

PERSPECTIVE

# On the Trail of Linked Selection

Emily B. Josephs<sup>1</sup>, Stephen I. Wright<sup>2\*</sup>

**1** Department of Evolution and Ecology, University of California, Davis, California, United States of America,

**2** Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada

\* [stephen.wright@utoronto.ca](mailto:stephen.wright@utoronto.ca)

Distinguishing the relative roles of positive and negative selection along with demographic history in shaping genetic diversity has been a decades-long endeavor. Understanding the forces structuring genetic variation informs us not only about the factors maintaining diversity but also about the fundamental evolutionary parameters that influence natural populations, including the rate and strength of positive selection, the deleterious genetic load experienced by populations, and the factors driving genome evolution. Most attempts at modeling demography, positive selection, or negative selection typically do so by ignoring the contribution of the other forces. Because multiple forces are acting simultaneously, these inferences likely overestimate the role of single evolutionary forces and can lead to biased interpretations. In this issue, Elyashiv et al. [1] make important advances towards addressing this problem by presenting a novel approach to simultaneously estimate the parameters of positive and negative selection based on the spatial patterns of neutral genetic variation and apply this method to *Drosophila*.

Elyashiv et al.'s [1] approach takes advantage of a long-held prediction that patterns of neutral variation can reveal the action of selection at linked sites. Maynard Smith and Haigh first recognized the potential impacts of linked selection in 1974 when they modeled the reduction in neutral genetic diversity that would occur near a site under positive selection, often termed a selective sweep, and argued that this process would reduce diversity genome-wide given sufficiently high rates of positive selection [2]. Support for their theory was first obtained when molecular population genetic data from *Drosophila* showed a correlation between diversity and recombination rate, which was consistent with the idea that selective sweeps are common and have a stronger impact on neutral diversity in regions of low recombination [3,4]. However, Charlesworth and colleagues proposed an alternate explanation: negative selection against the continuous influx of new deleterious mutations (background selection) will also reduce nearby neutral diversity, causing a similar correlation between recombination and genetic diversity [5]. This model has been used to show that many of the major patterns of variation in diversity across the *Drosophila* genome are explainable by this process [6]. Since this alternative mechanism was first proposed, understanding the relative importance of positive and negative selection on linked neutral diversity has been an ongoing challenge in population genetics [7].

Although evidence has continued to mount for the negative relationship between recombination and diversity due to linked selection [8], the expectation that both sweeps and background selection are strongest in regions of low recombination made it difficult to identify what selective forces are responsible for lower diversity in low recombination regions. A number of proposed signals in the data have been thought to be unique predictions that could help distinguish the models, including the allele frequency spectrum in regions of low recombination [3,9] and the exact shape of the relationship between recombination and diversity [10]. However, by broadening background selection predictions to include a class of slightly



 OPEN ACCESS

**Citation:** Josephs EB, Wright SI (2016) On the Trail of Linked Selection. *PLoS Genet* 12(8): e1006240. doi:10.1371/journal.pgen.1006240

**Editor:** Nick H. Barton, Institute of Science and Technology Austria (IST Austria), AUSTRIA

**Published:** August 18, 2016

**Copyright:** © 2016 Josephs, Wright. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** EBJ is supported by an NSF Postdoctoral Research Fellowship in Biology (IOS-1523733). SIW is supported by an NSERC Discovery Grant. The funders had no role in preparation of the manuscript.

**Competing Interests:** I have read the journal's policy and have the following conflicts. Since beginning to co-author this manuscript, but prior to publication, Emily Josephs has joined the lab of Graham Coop, an author on the subject of this Perspective. The authors have declared that no other competing interests exist.

deleterious mutations, it became clear that many features thought to be unique to positive selection could also be explained by background selection [11]. Until recently, there has been little direct evidence that selective sweeps affect neutral diversity on a genome-wide scale.

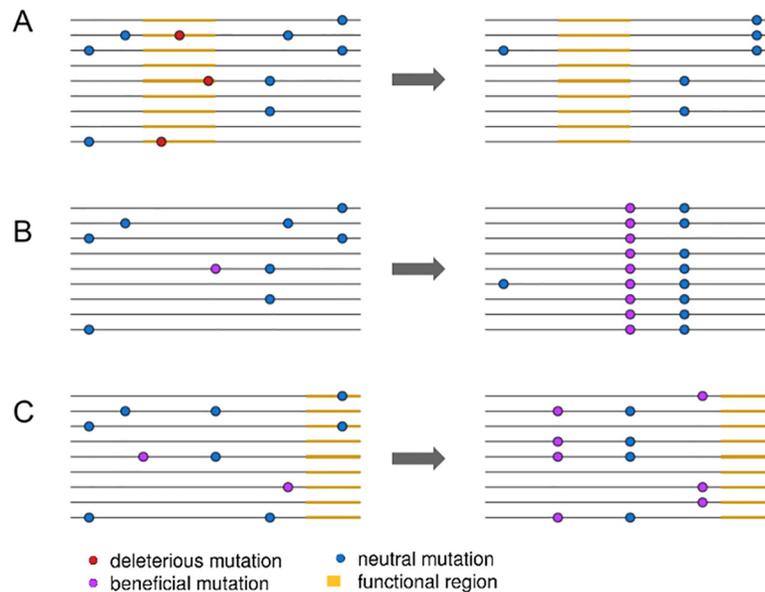
A key step forward came from the insight that, because selective sweeps result from the fixation of new mutations, fixed substitutions at functional sites provide a unique indicator for the presence of selective sweeps (Fig 1A and 1B) [12,13]. This approach is especially powerful if background selection can be accounted for by comparing to neutral substitutions [14]. For example, neutral diversity is lower around fixed nonsynonymous substitutions than fixed synonymous substitutions in the *Drosophila* genome, providing important evidence that selective sweeps have been common during *Drosophila* adaptation [14]. However, while these patterns were used to infer rates of positive selection, fully quantifying the extent of selective sweeps requires explicitly separating the effects of background selection from sweeps.

Elyashiv et al. [1] used the presence of fixed substitutions to distinguish between sweeps and background selection in a joint modelling framework. Specifically, they model background selection as a function of the density of sites expected to be under purifying selection, such as those in exons and untranslated regions (UTRs). In contrast, they model selective sweeps as a function of the density of fixed substitutions at functional sites. This approach then allows them to estimate deleterious mutation rates directly from data instead of relying on experimentally estimated rates, which often rely on many assumptions and can vary across genotypes [15]. They can then, ultimately, estimate selection coefficients and mutation rates for positively and negatively selected alleles. While Elyashiv et al. [1] are not the first to jointly model the effects of selective sweeps and background selection [16], their approach allows them to generally separate out the effects of sweeps from those of background selection and to do so in a manner that can be applied directly to genomic data.

Using their method, Elyashiv et al. [1] estimate that linked selection reduces diversity by 70%–90% across the genome and that both background selection and selective sweeps contribute in substantial ways to the structuring of genetic diversity. Their joint modeling approach validates previous conclusions that *Drosophila* experiences a high rate of positive selection but also highlights the important role of purifying selection in explaining large-scale patterns of genetic diversity across the genome. These findings cement the view that linked selection is crucial for shaping genomic diversity, so much so that it may be wise to begin including parameters of linked selection—essentially “genomic demography”—in basic population genomic models the same way that standard demographic models are included. Beyond this, approaches based on linked selection may open up new opportunities for investigating the distribution of fitness effects (DFE) of new mutations by allowing us to estimate more precisely the DFE of sites that are so rare they don’t segregate in population genomic samples but do have impacts on linked neutral diversity.

Although Elyashiv et al. [1] have made significant progress in our understanding of linked selection, there are still some avenues for improvement. Their method is dependent on knowledge of functional sites that accurately predict where purifying and positive selection act in the genome, and, as they note, selection outside of these regions will affect their predictions. Genes and genomic regions also differ in their fraction of selectively constrained sites, and so explicitly incorporating information about cross-species divergence into models of purifying selection [17] may also help refine the model parameters.

More crucially, while conditioning their inference of selective sweeps compared to background selection on the presence of fixed substitutions, other forms of positive selection that do not involve fixations, such as polygenic adaptation, are likely parameterized as background selection (Fig 1C). If polygenic adaptation is common [18], this should incorrectly increase the strength of background selection inferred in the model. Recent work in humans implies that



**Fig 1. An illustration of three linked selection scenarios.** In (A), the removal of deleterious mutations in a functional region, or background selection, reduces nearby linked diversity. In (B), the fixation of a new beneficial mutation by positive selection reduces nearby linked diversity in a process called a selective sweep. In (C), positive selection increases the frequency of two beneficial mutations, reducing diversity but not involving the fixation of any selected mutations.

doi:10.1371/journal.pgen.1006240.g001

large fractions of the genome may in fact be subject to directional selection on quantitative traits [19], further indicating that polygenic selection can have genome-wide effects. Indeed, Elyashiv et al. [1] estimate a much higher deleterious mutation rate than appears plausible given estimates from the literature, suggesting an important role for polygenic adaptation in driving down diversity near functional sites. Developing techniques to properly account for polygenic adaptation will not only improve accuracy for inferring selection parameters and the effects of linked selection but also provide further insights into the extent to which positive selection influences the structuring of genetic diversity genome-wide.

Beyond extending linked selection approaches to better deal with poor annotations and polygenic adaptation, estimating the effects of linked selection in additional species beyond *Drosophila* holds great promise. While patterns of linked selection generally differ between *Drosophila* and other species, the causes of these differences remain uncertain [8]. The test developed by Sattath and colleagues [14] to detect selective sweeps by comparing diversity around functional substitutions to diversity around nonfunctional substitutions has since been applied to additional species, including humans, mice, maize, and *Capsella grandiflora* [20–23], with mixed evidence for recurrent sweeps. However, this method’s use of synonymous substitutions as a neutral control may not be effective in larger genomes, in which nonsynonymous substitutions may be more common in regions of low constraint, reduced background selection, and higher neutral diversity, masking the signals of selective sweeps [24]. Because Elyashiv et al.’s [1] method can use the distribution of selected sites present in the actual genome under investigation, it may be more robust to variation in constraint across the genome and be useful for investigating linked selection in large genomes.

Patterns of linked selection are also likely to differ among species of varying population sizes. There is already evidence that the effects of linked selection on neutral diversity increase with species census size [25], and Elyashiv et al.’s [1] joint modelling approach may help tease

apart why these effects vary with effective population size ( $N_e$ ). For example, joint modelling approaches could help improve the parameterization of positive selection in species with smaller  $N_e$ , in which signals may be weaker due to lower rates of positive selection. However, we may also expect a larger proportion of adaptation in species with high  $N_e$  to involve multiple alleles, altering our expectations of how linked selection will shape nearby neutral diversity [26,27]. This further highlights the need to develop approaches that independently parameterize polygenic adaptation, although this remains a difficult challenge.

Furthermore, the relative importance of linked selection on neutral diversity may vary with mating system [5,28]. Reduced effective recombination rates in selfing species are expected to play an important role in limiting both the removal of deleterious alleles by purifying selection [29,30] and the fixation of adaptive alleles due to positive selection [31,32], but this is difficult to accurately quantify. Intriguingly, there is evidence that there is a much larger drop in genome-wide diversity in selfers than in outcrossers due to the strong effects of linked selection [33]. However, the effects of linked selection may be especially difficult to investigate in selfers because low recombination will mean that subsequent events of linked selection mask those that occurred previously. Joint modelling approaches may help investigate these patterns and possibly help quantify the extent to which positive and purifying selection are reduced in selfing lineages.

Overall, the comprehensive model of linked selection developed by Elyashiv et al. [1] both reinforces the growing consensus that linked selection is an important contributor to genetic variation and provides a useful framework for investigating not only the specific impacts of linked selection but also the general selective parameters that shape genomic variation. As we have outlined above, their work opens up a number of further opportunities to investigate the action of selection across a genome and provides an important step forward in the ongoing challenge of quantifying the relative importance of different evolutionary forces.

## References

1. Elyashiv E, Sattath S, Hu TT, Strustovsky A, McVicker G, Andolfatto P, et al. A genomic map of the effects of linked selection in *Drosophila*. *PLoS Genet* 12(8): e1006130.
2. Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. *Genet Res.* 1974; 23: 23–35.
3. Aguade M, Miyashita N, Langley CH. Reduced variation in the yellow-achaete-scute region in natural populations of *Drosophila melanogaster*. *Genetics*. 1989; 122: 607–615. PMID: [17246506](#)
4. Begun DJ, Aquadro CF. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature*. 1992; 356: 519–520. PMID: [1560824](#)
5. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. *Genetics*. 1993; 134: 1289–1303. PMID: [8375663](#)
6. Charlesworth B. Background selection and patterns of genetic diversity in *Drosophila melanogaster*. *Genet Res.* 1996; 68: 131–149. PMID: [8940902](#)
7. Stephan W. Genetic hitchhiking versus background selection: the controversy and its implications. *Philos Trans R Soc Lond B Biol Sci.* 2010; 365: 1245–1253. doi: [10.1098/rstb.2009.0278](#) PMID: [20308100](#)
8. Cutter AD, Payseur BA. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat Rev Genet.* 2013; 14: 262–274. doi: [10.1038/nrg3425](#) PMID: [23478346](#)
9. Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics*. 1995; 140: 783–796. PMID: [7498754](#)
10. Innan H, Stephan W. Distinguishing the hitchhiking and background selection models. *Genetics*. 2003; 165: 2307–2312. PMID: [14704207](#)
11. 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: e1002326. doi: [10.1371/journal.pgen.1002326](#) PMID: [22022285](#)
12. Andolfatto P. Hitchhiking effects of recurrent beneficial amino acid substitutions in the *Drosophila melanogaster* genome. *Genome Res.* 2007; 17: 1755–1762. PMID: [17989248](#)

13. Macpherson JM, Sella G, Davis JC, Petrov DA. Genomewide spatial correspondence between nonsynonymous divergence and neutral polymorphism reveals extensive adaptation in *Drosophila*. *Genetics*. 2007; 177: 2083–2099. PMID: [18073425](#)
14. Sattath S, Elyashiv E, Kolodny O, Rinott Y, Sella G. Pervasive adaptive protein evolution apparent in diversity patterns around amino acid substitutions in *Drosophila simulans*. *PLoS Genet*. 2011; 7: e1001302. doi: [10.1371/journal.pgen.1001302](#) PMID: [21347283](#)
15. Schrider DR, Houle D, Lynch M, Hahn MW. Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster*. *Genetics*. 2013; 194: 937–954. doi: [10.1534/genetics.113.151670](#) PMID: [23733788](#)
16. Kim Y, Stephan W. Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics*. 2000; 155: 1415–1427. PMID: [10880499](#)
17. McVicker G, Gordon D, Davis C, Green P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet*. 2009; 5: e1000471. doi: [10.1371/journal.pgen.1000471](#) PMID: [19424416](#)
18. Pritchard JK, Di Rienzo A. Adaptation—not by sweeps alone. *Nat Rev Genet*. 2010; 11: 665–667. doi: [10.1038/nrg2880](#) PMID: [20838407](#)
19. Field Y, Boyle EA, Telis N, Gao Z, Gaulton KJ, Golan D, et al. Detection of human adaptation during the past 2,000 years [Internet]. *bioRxiv*. 2016. p. 052084. doi: [10.1101/052084](#)
20. 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: 920–924. doi: [10.1126/science.1198878](#) PMID: [21330547](#)
21. Halligan DL, Kousathanas A, Ness RW, Harr B, Eöry L, Keane TM, et al. Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. *PLoS Genet*. 2013; 9: e1003995. doi: [10.1371/journal.pgen.1003995](#) PMID: [24339797](#)
22. Williamson RJ, Josephs EB, Platts AE, Hazzouri KM, Haudry A, Blanchette M, et al. Evidence for widespread positive and negative selection in coding and conserved noncoding regions of *Capsella grandiflora*. *PLoS Genet*. 2014; 10: e1004622. doi: [10.1371/journal.pgen.1004622](#) PMID: [25255320](#)
23. Beissinger TM, Wang L, Crosby K, Durvasula A, Hufford MB, Ross-Ibarra J. Recent demography drives changes in linked selection across the maize genome. *Nat Plants*. 2016; 2: 16084. doi: [10.1038/nplants.2016.84](#) PMID: [27294617](#)
24. Enard D, Messer PW, Petrov DA. Genome-wide signals of positive selection in human evolution. *Genome Res*. 2014; 24: 885–895. doi: [10.1101/gr.164822.113](#) PMID: [24619126](#)
25. Corbett-Detig RB, Hartl DL, Sackton TB. Natural Selection Constrains Neutral Diversity across A Wide Range of Species. *PLoS Biol*. 2015; 13: e1002112. doi: [10.1371/journal.pbio.1002112](#) PMID: [25859758](#)
26. Pennings PS, Hermisson J. Soft Sweeps II—Molecular Population Genetics of Adaptation from Recurrent Mutation or Migration. *Mol Biol Evol*. 2006; 23: 1076–1084. PMID: [16520336](#)
27. Barton N. Understanding adaptation in large populations. *PLoS Genet*. 2010; 6: e1000987. doi: [10.1371/journal.pgen.1000987](#) PMID: [20585547](#)
28. Charlesworth D, Wright SI. Breeding systems and genome evolution. *Curr Opin Genet Dev*. 2001; 11: 685–690. PMID: [11682314](#)
29. Muller HJ. THE RELATION OF RECOMBINATION TO MUTATIONAL ADVANCE. *Mutat Res*. 1964; 106: 2–9. PMID: [14195748](#)
30. Arunkumar R, Ness RW, Wright SI, Barrett SCH. The evolution of selfing is accompanied by reduced efficacy of selection and purging of deleterious mutations. *Genetics*. 2015; 199: 817–829. doi: [10.1534/genetics.114.172809](#) PMID: [25552275](#)
31. Hartfield M, Glémin S. Hitchhiking of deleterious alleles and the cost of adaptation in partially selfing species. *Genetics*. 2014; 196: 281–293. doi: [10.1534/genetics.113.158196](#) PMID: [24240529](#)
32. Kamran-Disfani A, Agrawal AF. Selfing, adaptation and background selection in finite populations. *J Evol Biol*. 2014; 27: 1360–1371. doi: [10.1111/jeb.12343](#) PMID: [24601989](#)
33. Coop G. Does linked selection explain the narrow range of genetic diversity across species? [Internet]. *bioRxiv*. 2016. p. 042598. doi: [10.1101/042598](#)