The short-term, genome-wide effects of indirect selection deserve study: a response to Charlesworth and Jensen (2022)

Zachariah Gompert[1,2*], Jeffrey L. Feder[3], Patrik Nosil[4]

[1] Department of Biology, Utah State University, Logan, UT 84322, USA

[2] Ecology Center, Utah State University, Logan, UT 84322, USA

[3] Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556, USA

[4] CEFE, Univ Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier 3, Montpellier, 34293, France

Corresponding author: Zachariah Gompert
Department of Biology
5305 Old Main Hill
Utah State University
Logan, UT 84322-5305
Phone: (307) 760-3858
Email: zach.gompert@usu.edu

Running Title: Response to Charlesworth and Jensen

We recently published a paper quantifying the genome-wide consequences of natural selection, including the effects of indirect selection due to the correlation of genetic regions (neutral or selected) with directly-selected regions (Gompert *et al.*, 2022). In their critique of our paper, Charlesworth & Jensen (2022) make two main points: (i) indirect selection is equivalent to hitchhiking and thus well documented (i.e., our results are not novel), and (ii) that we do not demonstrate the source of linkage disequilibrium (LD) between SNPs and the *Mel-Stripe* locus in the *Timema cristinae* experiment we analyze. As we discuss in detail below, neither of these are substantial criticisms of our work.

First, indirect selection and hitchhiking are related but not equivalent concepts. Our focus on the short-term consequences of long-range LD (i.e., across the genome and not restricted to tightly-linked regions) sets our work apart from most population genetic studies of hitchhiking, which emphasize longer-term consequences of short-range LD. In this context, we take the opportunity here to expand on why readers of *Molecular Ecology* might be interested in the short-term consequences of indirect selection.

Second, the consequences of LD for indirect selection that our study focused on hold regardless of the source(s) of LD. Drift and local migration, as well as natural selection itself, can all create a degree of long-range LD between selected sites and neutral unlinked variants, particularly in small populations (Fig. 1a). We demonstrated the existence of this long-range LD with respect to the *Mel-Stripe* locus in *Timema cristinae* (which controls color), as well as in two other cases involving the *Agouti* gene and coat color in deer mice (*Peromyscus maniculatus*) and the *Ectodysplasin* gene and body armor in stickleback fish (*Gasterosteus aculeatus*). Although this long-range LD is transient compared to that between physically linked sites, our point was that selection on a target locus can deterministically affect the direction of allele frequency change at even unlinked neutral sites for short periods of time (one or a few generations), before being dissipated by recombination and random assortment. Moreover, new long-range associations are constantly reforming between selected and other sites in finite populations due, in part, to the sources creating LD discussed by Charlesworth

& Jensen (2022). In a metaphoric sense, selected sites will be constantly picking up new unlinked passengers in the genome and taking them for short rides. Thus, while we appreciate the points raised by Charlesworth & Jensen (2022), they do not detract from our main thesis. We elaborate on these misunderstandings below in hopes that it clarifies the role and significance of indirect selection in evolution for the readers of *Molecular Ecology*.

Indirect selection has a long history and is related to but distinct from hitchhiking

The concept of indirect selection was articulated by Pearson (1903) over 70 years before the term hitchhiking was introduced by Maynard Smith & Haigh (1974). Indirect selection was first used to describe the within-generation change in the distribution of phenotypes for traits not directly affecting fitness but instead that were correlated with traits directly under selection. This terminology thus distinguishes *selection for* specific traits because of their causal effect on fitness (direct selection) from *selection of* traits because of a non-causal correlation with fitness (indirect selection) (Sober, 1984). Such correlations might arise because of LD between loci affecting different traits, pleiotropy, or a common effect of the environment on sets of traits. Thus, in contrast to standard usage of hitchhiking (discussed more below), indirect selection does not emphasize physical linkage nor does it focus on the long-term consequences of selection on patterns of molecular variation.

Indirect selection was then brought to the widespread attention of evolutionary biologists in the 1980s, for example, by Lande & Arnold (1983) introducing the use of multiple regression to disentangle direct versus indirect selection on traits. Lande & Arnold (1983) demonstrated that the selection gradient ($\boldsymbol{\beta}$) obtained from regressing fitness measurements on a set of possibly correlated traits predicts the short-term (i.e., one generation) evolutionary response to selection for those traits in terms of changes in breeding values (or similarly, polygenic scores): $\Delta\bar{\boldsymbol{z}} = \mathbf{G}\boldsymbol{\beta}$ (this is the multivariate breeder's equation). Here, $\bar{\boldsymbol{z}}$ denotes the vector of mean breeding values (across traits) and **G** is the genotypic variance-covariance matrix. Lande & Arnold (1983) further showed how the selection gradient estimated in this way can be used to partition the selection differential for each trait (S_i, i.e., the change in the mean trait value following selection) into selection caused by the direct effect of the trait on fitness and indirect selection caused by selection on other correlated characters: $S_i = P_{ii}\beta_i + \sum_{j\neq i} P_{ij}\beta_j$ (here $\boldsymbol{P}$ denotes the phenotypic variance-covariance matrix). If the phenotypic effects of genetic loci are known, additional theory exists to connect direct and indirect selection on traits to selection on loci (Kimura & Crow, 1978; Walsh & Lynch, 2018). These and other models comprise a well-developed quantitative theory for evolutionary change caused by direct and indirect selection. Since this time, thousands of studies have applied the methods of Lande & Arnold (1983) to measure indirect selection (reviewed in Kingsolver *et al.*, 2001, 2012) or used such measurements to analyze patterns of evolutionary change within species or divergence among them (e.g., Schluter, 1996; Walsh & Blows, 2009; Lucas *et al.*, 2018), with recent extensions in evolutionary biology and agriculture involving genomic prediction of breeding values and measuring indirect selection on genetic loci (e.g., Gompert *et al.*, 2014; Thurman & Barrett, 2016; Gompert *et al.*, 2017; Exposito-Alonso *et al.*, 2019; Watson *et al.*, 2019; Stern *et al.*, 2021).

The concept of hitchhiking was introduced by Maynard Smith & Haigh (1974) to describe the increase in frequency of an allele present on a chromosome where a new beneficial mutation arises that is caused by LD between the allele and the beneficial mutation. This concept emphasizes physical linkage and the evolutionary consequences of new mutations, especially in the context of genetic diversity. As such, hitchhiking can be viewed as a consequence of indirect selection, where physical linkage is involved. Hitchhiking was extended to the case of linked, indirect selection against deleterious alleles, that is background selection, by Charlesworth *et al.* (1993). More recent work has investigated soft or partial selective sweeps and hitchhiking caused by selection on standing genetic variation (Hermisson & Pennings, 2005; Prezeworski *et al.*, 2005; Barrett & Schluter, 2008; Hermisson & Pennings, 2017). This combined body of theory led to the development of various statistical methods to detect selective sweeps (including recurrent sweeps) (e.g., Kim & Stephan, 2002; Jensen *et al.*, 2005; Nielsen *et al.*, 2005; DeGiorgio *et al.*, 2016), which have been successfully used to identify evidence of sweeps in a number of empirical systems (e.g., Jensen *et al.*, 2008; Carneiro *et al.*, 2014; Garud *et al.*, 2015; Moest *et al.*, 2020).

Thus, while indirect selection and hitchhiking are clearly closely related concepts, they have distinct historical usage and the distinction between them is more than semantic. We here argue that indirect selection is best used to describe the immediate indirect effects of selection on one trait or gene caused by correlations with fitness (or other selected traits or genes), whereas hitchhiking best describes the longer-term consequences of this for patterns of molecular variation especially when physical linkage is involved. With that said, the recent (alternative) definition of hitchhiking given by Charlesworth & Jensen (2021) based on the additive genetic covariance between a trait (or gene) and fitness via the Price-Robertson equation (Robertson, 1968; Price *et al.*, 1970) is in essence equivalent to our treatment of indirect selection and the way in which the term indirect selection has been mostly used in evolutionary biology since Pearson (1903). Nonetheless, this newer definition does not coincide with the historical or dominant usage of hitchhiking in the literature nor does this affect the substance of the arguments in our original manuscript.

Why study short-term long-range consequences of indirect selection?

We think that this (admittedly subtle) distinction between indirect selection and hitchhiking helps to clarify the novelty of our recent work (Gompert *et al.*, 2022), and motivates further empirical work on indirect selection. As noted by Charlesworth & Jensen (2022), and as made clear in our original manuscript (Gompert *et al.*, 2022), the long-term consequences of indirect selection, and specifically of hitchhiking (including background selection) on patterns of genetic variation are well established, with especially compelling evidence for the impact of background selection on patterns of diversity in many organisms (e.g., Begun & Aquadro, 1992; Comeron, 2014; Charlesworth & Campos, 2014; Pouyet *et al.*, 2018). Also as noted by Charlesworth & Jensen (2022), this of course implies that indirect selection must operate on shorter timescales. Despite this theoretical truism, the importance of indirect selection for short-term evolutionary or ecological dynamics (e.g., one or a few geneations) has received far less empirical attention. Specifically, this cannot be determined from static studies of molecular variation alone but instead requires studies of contemporary evolution in natural populations, ideally combined with field and lab experiments that measure selection and its immediate consequences across the genome.

In Gompert *et al.* (2022) we aimed to begin to fill this knowledge gap and provide evidence that indirect selection has the potential to have measurable short-term impacts on evolution. These impacts include indirect selection on genetic variants not physically linked to a focal directly selected locus. Although we do focus on the color locus *Mel-Stripe* in *T. cristinae*, the only system criticized by Charlesworth & Jensen (2022), we stress that the conclusions noted above rest on the collective analysis of data sets in stick-insects, stickleback and mice. Moreover, the existence (though not quantitative extent) of genome-wide indirect selection is a mathematical certainty in any finite population experiencing selection. In the stick-insect example, we showed that LD (measured by r^2) was > 0.01 for 3% of all SNPs not on the same chromosome as *Mel-Stripe* and that LD exceeded 0.10 for 64 SNPs. Importantly and also distinct from most work on hitchhiking, we extended the results in stick insects to show that LD with numerous (unknown) causal variants likely affecting fitness results in even more genomically widespread indirect selection (i.e., due to polygenic selection, see Gompert *et al.*, 2022 for details).

Should we invest effort in investigating possible short-term (i.e., one or several generations) long-distance (i.e., among distant loci including unlinked loci) effects of indirect selection? As pointed out both by Charlesworth & Jensen (2022) and ourselves (Gompert *et al.*, 2022), recombination and independent assortment cause LD to decay over time. Thus, long-distance LD is not expected to persist for long periods of time. Nonetheless, the short-term, long-distance effects of indirect selection should be considered for several core reasons. First, incorporating the short-term effects of indirect selection provides a more mathematically precise and conceptually appropriate model of evolution. For example, with indirect selection, the expected single generation change in allele frequency for a neutral locus in LD with loci affecting fitness is not 0 (i.e., $E[\Delta p] \neq 0$), as it would be in standard models of genetic drift (Wright, 1931). This difference might be slight in most cases, but still it is real, even if past drift was the cause of LD. Such indirect selection might be especially important in field or lab studies, and more generally in any situation involving strong selection in small populations, such as natural colonization events of new environments or accidental introductions or organisms to new habitats by humans. Second, in some cases, the short-term effects of indirect selection could have meaningful ecological consequences. Even for physically unlinked loci, LD can persist for several generations (it decays by half each generation), and there is now compelling evidence that even short-term evolutionary change, especially in fluctuating or heterogeneous environments, can have ecological consequences for entire communities (e.g., Farkas *et al.*, 2013; Hendry, 2020). Third, some species have long generation times (e.g., many tree species), and thus patterns of LD might persist for hundreds or thousands of years in absolute time (i.e., the entire modern period of human-induced climate change). Thus, although further empirical and theoretical work is required to compare the effects of indirect selection across timescales and genomic regions, we do not think there is sufficient evidence at present to completely dismiss the study of short-term, long-distance effects.

The consequences of LD do not depend on the causes of LD

The other major critique raised by Charlesworth & Jensen (2022) was that we did not determine the source of LD in the *T. cristinae* experiment. As noted above, this was because the consequences of LD for indirect selection, the focus of our study, are not conditional on the causes of LD in the population (Fig. 1b). Thus, this is not a valid criticism of our study. Still, we do agree that this is an interesting question worth considering and we take the opportunity to do so here. We agree that each of the six mechanisms identified by Charlesworth & Jensen (2022) could contribute to the observed LD. We think most of these are unlikely, but comment on all briefly here, before discussing what we think are the most likely explanations in more detail. The first possibility raised was that SNPs in LD with *Mel-Stripe* were in fact physically linked to *Mel-Stripe* but were spuriously placed on other chromosomes. The overall quality of our reference genome (mostly large scaffolds derived from a Chicago genomic library) and linkage groups (constructed from crosses) makes this unlikely to be the primary cause of LD (Nosil *et al.*, 2018) (more recent, but as of yet unpublished chromosome-level genome assemblies further support our conclusion). Moreover, the deer mouse (*P. maniculatus*) and threespine stickleback (*G. aculeatus*) genomes would also have to be of quite poor quality for this mechanism to explain our full results, which is unlikely given the data and resources invested in these reference genomes (Kenney-Hunt *et al.*, 2014; Brown *et al.*, 2018; Long *et al.*, 2019; Reid *et al.*, 2021). Lastly, even if this were true, the observed LD would still result in indirect selection, but such selection would not be spread as widely across the genome.

The second possibility raised is genetic drift, which Charlesworth & Jensen (2022) dismiss but we think is a highly probable mechanism as we discuss below. Their third mechanism is very recent admixture. While this is unlikely in light of past genetic analyses in *T. cristinae* (which were not considered by Charlesworth & Jensen, 2022) (e.g., Nosil *et al.*, 2012; Soria-Carrasco *et al.*, 2014; Riesch *et al.*, 2017), we do think that population structure and ongoing gene flow (combined with selection) contribute to the observed LD. The fourth mechanism was a recent bottleneck, which is possible but not supported by observations in the field over the last 29 years, where population size of the source population fluctuated but has not dropped extensively (e.g., Nosil *et al.*, 2018 for a long-term study).

Fifth, Charlesworth & Jensen (2022) suggested but mostly dismissed an ongoing selective sweep at *Mel-Stripe*; we agree that this is unlikely given evidence for long-term balancing selection on this locus (Lindtke *et al.*, 2017; Nosil *et al.*, 2018; Villoutreix *et al.*, 2020). Lastly, Charlesworth & Jensen (2022) suggest a possible role for epistatic interactions for fitness between *Mel-Stripe* and other genetic variants genome wide. We agree with them that this is unlikely to be a major mechanism (i.e., it could apply, but most likely to only one or a few loci) (e.g., Villoutreix *et al.*, 2022), nonetheless evidence for a non-trivial role in selection generating long-distance LD does exist in other systems, such as lodgepole pine (MacLachlan *et al.*, 2021).

Thus, in terms of likely explanations for the observed LD, we think genetic drift and gene flow warrant prime consideration. In their critique, Charlesworth & Jensen (2022) are mostly dismissive of genetic drift because of the size of the source *T. cristinae* population from which the experimental population was sampled ($\sim$1000 individuals). However, their criticism fails to distinguish between census and effective population size. Past work in *T. cristinae* suggests that Ne is considerably smaller than the census population size, and probably on the order of 10% of the census size (Soria-Carrasco *et al.*, 2014; Nosil *et al.*, 2018), consistent with broad patterns in other organisms (Frankham, 1995). Simple simulations show that with Ne $\sim$100–200, drift could readily create the patterns observed (Fig. 2). Second, gene flow likely contributes to the observed patterns of LD. Past work, including demographic modeling from population genetic data suggest gene flow at least at smaller spatial scales, including for differentiated regions under divergent selection where gene flow will create LD (Nosil & Crespi, 2004; Nosil *et al.*, 2006, 2012, 2018). However, we re-iterate that the causes of LD have little bearing on the core conclusions of our original study (Gompert *et al.*, 2022).

Conclusions

We want to conclude by thanking Charlesworth & Jensen (2022) for their useful comments on Gompert *et al.* (2022), stimulating us to clarify the issues discussed above concerning our original manuscript. We think that our disagreements with them mostly reflect differences in emphasis between studies of indirect selection (prominent in field studies of natural selection) and hitchhiking (prominent in population genetics and molecular evolution). We hope that this discussion provides cross-talk between these important sub-disciplines. We fully agree with Charlesworth & Jensen (2022) that the long-term consequences of hitchhiking, including background selection, are well known and that we are now at the point where background selection should be part of our standard null models in molecular evolution (Comeron, 2017). We also concur that resolving the processes creating long-range LD in *Timema* and other systems is important. However, it is the subsequent effects that selection may have indirectly on unlinked sites due to the existence of long-range LD, as demonstrated in the manipulative transplant experiment for *T. cristinae*, that we highlight as the take-home message of our paper. Thus, we stand by our original argument that the short-term, long-distance consequences of indirect selection on genetic loci deserve more attention. Despite the rich body of theory associated with indirect selection and hitchhiking, we think additional theoretical work would be useful on the conditions under which sufficient LD is expected to arise among unlinked or loosely linked loci to have non-trivial effects on short-term evolution via indirect selection. Ultimately, it will be such theoretical developments combined with further empirical work that reveals the relevance of this process for ecology and evolution, particularly for small populations experiencing polygenic and varying selection on many traits.

References

Barrett RD, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**, 38–44.

Begun DJ, Aquadro CF (1992) Levels of naturally occurring dna polymorphism correlate with recombination rates in *D. melanogaster*. *Nature*, **356**, 519–520.

Brown J, Crivello J, O'Neill RJ (2018) An updated genetic map of *Peromyscus* with chromosomal assignment of linkage groups. *Mammalian Genome*, **29**, 344–352.

Carneiro M, Rubin CJ, Di Palma F, *et al.* (2014) Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science*, **345**, 1074–1079.

Charlesworth B, Campos JL (2014) The relations between recombination rate and patterns of molecular variation and evolution in *Drosophila*. *Annual Review of Genetics*, **48**, 383–403.

Charlesworth B, Jensen J (2022) Some complexities in interpreting apparent effects of hitchhiking: a commentary on Gompert, Feder & Nosil (2021). *Molecular Ecology*, **XX**, XXX–XXX.

Charlesworth B, Jensen JD (2021) Effects of selection at linked sites on patterns of genetic variability. *Annual Review of Ecology, Evolution, and Systematics*, **52**, 177–197.

Charlesworth B, Morgan M, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics*, **134**, 1289–1303.

Comeron JM (2014) Background selection as baseline for nucleotide variation across the *Drosophila* genome. *PLoS Genetics*, **10**, e1004434.

Comeron JM (2017) Background selection as null hypothesis in population genomics: insights and challenges from *Drosophila* studies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **372**, 20160471.

DeGiorgio M, Huber CD, Hubisz MJ, Hellmann I, Nielsen R (2016) SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics*, **32**, 1895–1897.

Exposito-Alonso M, Burbano HA, Bossdorf O, Nielsen R, Weigel D (2019) Natural selection on the *Arabidopsis thaliana* genome in present and future climates. *Nature*, **573**, 126–129.

Farkas TE, Mononen T, Comeault AA, Hanski I, Nosil P (2013) Evolution of camouflage drives rapid ecological change in an insect community. *Current Biology*, **23**, 1835–1843.

Frankham R (1995) Effective population size/adult population size ratios in wildlife: a review. *Genetics Research*, **66**, 95–107.

Garud NR, Messer PW, Buzbas EO, Petrov DA (2015) Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLoS Genetics*, **11**, e1005004.

Gompert Z, Comeault AA, Farkas TE, *et al.* (2014) Experimental evidence for ecological selection on genome variation in the wild. *Ecology Letters*, **17**, 369–379.

Gompert Z, Egan SP, Barrett RD, Feder JL, Nosil P (2017) Multilocus approaches for the measurement of selection on correlated genetic loci. *Molecular Ecology*, **26**, 365–382.

Gompert Z, Feder JL, Nosil P (2022) Natural selection drives genome-wide evolution via chance genetic associations. *Molecular Ecology*, **31**, 467–481.

Hendry AP (2020) *Eco-evolutionary dynamics*. Princeton university press.

Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, **169**, 2335–2352.

Hermisson J, Pennings PS (2017) Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods in Ecology and Evolution*, **8**, 700–716.

Jensen JD, Kim Y, DuMont VB, Aquadro CF, Bustamante CD (2005) Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics*, **170**, 1401–1410.

Jensen JD, Thornton KR, Andolfatto P (2008) An approximate Bayesian estimator suggests strong, recurrent selective sweeps in *Drosophila*. *PLoS Genetics*, **4**, e1000198.

Kenney-Hunt J, Lewandowski A, Glenn TC, *et al.* (2014) A genetic map of *Peromyscus* with chromosomal assignment of linkage groups (a *Peromyscus* genetic map). *Mammalian Genome*, **25**, 160–179.

Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, **160**, 765–777.

Kimura M, Crow JF (1978) Effect of overall phenotypic selection on genetic change at individual loci. *Proceedings of the National Academy of Sciences*, **75**, 6168–6171.

Kingsolver JG, Diamond SE, Siepielski AM, Carlson SM (2012) Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evolutionary Ecology*, **26**, 1101–1118.

Kingsolver JG, Hoekstra HE, Hoekstra JM, *et al.* (2001) The strength of phenotypic selection in natural populations. *The American Naturalist*, **157**, 245–261.

Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution*, pp. 1210–1226.

Lindtke D, Lucek K, Soria-Carrasco V, *et al.* (2017) Long-term balancing selection on chromosomal variants associated with crypsis in a stick insect. *Molecular Ecology*, **26**, 6189–6205.

Long AD, Baldwin-Brown J, Tao Y, *et al.* (2019) The genome of *Peromyscus leucopus*, natural host for lyme disease and other emerging infections. *Science Advances*, **5**, eaaw6441.

Lucas LK, Nice CC, Gompert Z (2018) Genetic constraints on wing pattern variation in *Lycaeides* butterflies: A case study on mapping complex, multifaceted traits in structured populations. *Molecular Ecology Resources*, **18**, 892–907.

MacLachlan IR, McDonald TK, Lind BM, Rieseberg LH, Yeaman S, Aitken SN (2021)

Genome-wide shifts in climate-related variation underpin responses to selective breeding in a widespread conifer. *Proceedings of the National Academy of Sciences*, **118**, e2016900118.

Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetics Research*, **23**, 23–35.

Moest M, Van Belleghem SM, James JE, *et al.* (2020) Selective sweeps on novel and introgressed variation shape mimicry loci in a butterfly adaptive radiation. *PLoS Biology*, **18**, e3000597.

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C (2005) Genomic scans for selective sweeps using SNP data. *Genome Research*, **15**, 1566–1575.

Nosil P, Crespi B (2004) Does gene flow constrain adaptive divergence or vice versa? a test using ecomorphology and sexual isolation in timema cristinae walking-sticks. *Evolution*, **58**, 102–112.

Nosil P, Crespi B, Sandoval C, Kirkpatrick M (2006) Migration and the genetic covariance between habitat preference and performance. *The American Naturalist*, **167**, E66–E78.

Nosil P, Gompert Z, Farkas TE, *et al.* (2012) Genomic consequences of multiple speciation processes in a stick insect. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 5058–5065.

Nosil P, Villoutreix R, de Carvalho CF, *et al.* (2018) Natural selection and the predictability of evolution in *Timema* stick insects. *Science*, **359**, 765–770.

Pearson K (1903) Mathematical contributions to the theory of evolution.—XI. on the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **200**, 1–66.

Pouyet F, Aeschbacher S, Thiéry A, Excoffier L (2018) Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. *Elife*, **7**, e36317.

Prezeworski M, Coop G, Wall JD (2005) The signature of positive selection on standing genetic variation. *Evolution*, **59**, 2312–2323.

Price GR, *et al.* (1970) Selection and covariance. *Nature*, **227**, 520–521.

Reid K, Bell MA, Veeramah KR (2021) Threespine stickleback: a model system for evolutionary genomics. *Annual Review of Genomics and Human Genetics*, **22**, 357–383.

Riesch R, Muschick M, Lindtke D, *et al.* (2017) Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology & Evolution*, **1**, 1–13.

Robertson A (1968) The spectrum of genetic variation, pp. 5–16 in population biology and evolution, edited by lewontin rc.

Schluter D (1996) Adaptive radiation along genetic lines of least resistance. *Evolution*, **50**, 1766–1774.

Sober E (1984) *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. MIT Press.

Soria-Carrasco V, Gompert Z, Comeault AA, *et al.* (2014) Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, **344**, 738–742.

Stern AJ, Speidel L, Zaitlen NA, Nielsen R (2021) Disentangling selection on genetically correlated polygenic traits via whole-genome genealogies. *The American Journal of Human Genetics*, **108**, 219–239.

Thurman TJ, Barrett RD (2016) The genetic consequences of selection in natural populations. *Molecular Ecology*, **25**, 1429–1448.

Villoutreix R, de Carvalho CF, Gompert Z, Parchman TL, Feder JL, Nosil P (2022) Testing for fitness epistasis in a transplant experiment identifies a candidate adaptive locus in *Timema* stick insects. *Philosophical Transactions of the Royal Society B*, **377**, 20200508.

Villoutreix R, de Carvalho CF, Soria-Carrasco V, *et al.* (2020) Large-scale mutation in the evolution of a gene complex for cryptic coloration. *Science*, **369**, 460–466.

Walsh B, Blows MW (2009) Abundant genetic variation + strong selection = multivariate genetic constraints: a geometric view of adaptation. *Annual Review of Ecology, Evolution and Systematics*, **40**, 41.

Walsh B, Lynch M (2018) *Evolution and selection of quantitative traits*. Oxford University Press.

Watson A, Hickey LT, Christopher J, Rutkoski J, Poland J, Hayes BJ (2019) Multivariate genomic selection and potential of rapid indirect selection with speed breeding in spring wheat. *Crop Science*, **59**, 1945–1959.

Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97.

Data availability

No new data were analyzed in this paper, but the *T. cristinae* DNA sequence data from Gompert et al. (2022) have been archived on NCBIs SRA (PRJNA356801) and the phenotypic data and genotype estimates for *T. cristinae* are available on Dryad (`https://doi.org/10.5061/dryad.m905qfv26`).

Code availability

Computer code for the new simulation analysis is available from `github` (`https://github.com/zgompert/TimemaPolygenicSelection/blob/main/SimMsLD.R`).

Author contributions

ZG, JF and PN wrote and revised the manuscript.

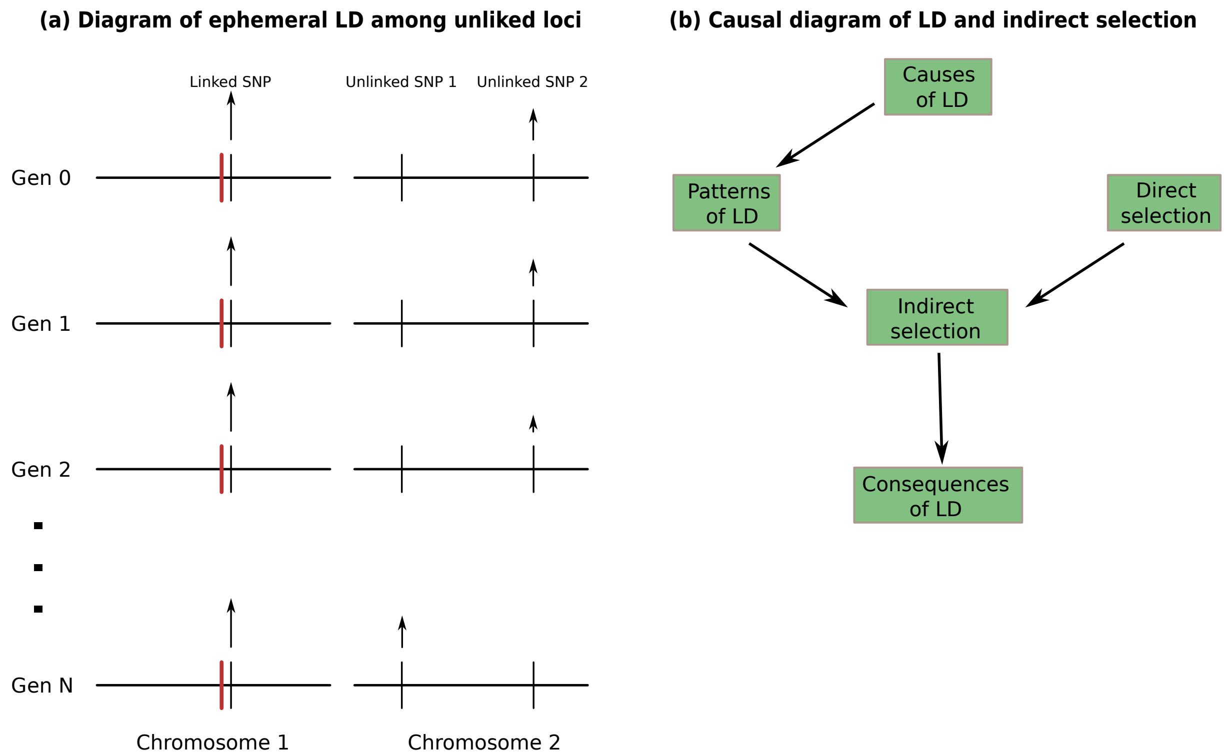

Figure 1: Hypothetical illustration of short-term, long-distance indirect selection (a). The vertical red line denotes a genetic variant causally affecting fitness (i.e., under direct selection). Vertical black lines denote a linked SNP and two unlinked SNPs. Arrows denote the degree of linkage disequilibrium (LD) between each SNP and the causal genetic variant, with larger arrows denoting higher levels of LD. Because of limited recombination, the linked SNP remains in high LD with the causal variant for an extended period of time (i.e., up to generation N, where N might denote tens or hundreds of generations). In contrast, because of free recombination, LD between unlinked SNP 2 and the causal variant decays over several generations. Nonetheless, during these few generations, indirect selection causes some of the allele frequency change at this locus to be deterministic and directional. Moreover, in this hypothetical example, LD is later created between the causal variant and a different unlinked SNP (unlinked SNP 1) because of some combination of drift, gene flow and selection. Thus, LD with unlinked variants is ephemeral but also constantly recreated, leading to perpetual indirect selection. Panel (b) depicts a causal diagram showing that, after accounting for patterns of LD, the causes of LD have no causal affect on the consequences of LD for indirect selection. These consequences only depend on patterns of LD and direct selection.

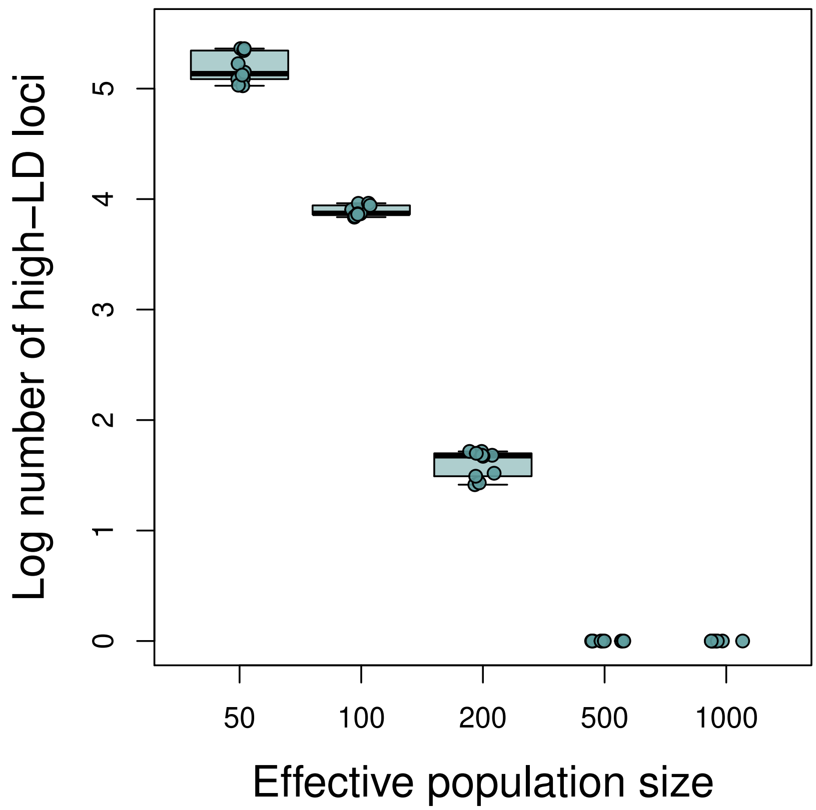

Figure 2: The effect of effective population size on drift's capacity to generate LD. Boxplots with overlain points show the number of genetic loci (e.g., SNPs) in high linkage disequilibrium (LD), here defined as $r^2 > 0.1$, with a focal locus in simulations of genetic drift. Results are shown for each of 10 simulations with effective population sizes of 50, 100, 200, 500 or 1000 individuals. Numbers are shown on a $\log_{10}$ scale, with the exception that all simulations with effective sizes of 500 and 1000 resulted in 0 high-LD SNPs ($\log_{10}(0)$ is $-\infty$ but 0 values are shown instead). Simulations involved a focal locus with minor allele frequency of 0.32 (our estimate of the minor allele frequency for *Mel-Stripe* in the *T. cristinae* population) and 7 million unlinked SNPs with allele frequencies drawn from the estimated allele frequencies at genome-wide SNPs in *T. cristinae*. We used binomial sampling to draw genotypes at each locus for each individual with independent draws across loci (i.e., we simulated unlinked loci). LD was then computed as the squared genotypic correlation (r^2). We conducted these simulations in R (version 4.0.2).