth 2012b; Charlesworth and Campos 2014; Comeron 2014; Renzette et al. 2016; Pouyet et al. 2018). However, model-based approaches have often lacked comparative assessments of the explanatory power of sweeps-only models vs. joint models of background selection and sweeps together (though see Elyashiv et al. 2016). Modeling approaches also depend on expectations built from empirical observations, but we have little empirical data on linked selection outside of hard selective sweeps. To our knowledge, there are currently no standard approaches for gathering direct empirical evidence of the operation of background selection. Despite the limitations of the approach applied in the case study above, it provides an empirical strategy for examining how selection reduces linked neutral variation after minimizing possible effects of recent hard sweeps. Much work remains to resolve the long-standing debate over the relative contributions of different types of linked selection to patterns of genetic diversity (e.g., Stephan 2010; Jensen 2014; Renzette et al. 2016). To fully understand the impacts and interactions of types of linked selection, we need to thoughtfully combine theoretical and empirical approaches.

Methods such as the one discussed in the case study above are necessary to gather important empirical observations to inform and test theoretical work toward disentangling the multiple types of linked selection that affect patterns of variation throughout *Drosophila* genomes and in insect genomes more broadly. Simulations allow us to vary one parameter at a time, but empirical studies allow us to develop our intuition of the range of various parameters in nature and create simulations that model after empirically observed patterns. For instance, recent simulations show that

understanding how the spatial arrangement of genes changes background selection may help in learning how its signatures differ from those of sweeps (Schrider 2020). Similarly, we know that the landscape of recombination shapes the action of linked selection, so we need empirical assessments of recombination rate. Flies of the *pseudoobscura* group examined in the case study above are established model systems for the evolutionary genetics of recombination, and multiple empirical studies have analyzed and discussed the *D. pseudoobscura* recombination landscape and its evolution (Korunes and Noor 2019; Kulathinal et al. 2008; Samuk et al. 2020; Smukowski and Noor 2011; Smukowski Heil et al. 2015; Stevison et al. 2011; Stevison and Noor 2010). For understanding linked selection in diverse insect systems, we need broader sampling of recombination maps and their variation within insect species. For example, several species of social insects have been discovered to have some of the highest recombination rates among plants and animals, and the effects of recombination and linked selection in social insects may be an exciting area of future work (Jones et al. 2019; Liu et al. 2017; Meznar et al. 2010; Sirviö et al. 2011; Wallberg et al. 2015; Wilfert et al. 2007). As recombination rates vary greatly across species and even within species (Stapley et al. 2017), empirical studies of recombination and recombination rate variation are likely to play a large role in the future of linked selection studies.

4 Future Perspectives

In the field of population genomics, both within insects and more broadly, the challenges highlighted here exemplify recent advances and important areas for future work in the study of linked selection and its roles in driving patterns of nucleotide variation. Recent methodological and empirical work has improved our understanding of linked selection, but we are just beginning to appreciate the complex interactions of genetic variants with each other, with the recombination landscape, and with environments. While there has been major progress in understanding and documenting some forms of selective sweeps, understanding background selection and soft selective sweeps remains an area with substantial room for growth. Here, we have explored the study of these different types of linked selection, with particular emphasis on the body of work devoted to linked selection in *Drosophila*. As demonstrated with our case study of *Drosophila pseudoobscura*, many challenges remain in distinguishing types of linked selection. Moving forward, the field of insect population genetics must emphasize developing new empirical methods for interpreting signals of linked selection, jointly considering models of background selection and sweeps together, and understanding recombination rate variation. New empirical insights will continue to emerge with increased genomic sampling of diverse taxa, including the sampling of species with diverse recombination landscapes. As discussed above, the role of background selection in non-recombining regions and the diversity of sex determination systems in insects suggest that an area of exciting future development may be linked selection on

non-recombining or dosage-dependent sex chromosomes. Future empirical work in these areas will help inform methodological and theoretical advances as we aim to understand the impacts and interactions of types of linked selection.

References

Agrawal AF, Hartfield M. Coalescence with background and balancing selection in systems with bi- and uniparental reproduction: contrasting partial asexuality and selfing. Genetics. 2016;202 (1):313–26. https://doi.org/10.1534/genetics.115.181024.

Aguade M, Miyashita N, Langley CH. Reduced variation in the yellow-achaete-scute region in natural populations of *Drosophila melanogaster*. Genetics. 1989;122(3):607–15. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1203734&tool=pmcentrez&rendertype=abstract

Alves I, Šrámková Hanulová A, Foll M, Excoffier L. Genomic data reveal a complex making of humans. PLoS Genet. 2012;8(7):e1002837. https://doi.org/10.1371/journal.pgen.1002837.

Andolfatto P. Contrasting patterns of X-linked and autosomal nucleotide variation in Drosophila melanogaster and *Drosophila simulans*. Mol Biol Evol. 2001;18(3):279–90. http://www.ncbi.nlm.nih.gov/pubmed/11230529

Andolfatto P. Adaptive evolution of non-coding DNA in Drosophila. Nature. 2005;437 (7062):1149–52. https://doi.org/10.1038/nature04107.

Begun DJ, Aquadro CF. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature. 1992;356(6369):519–20. https://doi.org/10.1038/356519a0.

Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh Y-P, Hahn MW, et al. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. PLoS Biol. 2007;5(11):e310. https://doi.org/10.1371/journal.pbio.0050310.

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, et al. Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet. 2004;74 (6):1111–20. https://doi.org/10.1086/421051.

Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, et al. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome Res. 2015;25(11):1656–65. https://doi.org/10.1101/gr.196485.115.

Charlesworth B. The effects of deleterious mutations on evolution at linked sites. Genetics. 2012a;190(1):5–22. https://doi.org/10.1534/genetics.111.134288.

Charlesworth B. The role of background selection in shaping patterns of molecular evolution and variation: evidence from variability on the Drosophila X chromosome. Genetics. 2012b;191 (1):233–46. https://doi.org/10.1534/genetics.111.138073.

Charlesworth B, Campos JL. The relations between recombination rate and patterns of molecular variation and evolution in Drosophila. Annu Rev Genet. 2014;48(1):383–403. https://doi.org/10.1146/annurev-genet-120213-092525.

Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. Genetics. 1993;134(4):1289–303. http://www.genetics.org/content/134/4/1289.short

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, et al. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science. 2005;307(5717):1928–33. https://doi.org/10.1126/science.1107239.

Comeron JM. Background selection as baseline for nucleotide variation across the Drosophila genome. PLoS Genet. 2014;10(6):e1004434. https://doi.org/10.1371/journal.pgen.1004434.

Cutter AD, Choi JY. Natural selection shapes nucleotide polymorphism across the genome of the nematode *Caenorhabditis briggsae*. Genome Res. 2010;20(8):1103–11. https://doi.org/10.1101/gr.104331.109.

Cutter AD, Payseur BA. Genomic signatures of selection at linked sites: unifying the disparity among species. Nat Rev Genet. 2013;14(4):262–74. https://doi.org/10.1038/nrg3425.

de Groot NG, Otting N, Doxiadis GGM, Balla-Jhagjhoorsingh SS, Heeney JL, van Rood JJ, et al. Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. Proc Natl Acad Sci. 2002;99(18):11748–53. https://doi.org/10.1073/pnas.182420799.

DeGiorgio M, Huber CD, Hubisz MJ, Hellmann I, Nielsen R. SweepFinder2: increased sensitivity, robustness and flexibility. Bioinformatics. 2016;32(12):1895–7. https://doi.org/10.1093/bioinformatics/btw051.

Elyashiv E, Sattath S, Hu TT, Strutsovsky A, McVicker G, Andolfatto P, et al. A genomic map of the effects of linked selection in Drosophila. PLoS Genet. 2016;12(8):e1006130. https://doi.org/10.1371/journal.pgen.1006130.

Fay JC, Wu C-I. Hitchhiking under positive Darwinian selection. Genetics. 2000;155(3):1405–13.

Felsenstein J. The evolutionary advantage of recombination. Genetics. 1974;78(2):737–56.

Fu YX, Li WH. Statistical tests of neutrality of mutations. Genetics. 1993;133(3):693–709.

Halldorsson BV, Palsson G, Stefansson OA, Jonsson H, Hardarson MT, Eggertsson HP, et al. Characterizing mutagenic effects of recombination through a sequence-level genetic map. Science. 2019;363(6425):eaau1043. https://doi.org/10.1126/science.aau1043.

Harris RB, Sackman A, Jensen JD. On the unfounded enthusiasm for soft selective sweeps II: examining recent evidence from humans, flies, and viruses. PLoS Genet. 2018;14(12):e1007859. https://doi.org/10.1371/journal.pgen.1007859.

Hermisson J, Pennings PS. Soft sweeps. Genetics. 2005;169(4):2335–52. https://doi.org/10.1534/genetics.104.036947.

Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, et al. Classic selective sweeps were rare in recent human evolution. Science. 2011;331(6019):920–4. https://doi.org/10.1126/science.1198878.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science. 2006;313(5783):101–4. https://doi.org/10.1126/science.1126121.

Huber CD, DeGiorgio M, Hellmann I, Nielsen R. Detecting recent selective sweeps while controlling for mutation rate and background selection. Mol Ecol. 2016;25(1):142–56. https://doi.org/10.1111/mec.13351.

Hutter S, Li H, Beisswanger S, De Lorenzo D, Stephan W. Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosomewide single nucleotide polymorphism data. Genetics. 2007;177(1):469–80. https://doi.org/10.1534/genetics.107.074922.

Jackson BC, Campos JL, Haddrill PR, Charlesworth B, Zeng K. Variation in the intensity of selection on codon bias over time causes contrasting patterns of base composition evolution in Drosophila. Genome Biol Evol. 2017;9(1):102–23. https://doi.org/10.1093/gbe/evw291.

Jensen JD. On the unfounded enthusiasm for soft selective sweeps. Nat Commun. 2014;5:5281. https://doi.org/10.1038/ncomms6281.

Jones JC, Wallberg A, Christmas MJ, Kapheim KM, Webster MT. Extreme differences in recombination rate between the genomes of a solitary and a social bee. Mol Biol Evol. 2019;36(10):2277–91. https://doi.org/10.1093/molbev/msz130.

Kaiser VB, Charlesworth B. The effects of deleterious mutations on evolution in non-recombining genomes. Trends Genet. 2009;25(1):9–12. https://doi.org/10.1016/j.tig.2008.10.009.

Kaplan NL, Hudson RR, Langley CH. The "hitchhiking effect" revisited. Genetics. 1989;123(4):887–99. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1203897&tool=pmcentrez&rendertype=abstract

Kohn MH, Fang S, Wu C-I. Inference of positive and negative selection on the 5′ regulatory regions of Drosophila genes. Mol Biol Evol. 2004;21(2):374–83. https://doi.org/10.1093/molbev/msh026.

Kondrashov AS. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. Hum Mutat. 2003;21(1):12–27. https://doi.org/10.1002/humu.10147.

Korunes KL, Noor MAF. Pervasive gene conversion in chromosomal inversion heterozygotes. Mol Ecol. 2019;28(6):1302–15. https://doi.org/10.1111/mec.14921.

Kulathinal RJ, Bennett SM, Fitzpatrick CL, Noor MAF. Fine-scale mapping of recombination rate in Drosophila refines its correlation to diversity and divergence. Proc Natl Acad Sci. 2008;105(29):10051–6. https://doi.org/10.1073/pnas.0801848105.

Langley CH, Lazzaro BP, Phillips W, Heikkinen E, Braverman JM. Linkage disequilibria and the site frequency spectra in the su(s) and su(wa) regions of the *Drosophila melanogaster* X chromosome. Genetics. 2000;156(4):1837–52. http://www.genetics.org/content/156/4/1837.short

Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics. 2009;25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324.

Liu H, Jia Y, Sun X, Tian D, Hurst LD, Yang S. Direct determination of the mutation rate in the bumblebee reveals evidence for weak recombination-associated mutation and an approximate rate constancy in insects. Mol Biol Evol. 2017;34(1):119–30. https://doi.org/10.1093/molbev/msw226.

Lohmueller KE, Albrechtsen A, Li Y, Kim SY, Korneliussen T, Vinckenbosch N, et al. Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. PLoS Genet. 2011;7(10):e1002326. https://doi.org/10.1371/journal.pgen.1002326.

Maynard Smith J, Haigh J. The hitch-hiking effect of a favourable gene. Genet Res. 1974;23(1):23–35. http://www.ncbi.nlm.nih.gov/pubmed/4407212

McGaugh SE, Heil CSS, Manzano-Winkler B, Loewe L, Goldstein S, Himmel TL, et al. Recombination modulates how selection affects linked sites in Drosophila. PLoS Biol. 2012;10(11):e1001422. https://doi.org/10.1371/journal.pbio.1001422.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303. https://doi.org/10.1101/gr.107524.110.

McVicker G, Gordon D, Davis C, Green P. Widespread genomic signatures of natural selection in hominid evolution. PLoS Genet. 2009;5(5):e1000471. https://doi.org/10.1371/journal.pgen.1000471.

Messer PW, Petrov DA. Population genomics of rapid adaptation by soft selective sweeps. Trends Ecol Evol. 2013;28(11):659–69. https://doi.org/10.1016/j.tree.2013.08.003.

Meznar ER, Gadau J, Koeniger N, Rueppell O. Comparative linkage mapping suggests a high recombination rate in all honeybees. J Hered. 2010;101(Suppl 1):S118–26. https://doi.org/10.1093/jhered/esq002.

Nachman MW. Single nucleotide polymorphisms and recombination rate in humans. Trends Genet. 2001;17(9):481–5. https://doi.org/10.1016/S0168-9525(01)02409-X.

Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in humans. Genetics. 2000;156(1):297–304.

Nordborg M, Charlesworth B, Charlesworth D. The effect of recombination on background selection. Genet Res. 1996;67(2):159–74. http://www.ncbi.nlm.nih.gov/pubmed/8801188

Pennings PS, Hermisson J. Soft sweeps II – molecular population genetics of adaptation from recurrent mutation or migration. Mol Biol Evol. 2006;23(5):1076–84. https://doi.org/10.1093/molbev/msj117.

Phung TN, Huber CD, Lohmueller KE. Determining the effect of natural selection on linked neutral divergence across species. PLoS Genet. 2016;12(8):e1006199. https://doi.org/10.1371/journal.pgen.1006199.

Pouyet F, Aeschbacher S, Thiéry A, Excoffier L. Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. Elife. 2018;7:e36317. https://doi.org/10.7554/eLife.36317.

Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol. 2010;20(4):R208–15. https://doi.org/10.1016/j.cub.2009.11.055.

Renzette N, Kowalik TF, Jensen JD. On the relative roles of background selection and genetic hitchhiking in shaping human cytomegalovirus genetic diversity. Mol Ecol. 2016;25(1):403–13. https://doi.org/10.1111/mec.13331.

Roze D. Background selection in partially selfing populations. Genetics. 2016;203(2):937–57. https://doi.org/10.1534/genetics.116.187955.

Samuk K, Manzano-Winkler B, Ritz KR, Noor MAF. Natural selection shapes variation in genome-wide recombination rate in Drosophila pseudoobscura. Curr Biol. 2020;30(8):1517–1528.E6. https://doi.org/10.1016/j.cub.2020.03.053.

Schrider DR. Background selection does not mimic the patterns of genetic diversity produced by selective sweeps. BioRxiv. 2020; https://doi.org/10.1101/2019.12.13.876136.

Schrider DR, Kern AD. Soft sweeps are the dominant mode of adaptation in the human genome. Mol Biol Evol. 2017;34(8):1863–77. https://doi.org/10.1093/molbev/msx154.

Schrider DR, Shanku AG, Kern AD. Effects of linked selective sweeps on demographic inference and model selection. Genetics. 2016;204(3):1207–23. https://doi.org/10.1534/genetics.116.190223.

Sella G, Petrov DA, Przeworski M, Andolfatto P. Pervasive natural selection in the Drosophila genome? PLoS Genet. 2009;5(6):e1000495. https://doi.org/10.1371/journal.pgen.1000495.

Sirviö A, Johnston JS, Wenseleers T, Pamilo P. A high recombination rate in eusocial Hymenoptera: evidence from the common wasp Vespula vulgaris. BMC Genet. 2011;12:95. https://doi.org/10.1186/1471-2156-12-95.

Slotte T. The impact of linked selection on plant genomic variation. Brief Funct Genomics. 2014;13(4):268–75. https://doi.org/10.1093/bfgp/elu009.

Smukowski Heil CS, Ellison C, Dubin M, Noor MAF. Recombining without hotspots: a comprehensive evolutionary portrait of recombination in two closely related species of Drosophila. Genome Biol Evol. 2015;7(10):2829–42. https://doi.org/10.1093/gbe/evv182.

Smukowski CS, Noor MAF. Recombination rate variation in closely related species. Heredity. 2011;107(6):496–508. https://doi.org/10.1038/hdy.2011.44.

Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Philos Trans R Soc B. 2017;372(1736) https://doi.org/10.1098/rstb.2016.0455.

Stephan W. Genetic hitchhiking versus background selection: the controversy and its implications. Philos Trans R Soc Lond B Biol Sci. 2010;365(1544):1245–53. https://doi.org/10.1098/rstb.2009.0278.

Stephan W. Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. Mol Ecol. 2016;25(1):79–88. https://doi.org/10.1111/mec.13288.

Stephan W, Langley CH. Molecular genetic variation in the centromeric region of the X chromosome in three *Drosophila ananassae* populations. I. Contrasts between the vermilion and forked loci. Genetics. 1989;121(1):89–99. http://www.genetics.org/content/121/1/89.abstract

Stevison LS, Noor MAF. Genetic and evolutionary correlates of fine-scale recombination rate variation in *Drosophila persimilis*. J Mol Evol. 2010;71(5–6):332–45. https://doi.org/10.1007/s00239-010-9388-1.

Stevison LS, Hoehn KB, Noor MAF. Effects of inversions on within- and between-species recombination and divergence. Genome Biol Evol. 2011;3:830–41. https://doi.org/10.1093/gbe/evr081.

Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123(3):585–95. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1203831&tool=pmcentrez&rendertype=abstract

Takahashi A, Liu Y-H, Saitou N. Genetic variation versus recombination rate in a structured population of mice. Mol Biol Evol. 2004;21(2):404–9. https://doi.org/10.1093/molbev/msh030.

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc Natl Acad Sci. 2001;98(16):9161–6. https://doi.org/10.1073/pnas.151244298.

Torres R, Stetter MG, Hernandez RD, Ross-Ibarra J. The temporal dynamics of background selection in non-equilibrium populations. Genetics. 2020;214(4):1019–30. https://doi.org/10.1534/genetics.119.302892.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43(1):11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43.

Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. Annu Rev Genet. 2013;47(1):97–120. https://doi.org/10.1146/annurev-genet-111212-133526.

Wallberg A, Glémin S, Webster MT. Extreme recombination frequencies shape genome variation and evolution in the honeybee, *Apis mellifera*. PLoS Genet. 2015;11(4):e1005189. https://doi.org/10.1371/journal.pgen.1005189.

Wang RL, Hey J. The speciation history of Drosophila pseudoobscura and close relatives: inferences from DNA sequence variation at the period locus. Genetics. 1996;144(3):1113–26. http://www.genetics.org/content/144/3/1113.abstract

Wilfert L, Gadau J, Schmid-Hempel P. Variation in genomic recombination rates among animal taxa and the case of social insects. Heredity. 2007;98(4):189–97. https://doi.org/10.1038/sj.hdy.6800950.

Wilson Sayres MA, Lohmueller KE, Nielsen R. Natural selection reduced diversity on human Y chromosomes. PLoS Genet. 2014;10(1):e1004064. https://doi.org/10.1371/journal.pgen.1004064.

Wright S. Evolution in mendelian populations. Genetics. 1931;16(2):97–159.