Using Patterns of Signed Linkage Disequilibria to Test for Epistasis in Flies and Plants

George Sandler¹, Stephen I. Wright¹,²*, Aneil F. Agrawal¹,²*

¹Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, ON M5S 3B2, Canada; ²Center for Analysis of Genome Evolution and Function, University of Toronto, 25 Willcocks Street, Toronto, ON M5S 3B2, Canada

*These authors made equal contributions.

Corresponding author

George Sandler

Email: george.sandler@mail.utoronto.ca
Abstract

Whether deleterious mutations affect fitness independently, or synergistically, remains an open question in evolutionary genetics. Previous work by Sohail et al. (2017) reported an abundance of negative linkage disequilibrium (LD) values among loss-of-function (LOF) mutations in several human and fruit fly datasets, a pattern the authors interpreted as evidence of negative synergistic epistasis. Here we re-visit this question in a population genomic dataset of plants (Capsella grandiflora), and a fruit fly (Drosophila melanogaster) dataset previously used by Sohail et al. When using synonymous sites as a control, as Sohail et al., we find that both species have significantly less positive LD at LOF sites than synonymous sites. However, LD is not significantly different from 0 for LOF mutations but is significantly positive for synonymous mutations in both species. We question the use of synonymous sites as an appropriate control when attempting to make inferences about LD at selected sites. We use simulations to show how admixture or mating bias towards physically proximal individuals can cause positive LD to build up among neutral mutations but has a much weaker effect on selected sites, regardless of the presence of epistasis. Finally, we use information from published biological networks to explore whether there is evidence for negative synergistic epistasis between interacting radical missense mutations. We report no significant enrichment for negative inter-network LD in C. grandiflora. However, we note a modest but significant enrichment of negative LD in D. melanogaster networks, suggestive of intra-network negative synergistic epistasis in this species.
Introduction

Linkage disequilibrium (LD), the association of different alleles across the genome, is a general feature of population genomic datasets, often revealing clues of ongoing evolutionary or demographic processes (McEvoy et al. 2011). For example, in finite populations, drift can be a ready source of LD, generating both positive and negative associations between alleles (Hill and Robertson 1968). While LD is usually described through unsigned statistics such as \( r^2 \), the sign of allelic associations can also provide useful information. Here we refer to positive associations as those between two ancestral alleles (or equivalently between two derived alleles), and negative associations as those between ancestral and derived alleles. Demographic processes such as admixture and population structure can create LD, where unlike drift in a single panmictic population, an overabundance of positive associations is expected between pairs of migrant alleles (Chakraborty and Weiss 1988; Stephens et al. 1994; Pfaff et al. 2001). Selective processes can also be a source of LD; for example, ongoing strong selective sweeps can be characterised by an elevation of unsigned LD around the sweeping haplotype (McVean 2007). Finally, unsigned LD can also be used to analyze patterns of recombination across the genome, as recombination is expected to break down any existing LD (Auton and McVean 2007).

LD can also build up due to selection against deleterious mutations in one of two different ways. First, Hill Robertson interference (resulting from the interaction of selection and drift) can cause negative associations to build up among deleterious mutations, if recombination between them is limited (Hill and Robertson 1966). In sexually reproducing organisms such as humans, this process has recently been suggested to build up negative LD among physically proximal, missense mutations (Garcia and Lohmueller 2020). Second, negative selection can cause LD among deleterious mutations to build up if epistasis is present (Kondrashov 1995; Sohail et al. 2017). Under the null model of multiplicative fitness, where each mutation contributes to a reduction in fitness independently of other mutations, LD is not expected to accumulate. Synergistic epistasis, where each additional deleterious mutation reduces
fitness by a greater magnitude, creates negative LD among deleterious mutations and vice versa for antagonistic epistasis (Kimura and Maruyama 1966; Kondrashov 1982).

Synergistic epistasis among deleterious mutations is of particular interest because such epistasis has several evolutionary consequences. For example, synergistic epistasis leads to a lowered mutation load under mutation-selection balance compared to multiplicative fitness (Kimura and Maruyama 1966; Crow and Kimura 1970; Crow and Kimura 1979). Relatedly, synergistic epistasis is predicted to favour the evolution of recombination by allowing organisms to prevent the buildup of deleterious mutations, providing a possible explanation to the long-standing question of why sexual reproduction is evolutionarily favoured (Kondrashov, 1982; see also Barton, 1995).

Despite considerable interest in the topic, empirical data on epistasis among deleterious mutations is limited. Most data come from microorganisms assayed in a lab setting. These studies have found that synergistic and antagonistic interactions are both common so that mean epistasis is close to zero (Elena and Lenski 1997; Agrawal and Whitlock 2010; Lalić and Elena 2012; Bank et al. 2015; Puchta et al. 2016).

A recent study by Sohail et al. (2017) used a different approach to make inferences about epistasis. They examined existing population genomic datasets to show that negative LD is present among rare loss of function mutations (LOF) in humans and fruit flies, a class of likely unconditionally deleterious mutations (Sohail et al. 2017). The authors compared LD between LOF mutations to LD between synonymous mutations (a putatively neutral control), demonstrating that in several datasets LD appeared to be more negative in the LOF mutation set. This pattern of negative LD was evident among LOF mutations that were far apart (most on separate chromosomes). Given these results, the authors argued that the observed patterns of LD are likely the product of synergistic epistasis among LOF mutations, rather than other processes such as Hill Robertson interference.

Here we expand upon the analyses conducted by Sohail et al., investigating LD among synonymous and LOF mutations in a published dataset of 182 whole genomes from a population of *Capsella grandiflora*. We find that negative LD is not present on average among LOF mutations in *C. grandiflora* and that LD among synonymous mutations is
considerably positive. Based on these results we revisited one of the previously
analyzed datasets of Sohail et al. consisting of 191 *Drosophila melanogaster* flies
sampled from an ancestral population in Zambia (Lack et al. 2015). Like Sohail et al.,
we found some initial evidence of negative LD between LOF mutations in flies; however,
upon closer inspection of the data, we question the interpretation of this finding. We also
explored the possibility of epistasis arising between interacting deleterious mutations by
analyzing LD among radical missense mutations in previously defined biological
pathways. Theory predicts that synergistic epistasis could arise as a by-product of
modularity in biological networks (Chiu et al. 2012), and such patterns may be obscured
in attempts to assess epistasis among all genes across the genome. However, limited
synergism within networks could still act to considerably alter mutation load, highlighting
the need to search for signatures of epistasis within interacting biological networks (Rice
1998). We use data on published biological networks from the KEGG database
(Kanehisa et al. 2016) and report some tentative evidence of synergistic epistasis within
pathways in *D. melanogaster* but not in *C. grandiflora*. Finally, we use simulations to
show how various processes such as admixture and continuous isolation by distance
can interact with multiplicative selection to skew patterns of LD among neutral and
deleterious mutations differently, complicating the use of LD as a metric to detect
epistasis or selective interference.
We applied the method of Sohail et al. to assess patterns of LD among both LOF and synonymous mutations in a dataset of 182 outbred diploid *C. grandiflora* individuals, and in a previously analyzed dataset of 191 *D. melanogaster* haploid embryos (population DPGP3). Briefly, we measured LD by assessing the over- or under-dispersion of deleterious (or synonymous) variants among individual genomes (see Materials and Methods). An under-dispersion of the deleterious variants implies negative LD (i.e. deleterious variants are found together less often than expected by chance).

In *C. grandiflora*, we found little evidence of negative LD among LOF mutations (Figure 1, Supplementary Table 1). LD was generally positive for LOF mutations, although considerably less so than for synonymous mutations under all allele frequency cut-offs considered. To remove LD from physically neighbouring sites, we calculated LD among sites in different 100Kb blocks instead of SNP-by-SNP, in which case we found that LD was consistently less positive between blocks for both synonymous and LOF mutations (Figure 1, Supplementary Table 1). The only case where the point estimate of mean LD among LOF mutations in *C. grandiflora* was negative was at an allelic cut-off of precisely 5 copies when calculating LD in 100Kb blocks (see Materials and Methods for more details). Because this gave the result most similar to that reported for flies and humans by Sohail et al. and also matched the allelic count cut off used in their study, we used a cut-off of no more than 5 alleles for all downstream analyses.

In the *D. melanogaster* dataset, we recovered similar results to Sohail et al. (Figure 1, Supplementary Table 1). LD was positive for synonymous mutations under all allele frequency cut offs and the point estimate for mean LD was negative for LOF mutations, except in the case of singletons where it was positive. Unlike in *C. grandiflora*, repeating the analysis in 100Kb blocks did not have much of an effect on the mean value of LD per pair of mutations, however, the small difference in all cases was a shift towards less positive LD, qualitatively the same pattern observed in *C. grandiflora*. The stronger effect of ‘blocking’ on LD implies that, relative to fruit flies, *C. grandiflora* has proportionally more physically close-by sites with strong positive LD (see below).
Figure 1. Mean pair-wise LD among LOF (red) and synonymous (blue) mutations at different allele count cut-offs. Solid lines indicate mean LD among all SNPs, dashed lines indicate LD calculated in 100Kb non-overlapping genomic blocks. Left, results for *C. grandiflora*, right, results for *D. melanogaster*.

We used two approaches to test whether LD among LOF mutations was significantly negative at an allelic cut-off of no more than 5. First, following Sohail et al. we used synonymous mutations in both species as a control group for patterns of LD. We subsampled synonymous mutations to the count of LOF mutations 1000 times and recalculated LD (both site by site and in 100Kb bins) to generate a null distribution of LD values. Only synonymous mutations with the same allelic count cut-off as for LOFs were used. Employing this method, we find LD is significantly less positive for LOF mutations than synonymous mutations in both *C. grandiflora* and *D. melanogaster* when calculating LD SNP-by-SNP (p < 0.002 and p < 0.002, 2-tailed *C. grandiflora*, *D. melanogaster* respectively) but only marginally so in 100Kb bins (p = 0.030 and p = 0.062, 2-tailed).

We also tested whether the LD among LOF mutations was itself significantly different from 0. We did this by taking both sets of LOF mutations and permuting their assignment among all individuals in the dataset. This method preserves the allele frequency at each locus while randomizing the associations among loci. This test suggested that LD among LOF mutations was not significantly different from 0 in either *C. grandiflora* or *D. melanogaster* when calculating LD SNP-by-SNP (p = 0.112 and p =
When we applied this permutation approach to synonymous mutations, we found that LD was significantly different from 0 in both species, when calculating LD SNP-by-SNP (p < 0.002 both species, 2-tailed), or in 100Kb bins (p = 0.326, 2-tailed). When we applied this permutation approach to synonymous mutations, we found that LD was significantly different from 0 in both species, when calculating LD SNP-by-SNP (p < 0.002 both species, 2-tailed), or in 100Kb bins (p = 0.326, 2-tailed), further verifying that LD was significantly positive among this set of mutations.

We next used PLINK to further assess the relationship between inter-mutation distance and LD for LOF mutations. We binned distance between LOF mutation pairs into seven categories: 100bp or less, 101-1000bp, 1001-10,000bp, 10,001-100,000bp, 100,001-1,000,000bp, >100,000,000bp and inter-chromosome. We visualised the relationship between LD and distance for every possible pair of mutations in two ways. First, like Sohail et al. we plotted the sum of LD for every bin of distance (Supplementary Figure 1). When visualised this way, LD appears to be positive between physically proximal mutations in both species (particularly in C. grandiflora) before becoming very negative for pairs of mutations that are found on different chromosomes. The observation of the point estimate of summed LD being negative for inter-chromosomal pairs was used as a key piece of evidence supporting the inference of synergistic epistasis by Sohail et al.

Though the sum of LD appears very negative for inter-chromosomal pairs of mutations, the variance around this estimate is very high as is shown by the 95% confidence intervals that overlap 0 in this distance bin (Supplementary Figure 1). The comparison among bins using “summed LD” is somewhat complicated by the imbalance between the number of datapoints found in each distance bin; the inter-chromosome comparison has several orders of magnitude more observations than any other distance bin, greatly increasing the uncertainty in the final sum of LD because the variance of a sum scales with the number of observations (i.e., random variables) contributing to the sum. To make comparisons across distance more easily interpretable, we plotted mean LD per pair of sites rather than summed LD (Figure 2). When visualised this way, LD appears to be significantly positive for physically proximal LOF mutations but decays to 0 by 10,000bp in both species. As noted above, this method of viewing the data also shows that the signal of positive LD between nearby mutations is enriched in C. grandiflora.
compared to *D. melanogaster*, i.e., the average LD in the smallest distance bin is higher in *C. grandiflora* and LD takes longer to fully decay (1000bp vs. 100bp).

![Figure 2. Distribution of LD between pairs of LOF mutations across different distance bins. Left *C. grandiflora*, right *D. melanogaster*. Only LOF mutations with an allelic count of five or less are included. Error bars show 95% confidence interval for mean LD within each distance bin, asterisks denote bins where the confidence interval does not overlap 0. Values under bars represent proportion of pairs of LOF mutations that fall into the specified distance bins.]

Following Sohail et al, our analysis above was motivated by the idea that synergistic epistasis could leave a population genomic signature of negative LD. However, these results do not indicate that LD is significantly negative for LOF mutations. Rather, LD is significantly less positive among LOF mutations than among synonymous mutations, but this pattern appears to be due to significantly positive LD at synonymous sites that likely exists for other reasons (see below). However, the analysis above aims to detect a signature of epistasis averaged over all mutations, yet epistasis may be weak and/or variable in sign (synergistic vs. antagonistic) for random pairs of mutations, precluding a clear LD signature. Epistasis, and the resulting LD signature, may be more consistent in
direction and stronger in strength among mutations in genes that are functionally
related. To address this possibility, we asked whether we could detect signals of negative LD
among deleterious mutations found within interacting gene networks. For these network
analyses, we used low frequency (no more than 5 copies in the population) radical
missense mutations as our set of candidate deleterious mutations in both datasets (see
Materials and Methods for details; there were too few LOFs to perform this analysis
using those variants). We also only considered LD calculated in 100Kb bins to minimize
the effect of positive LD between nearby mutations and to remove measurement of LD
between mutations within genes. We found little evidence for a major bias in the sign of
LD when we analyzed synonymous mutations segregating in KEGG networks. When
LD is plotted against the number of genes in each network, a pattern emerges in both
species wherein small networks tend to have variable mean LD estimates while large
networks tend to have LD near 0 (Supplementary Figure 2, Supplementary Table 2),
presumably reflecting higher sampling error in smaller networks. We permuted the
assignment of genes to each KEGG network 1000x and confirmed that the average of
all network mean LD values (hereafter, “average network mean LD”) was not
significantly different from a null distribution of LD (p = 0.524 C. grandiflora, p = 0.406 D.
melanogaster, 2-tailed, Supplementary Figure 2).
Figure 3. A,B) Mean LD among radical missense mutations affecting genes within interacting biological networks plotted against network size (as defined by number of genes within each network). LD was calculated in 100Kb blocks to minimize the effects of short-distance, and intra-genic interactions. Left data from C. grandiflora, right from D. melanogaster. C,D) Average network mean LD among radical missense mutations for deciles based on network size. Box-plots show the null distribution from permuted networks. In each permutation, networks were split into deciles based on bin size and...
the average of mean LD of all networks in each decile was calculated. Whiskers represent quartiles of average mean LD of permuted networks. True average network mean LD of each decile is overlaid in red. Insets show enlarged version of last four size class bins of plot to aid in visualization. E,F) The fraction of networks within each decile for which mean network LD is positive. Box plots represent the null distribution from permutations. The fraction of true networks with positive values in each decile is overlaid in red.

We repeated our network analyses with radical missense mutations assigned to the same KEGG networks. Unlike for synonymous SNPs, LD among radical SNPs, especially within small networks, seemed to be enriched for negative mean values of LD in both species (Figure 3 A,B, Supplementary Table 2). Permutation tests indicated that the average mean LD of networks for C. grandiflora was not significantly different from the null (p = 0.898, 2-tailed). In D. melanogaster, however, the average mean LD of networks was more negative than expected under the null (p = 0.042, 2-tailed), though the true significance could be debated given we have performed this test in two species; the pattern should be viewed as intriguing rather than compelling.

To help visualize the patterns in these data, we split networks into deciles with respect to network size and calculated average network mean LD. This reveals that the tendency towards negative LD in D. melanogaster is not driven by just small networks (Figure 3 D). In D. melanogaster, 6 of 10 network deciles had true values fall in the lower quartile of their respective null distributions with two of these within the lower 0.025th percentile of their null distributions; none of the deciles have true values that fall in the largest quartile of their respective null distributions. Consistent with our main statistical analysis, there is no clear pattern in C. grandiflora; most true values lie within the central part of their respective null distributions (Figure 3C).

One visually striking feature of Figure 3 A,B is the apparent preponderance of negative LD estimates among small networks. However, this is largely an artifact that is also apparent in the permuted networks (Figure 3 E,F). A high incidence of negative LD in small networks is expected because (1) small networks have few mutations and (2) the
neutral distribution of LD among rare variants is characterized by weakly negative LD among most pairs of mutations and strongly positive LD among a small fraction of pairs. Nonetheless, in *D. melanogaster* there appears to be a trend towards a preponderance of negative values of network mean beyond the null expectation; in 4 of 10 deciles the proportion of positive network mean LD values is in the lowest quartiles of their null distributions and only 1/10 deciles is in the highest quartile.

Our observations of pervasive positive LD among synonymous sites ran counter to our initial expectation that demographic effects should be minimal in these two datasets. To further explore this finding, we used a series of simulations using SLiM (V3.2.1) (Haller and Messer 2019) to explore how different cases of non-equilibrium demography and population structure can affect LD among neutral and deleterious mutations (see Materials and Methods for more details). We first tested how a model of admixture might impact patterns of LD for rare (allelic count of at least 5) neutral and deleterious mutations under strictly multiplicative selection when we introduced admixture between a focal population and two previously isolated satellite populations. We found that admixture easily caused positive LD to build up among neutral mutations, particularly so if admixture started recently between populations that had previously been isolated (Figure 4A). However, this was not the case for deleterious mutations in these populations where LD remained very close to 0 albeit slightly positive on average if admixture was present.
Figure 4. A) Mean LD among simulated neutral and deleterious mutations under different scenarios of admixture. The x-axis represents the generation in which admixture between isolated populations started. All simulations were run of a total of 1.5 million generations. B) Mean LD among neutral mutations segregating in simulated populations existing on a 2D geographic landscape. The x-axis represents different scenarios of mating bias by distance with increasingly more random mating to the right of the x-axis.

To ask whether LD generated in our simulations decays with distance, we plotted r versus distance between mutations pairs in a randomly chosen simulation replicate (sampled from a population where admixture started at generation 1,300,000) using PLINK (Purcell et al. 2007). We then repeated this analysis with synonymous mutations in our two datasets and two additional ones; 152 Arabidopsis thaliana individuals from Spain (1001 Genomes Consortium), and 162 Amaranthus tuberculatus individuals from North America (Kreiner et al. 2019). The observed spline of r vs. distance in one of our simulation replicates qualitatively resembled the real data with r being substantially positive for nearby mutations (Figure 5). Our simulations appeared to match the pattern observed in the highly selfing A. thaliana closest where signed LD was strongly elevated between nearby mutations.
Figure 5. Spline of signed LD (r) by distance among synonymous mutations in four different species. Additionally, spline of signed LD (r) by distance between neutral mutations from a simulation with 200K generations of migration between previously isolated populations.

Because positive LD among synonymous mutations was present in such a wide variety of datasets, we explored isolation by distance due to continuous geography as a potentially more generalizable mechanism that could create positive LD in a similar way to admixture. Using SLiM to model populations on a continuous 2D landscape we again readily observed positive LD forming among neutral mutations under several scenarios of distance biased mate choice, demonstrating that spatial considerations alone might be able to explain patterns of LD among synonymous mutations (Figure 4B).

Discussion

In this study we analyzed patterns of LD in two population genomic datasets to ask whether we could find evidence of negative synergistic epistasis among deleterious mutations as reported by Sohail et al. (2017). We repeated the analysis performed by Sohail et al. in a C. grandiflora dataset but found no evidence that LOF mutations are on average in negative LD with each other as would be expected if negative synergistic
epistasis was prevalent. We also revisited one of the *D. melanogaster* datasets used by Sohail et al. finding results qualitatively similar to those previously reported; however, we ultimately conclude that the data do not show clear evidence for the prevalence of significantly negative LD among deleterious mutations, a possible signature of synergistic epistasis.

Our analysis uncovered an average LD value of 4.57 x 10E-7 and -2.68 x 10E-7 per pair of LOF mutations in *C. grandiflora* and *D. melanogaster*, respectively. To test whether the LD among LOF mutations was significant, we followed the approach of Sohail et al., using synonymous mutations as a control group for LD comparisons. When using this method, we recovered similar results in both *D. melanogaster* and *C. grandiflora*; LOF mutations had LD values significantly lower than subsampled, allele frequency matched synonymous mutations. However, we ultimately caution against using this approach as a formal test for the presence of negative LD among deleterious mutations because LD among synonymous mutations appears to be excessively positive in both datasets.

Processes such as admixture can easily explain this pattern; positive LD due to admixture is expected between loci that have different allele frequencies in their populations of origin (Chakraborty and Weiss 1988).

Sohail et al. had recognized the potential for migration to cause positive LD. Using simulations employing human historical demography, they demonstrated that under this scenario of non-equilibrium demography, positive LD builds up between mutations in a manner dependent on their selection coefficient; the more deleterious the mutations, the less positive LD builds up among them (Sohail et al. 2017), even under a multiplicative model of negative selection. Using our own set of simulations, we have shown that this effect applies to other cases such as when admixture exists between a focal randomly mating population and unobserved satellite populations under multiplicative selection.

To test an even more subtle form of spatial structure, we used simulations across a continuous 2D landscape to show that positive LD will build up among neutral mutations if mating is biased towards nearby neighbours within a single continuous population. Presumably, the reason that positive LD occurs for low frequency neutral but not selected SNPs is as follows. Low frequency neutral SNPs within a given region will tend
to be of two types: local variants of relatively recent origin but also migrant variants (of older origin), which will have come to the local population linked to migrant variants at other genomic sites (i.e., in positive LD). Deleterious variants are less likely to be of older (migrant) origin by virtue of the selection against them.

Population genomic datasets such as the *D. melanogaster* DPGP3 and the *C. grandiflora* dataset we used here seemed like ideal, outbred, high-diversity populations for analysis of patterns of LD. However, they likely include LD buildup due to mating among physically proximal individuals. To further illustrate this point, we visualised r vs distance between synonymous mutations in four datasets (*C. grandiflora, D. melanogaster, A. thaliana, A. tuberculatus*). All datasets showed the same pattern of LD whereby r was positive for nearby mutations but decayed with distance, similar to the patterns we observed for LD among LOF mutations but at a slower rate. LD was most positive in *A. thaliana*, which likely relates to its highly selfing mode of reproduction that greatly reduces the efficacy of recombination in breaking down LD. These results add to the list of datasets where such patterns of signed LD decay have been observed such as humans and gonorrhea (Sohail et al. 2017; Arnold et al. 2020). In conclusion, observations of positive LD among synonymous mutations are widespread among taxa suggesting that gene flow coupled with population structure is the most parsimonious explanation.

Both our own simulations and those by Sohail et al show that such gene flow effects on LD are much stronger among neutral than selected mutations, even in the absence of epistasis. As a consequence, it is problematic to compare LD between synonymous and selected variants if the goal is to test for epistasis. Thus, those components of the case for synergistic epistasis reported by Sohail et al. that are based on such contrasts should be regarded with caution. More generally, the widespread observations of positive LD among synonymous sites further bolsters a recent call for the need for explicit modelling of spatial geography in population genetic analyses (Battery et al. 2020).

Because a comparison to synonymous mutations is potentially unsuitable for testing for negative synergistic epistasis, we used a permutation approach to test whether the LD
among LOF mutations in our datasets was significantly different from zero. Using this approach, we did not find that LD among LOF mutations was significantly different from 0 in either dataset. It could be argued that this approach is underpowered since LD is rapidly eroded by recombination in outcrossing populations such as the ones studied here. To address this concern, we visualised the entire distribution of LD among all LOF mutations in both datasets, split across bins of inter-mutation distance. If synergistic epistasis is present, a signature of negative LD should be most readily observable between mutations with limited recombination, such as those that are physically proximal to each other. We found the opposite pattern in both datasets; nearby pairs of LOF mutations had significantly positive LD, while LD decayed to zero at longer distances. This runs counter to the expectation of negative LD that either synergistic negative epistasis or Hill-Robertson interference predict, suggesting that demographic factors are likely the most important force shaping LD among LOF mutations. Of course, our negative result does not necessarily mean that synergistic epistasis is absent. Perhaps, we are simply unable to detect evidence of its signature of negative LD, because a different force, gene flow, overwhelms it. Nonetheless, our main conclusion is that we do not find evidence of the negative LD expected under synergistic epistasis (without demographic effects). For further discussion of why our study and that of Sohail et al. reach such different conclusions despite reporting similar results please see Supplementary Text.

Despite the lack of evidence of genome-wide negative synergistic epistasis, epistatic interactions might still be common between specific groups of deleterious mutations. In particular, given that genes function as part of larger biological networks, negative synergistic epistasis may arise between deleterious mutations that affect the function of genes within the same networks (Chiu et al. 2012). To test this idea, we calculated mean LD among synonymous and among radical missense mutations present in genes within interacting biological networks defined by KEGG (Kanehisa et al. 2016). Permutation tests in D. melanogaster suggested that the observed intra-network LD among radical mutations was more negative than expected by chance, but this was not the case in C. grandiflora. The difference between species could be biologically meaningful or more mundane. For example, KEGG network delineation could be more
biologically meaningful in *Drosophila* compared to *Capsella* where network information has been obtained from a different, closely related species (*A. thaliana*). Alternatively, the difference between species could be simply be a statistical artifact (i.e., false positive in *Drosophila* or false negative in *Capsella*). Similar analyses in other species will shed light on what the particular characteristics of epistasis within networks are.

Our result that synonymous, rather than LOF mutations have LD significantly different from 0 poses an interesting question when it comes to efforts to detect synergistic epistasis purely from metrics of LD: Is detecting epistasis from patterns of LD feasible in real world population genetics datasets of sexually reproducing organisms? We are uncertain that this is the case for two main reasons. One, detecting synergistic epistasis requires focusing on patterns of LD between deleterious mutations. Such mutations are expected to segregate at low frequencies in a population. If synergistic epistasis is present, the absolute amount of negative LD that is expected to build up is low due to the low frequency of such mutations. Next, the small amount of LD that will build up due to epistasis, is quickly broken apart by recombination every generation. As such, only LD built up by very recent selection, or in areas of low recombination, will be observable in a population genetic dataset. Together, we suspect these processes greatly reduce the power to detect synergistic epistasis through LD in most population genomics datasets.

In conclusion, we report no convincing evidence of genome-wide negative LD among deleterious mutations in either *C. grandiflora* or *D. melanogaster*. There is some evidence of negative LD within networks in *Drosophila*, but the result is not particularly strong. Rather, positive LD between nearby mutations seems to be a genome-wide phenomenon in both datasets analyzed. We suggest that these patterns are indicative of residual effects of demography and population structure, despite our presumption of the minimal presence of their effects in the datasets analyzed. Finally, we caution other researchers from using synonymous mutations as a control when analysing patterns of LD.
Materials and Methods

Population genomics datasets

We retrieved SNP calls from whole genome sequencing of 182 C. grandiflora individuals from (Josephs et al. 2015) and SNP calls from 197 haploid D. melanogaster embryos from the Drosophila population genomics project (DPGP3)(Lack et al. 2015). Both data sets are a result of thorough sampling from single populations with low population structure, making them ideal candidates for detecting signs of epistasis from patterns of LD. To ensure that recent migrants did not affect our LD analyses we used the R package SNPrelate (Li 2011) to visualize relatedness through PCA between C. grandiflora samples. This revealed six divergent genotypes that we eliminated from our downstream analysis leaving us with a total of 176 individuals. As Sohail et al. had already used the DPGP3 dataset to analyze patterns of LD we used the 190 individuals they retained after their filtering in our own analyses. We then further filtered both datasets by only considering bi-allelic sites where all individuals had genotype information.

SNP annotation

We used SNPeff (Cingolani et al. 2012) and the genome annotations of the reference genomes (Slotte et al., 2013 for Capsella rubella; D. melanogaster release 5.57 from Thurmond et al., 2019) to functionally annotate SNPs in both datasets as either LOF, synonymous or missense (non-synonymous). We included stop-gain and splice-disrupting SNPs in our set of LOF mutations based on the method of Sohail et al. We also classified missense SNPs as either radical or conservative. Missense SNPs were considered radical if they changed both the volume and polarity of an amino acid based on the classification of amino acids proposed by previous work (Sainudiin et al. 2005; Weber and Whelan 2019). In total we uncovered 3501 LOF SNPs in C. grandiflora and 2553 LOF SNPs in D. melanogaster at an allelic count cut-off of less than 6. This included 1219 of the 1652 mutations identified by Sohail et al. as candidate LOF mutations in their analysis of the D. melanogaster dataset. The difference in the number
of *D. melanogaster* LOF SNPs between our study, and the results reported by Sohail et al. likely stems from our use of a different SNP annotation tool (SNPeff vs. VEP).

### Calculating LD

We calculated LD values in two ways. First, we used the same method as Sohail et al. by calculating a point estimate of average LD among all mutations. For a genome with *K* loci, let *X*<sub>*i*</sub> be a discrete, random variable representing the number of derived alleles present at locus *i*, which can take values 0, 1 for a haploid population or alternatively 0, 1, 2 for a diploid population. The variance in the total number of derived mutations carried by each individual in the population can be expressed as:

\[
\text{Var} \left( \sum_{i=1}^{K} X_i \right) = \sum_{i=1}^{K} \text{Var}(X_i) + 2 \sum_{i<j}^{K} \text{Cov}(X_j, X_i)
\]

Because LD is, by definition, a covariance in the allelic state between two loci, we can use this equation to estimate the sum of all covariances across all loci by subtracting the first term of the right-hand side from the term on the left-hand side (and then dividing by 2). The term on the left-hand side represents the genome-wide variance in mutation burden; the first term on the right-hand side is the sum of the variance in mutation burden at each locus. We can then estimate a mean value of LD per pair of loci by dividing by the number of possible two-way interactions in the dataset

\[
\text{mean } LD = \frac{\left( \text{Var} \left( \sum_{i=1}^{K} X_i \right) - \sum_{i=1}^{K} \text{Var}(X_i) \right)}{2 \binom{K}{2}}
\]

We also modified this approach to calculate LD on a block by block basis instead of SNP by SNP. This measure of average LD largely eliminates LD between physically close sites, which could initially arise via random mutation. We first split the genome into 100Kb non-overlapping blocks. For a given genotype, we define *B*<sub>*g*</sub> as the number of derived variants in block *g*. This new variable can take values from 0 to 2*(number of segregating derived alleles in the given genomic block). To calculate total LD among all blocks, we infer the covariance in mutation burden between all blocks as follows
where $W$ refers to the total number of 100Kb blocks in the genome. Consider for example the simple case where we compare two blocks $(B_g, B_h)$, each with two segregating sites, $B$ can be represented as $B = X_1 + X_2, B = X_3 + X_4$.

The number of covariance terms for these two genomic blocks is $Cov(B_g, B_h) = Cov(X_1, X_3) + Cov(X_1, X_4) + Cov(X_2, X_3) + Cov(X_2, X_4)$.

The within-block LD (e.g., $Cov(X_1, X_2)$ and $Cov(X_3, X_4)$) from physically neighbouring sites contributes to the block-level variances (e.g., $Var(B_g)$ and $Var(B_h)$) but not the between-block covariances. For an arbitrary number of blocks, $Cov(B_g, B_h)$ can therefore be standardized per pair of interacting blocks as follows:

$$\text{mean LD}_{\text{blocks}} = \frac{\left( \text{Var}\left(\sum_{g=1}^{W} B_g\right) - \sum_{g=1}^{W} \text{Var}(B_g) \right)}{2 \left( \sum_{g \neq h}^{W} \sum_{n_g}^{W} n_g n_h \right)}$$

where $n_g$ and $n_h$ represent the number of sites with segregating derived variants in block $g$ and $h$ respectively.

We calculated mean LD using the above formula by transforming genotypes in our VCF files into tables of non-reference allele counts (0, 1, 2 for C. grandiflora and 0, 1 for D. melanogaster) and calculating the relevant statistics in R using the package matrixStats (Bengtsson 2017). We assumed that the non-reference alleles were generally the derived alleles in the two datasets. In principle, a reference genome assembled from a randomly sampled haplotype will contain some derived alleles that we will incorrectly assume are ancestral in our method. This issue however should be minimal since our analyses exclusively focus on rare mutations (<5% frequency) that are unlikely to be included in a reference assembly and will be filtered out as high frequency variants by
our analysis even if they are included. This is especially true for most putatively deleterious mutations such as LOF mutations which are likely maintained at low frequency by mutation-selection balance.

We also calculated LD using PLINK (Purcell et al. 2007) for LOF mutations as there were few enough mutations in these datasets to allow us to calculate a value of LD for each possible pair of mutations. We did this by first estimating $r$ between every single pair of mutations in our dataset in PLINK and then back-calculating a raw value of LD by multiplying $r$ by the square root of the product of allele frequencies at the two loci being compared. This approach allows us to observe the entire distribution of LD values rather than one summary statistic and to assess the effect of other factors that could affect patterns of LD such as the distance between two loci being compared. Back-calculating a raw value of LD from $r$ allows us to compare values from our two methods directly.

Gene network analysis

We used the R package graphite (Sales et al. 2012) to obtain lists of genes from biological pathways described in the KEGG database (Kanehisa et al. 2016). Network information from KEGG was directly available for D. melanogaster but not for C. grandiflora where we instead used network information from Arabidopsis thaliana. We used information on C. grandiflora - A. thaliana orthologs from (Josephs et al. 2015) to generate lists of interacting genes in C. grandiflora. Due to the low number of LOF mutations in each dataset we used low frequency (count of less than 5) radical missense mutations (definition described in SNP annotation section) as our set of candidate deleterious mutations. We then calculated a value of mean LD for each network defined by KEGG for our two species, generating separating sets of networks for synonymous and radical mutations.
Simulations

We used SLiM (V3.2.1) (Haller and Messer 2019) to run forward time simulations of population admixture to ask how signed LD can be affected by various demographic processes. We simulated three populations of 100,000 individuals each: one focal population that was sampled at the end of the simulation and two satellite populations with symmetrical migration to the focal population (10,000 individuals per generation).

Each diploid individual in our simulation contained two 1Mb chromosomes with recombination and mutation rates both 1E-08 per bp per generation. Mutations were sampled from two categories: neutral ($s = 0$) with a probability 1%, or deleterious ($s = -0.001$) with a probability of 99%. Fitness was determined by the multiplicative effect of deleterious load in each individual genome, dominance was also assumed to be additive. We ran all simulations for 1.5 million generations altering the generation where continuous admixture was started in several treatment groups: no admixture, admixture starting at generation 200,000, 600,000, and 1,300,000. Each treatment group was made of 20 simulated replicates. After 1.5 million generations, we sampled 100 individuals from the focal population in each replicate. Next, we filtered out recent migrants in our focal population by performing a PCA on genotype of our samples and eliminating individuals with PC values greater than 1 SD away from the mean of PC1 or PC2. This mimics how we treated our real-world data where we eliminated outlier samples using PCA. Next, to replicate how we defined ancestral/derived alleles in the real-world data, we assigned all mutations with frequencies over 50% in our samples as the ancestral variant. Finally, we filtered out sites with an allele count over 5 and calculated LD separately for neutral and deleterious mutations in each replicate.

We ran a second set of simulations consisting of only one focal population where individuals were placed on a 2D landscape to simulate the effects of isolation by distance due to limited dispersal. We used the “Mate choice with a spatial kernel” recipe provided in the SLiM manual for this set of simulations. Briefly, 10,000 individuals were randomly placed on an ($x,y$) plane, with coordinate ranges [0,1] for both axes. To avoid clumping, individual fitness was calculated as a function of spatial competition with neighbouring individuals exerting the most costs to each other (see SLiM manual for
more details URL: http://benhaller.com/slim/SLiM_Manual.pdf). Individuals chose mates a gaussian-distributed distance away, with mean 0, SD $\sigma$, and maximum value $\tau$. We ran simulations with three sets of parameter values for $\sigma$ and $\tau$: (0.1,0.02), (0.3,0.06), (0.5,0.5). This range of values was selected to explore various levels of bias towards localized mating much like might occur in plant populations with limited pollen dispersal. Finally, offspring dispersed a gaussian distance away from their first parent. Each individual contained two 1Mb chromosomes containing only neutral mutations with a recombination and mutation rate of 1E-08 per bp per generation. The simulations were terminating after 100,000 generations and 100 individuals were sampled per simulation replicate. Each mate choice condition was replicated 10 times. After sampling, LD was calculated as described for the other simulations with the exception of PCA analysis as no migrant filtering was necessary due to the absence of cross-population migration.

Acknowledgements

The authors would like to thank Tyler Kent for helpful discussions and providing guidance on using C. grandiflora population genomic data. This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grants (S.I.W., and A.F.A.) and an NSERC Alexander Graham Bell Canada Graduate Scholarship (G.S.). Organismal cartoons included in figures were licensed and created with BioRender.com.

Author Contributions

All authors designed the research. G.S. performed all analyses and wrote the draft manuscript. All authors revised the manuscript.
Literature Cited

1001 Genomes Consortium. 2016. 1,135 Genomes Reveal the Global Pattern of Polymorphism in *Arabidopsis thaliana*. Cell 166:481–491.

Agrawal AF, Whitlock MC. 2010. Environmental duress and epistasis: how does stress affect the strength of selection on new mutations? Trends Ecol. Evol. 25:450–458.

Arnold B, Sohail M, Wadsworth C, Corander J, Hanage WP, Sunyaev S, Grad YH. 2020. Fine-Scale Haplotype Structure Reveals Strong Signatures of Positive Selection in a Recombining Bacterial Pathogen. Mol. Biol. Evol. 37:417–428.

Auton A, McVean G. 2007. Recombination rate estimation in the presence of hotspots. Genome Res. 17:1219–1227.

Bank C, Hietpas RT, Jensen JD, Bolon DNA. 2015. A Systematic Survey of an Intragenic Epistatic Landscape. Mol. Biol. Evol. 32:229–238.

Barton NH. 1995. A general model for the evolution of recombination. Genet. Res. 65:123–145.

Battery C, Ralph PL, Kern AD. 2020. Space is the Place: Effects of Continuous Spatial Structure on Analysis of Population Genetic Data. Genetics 215:193–214.

Chakraborty R, Weiss KM. 1988. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc. Natl. Acad. Sci. 85:9119–9123.

Chiu H-C, Marx CJ, Segrè D. 2012. Epistasis from functional dependence of fitness on underlying traits. Proc. R. Soc. B Biol. Sci. 279:4156–4164.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6:80–92.

Crow JF, Kimura M. 1970. An introduction to population genetics theory. New York: Harper & Row.

Crow JF, Kimura M. 1979. Efficiency of truncation selection. Proc. Natl. Acad. Sci. 76:396–399.

Elena SF, Lenski RE. 1997. Test of synergistic interactions among deleterious mutations in bacteria. Nature 390:395–398.

Garcia JA, Lohmueller KE, unpublished data. Negative linkage disequilibrium between amino acid changing variants reveals interference among deleterious mutations.
in the human genome.

https://www.biorxiv.org/content/10.1101/2020.01.15.907097v1.full

Haller BC, Messer PW. 2019. SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. Mol. Biol. Evol. 36:632–637.

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8:269–294.

Hill WG, Robertson A. 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38:226–231.

Josephs EB, Lee YW, Stinchcombe JR, Wright SI. 2015. Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. Proc. Natl. Acad. Sci. U. S. A. 112:15390–15395.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44:D457–D462.

Kimura M, Maruyama T. 1966. The Mutational Load with Epistatic Gene Interactions in Fitness. Genetics 54:1337–1351.

Kondrashov AS. 1982. Selection against harmful mutations in large sexual and asexual populations. Genet. Res. 40:325–332.

Kondrashov AS. 1995. Dynamics of unconditionally deleterious mutations: Gaussian approximation and soft selection. Genet. Res. 65:113–121.

Kreiner JM, Giacomini DA, Bemm F, Waithaka B, Regalado J, Lanz C, Hildebrandt J, Sikkema PH, Tranel PJ, Weigel D, et al. 2019. Multiple modes of convergent adaptation in the spread of glyphosate-resistant Amaranthus tuberculatus. Proc. Natl. Acad. Sci. 116:21076–21084.

Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, Langley CH, Pool JE. 2015. The Drosophila Genome Nexus: A Population Genomic Resource of 623 Drosophila melanogaster Genomes, Including 197 from a Single Ancestral Range Population. Genetics 199:1229–1241.

Lalić J, Elena SF. 2012. Magnitude and sign epistasis among deleterious mutations in a positive-sense plant RNA virus. Heredity 109:71–77.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinforma. Oxf. Engl. 27:2987–2993.
McEvoy BP, Powell JE, Goddard ME, Visscher PM. 2011. Human population dispersal “Out of Africa” estimated from linkage disequilibrium and allele frequencies of SNPs. Genome Res. 21:821–829.

McVean G. 2007. The Structure of Linkage Disequilibrium Around a Selective Sweep. Genetics 175:1395–1406.

Pfaff CL, Parra EJ, Bonilla C, Hiester K, McKeigue PM, Kamboh MI, Hutchinson RG, Ferrell RE, Boerwinkle E, Shriver MD. 2001. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. Am. J. Hum. Genet. 68:198–207.

Puchta O, Cseke B, Czaja H, Tollervey D, Sanguinetti G, Kudla G. 2016. Network of epistatic interactions within a yeast snoRNA. Science 352:840–844.

Rice WR. 1998. Requisite mutational load, pathway epistasis, and deterministic mutation accumulation in sexual versus asexual populations. Genetica 102:71.

Sainudiin R, Wong WSW, Yogeeswaran K, Nasrallah JB, Yang Z, Nielsen R. 2005. Detecting Site-Specific Physicochemical Selective Pressures: Applications to the Class I HLA of the Human Major Histocompatibility Complex and the SRK of the Plant Sporophytic Self-Incompatibility System. J. Mol. Evol. 60:315–326.

Sales G, Calura E, Cavalieri D, Romualdi C. 2012. graphite - a Bioconductor package to convert pathway topology to gene network. BMC Bioinformatics 13:20.

Slotte T, Hazzouri KM, Ågren JA, Koenig D, Maumus F, Guo Y-L, Steige K, Platts AE, Escobar JS, Newman LK, et al. 2013. The Capsella rubella genome and the genomic consequences of rapid mating system evolution. Nat. Genet. 45:831–835.

Sohail M, Vakhrusheva OA, Sul JH, Pulit SL, Francioli LC, Consortium G of the N, Initiative ADN, Berg LH van den, Veldink JH, Bakker PIW de, et al. 2017. Negative selection in humans and fruit flies involves synergistic epistasis. Science 356:539–542.

Stephens JC, Briscoe D, O’Brien SJ. 1994. Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. Am. J. Hum. Genet. 55:809–824.

Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, Matthews BB, Millburn G, Antonazzo G, Trovisco V, et al. 2019. FlyBase 2.0: the next generation. Nucleic Acids Res. 47:D759–D765.
Weber CC, Whelan S. 2019. Physicochemical Amino Acid Properties Better Describe Substitution Rates in Large Populations. Mol. Biol. Evol. 36:679–690.