Mutation Load in Sunflower Inversions Is Negatively Correlated with Inversion Heterozygosity

Kaichi Huang [ID],*,[1] Kate L. Ostevik [ID],[1,2] Cassandra Elphinstone,[1] Marco Todesco [ID],[1] Natalia Bercovich [ID],[1] Gregory L. Owens [ID],[1,3] and Loren H. Rieseberg [ID]*,[1]

[1]Department of Botany and Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada
[2]Department of Evolution, Ecology, and Organismal Biology, University of California, Riverside, CA
[3]Department of Biology, University of Victoria, Victoria, British Columbia, Canada

*Corresponding authors: E-mails: kaichi.huang@botany.ubc.ca; lriesebe@mail.ubc.ca.
Associate editor: Dr. Aya Takahashi

Abstract

Recombination is critical both for accelerating adaptation and purging deleterious mutations. Chromosomal inversions can act as recombination modifiers that suppress local recombination in heterozygotes and thus, under some conditions, are predicted to accumulate such mutations. In this study, we investigated patterns of recombination, transposable element abundance, and coding sequence evolution across the genomes of 1,445 individuals from three sunflower species, as well as within nine inversions segregating within species. We also analyzed the effects of inversion genotypes on 87 phenotypic traits to test for overdominance. We found significant negative correlations of long terminal repeat retrotransposon abundance and deleterious mutations with recombination rates across the genome in all three species. However, we failed to detect an increase in these features in the inversions, except for a modest increase in the proportion of stop codon mutations in several very large or rare inversions. Consistent with this finding, there was little evidence of overdominance of inversions in phenotypes that may relate to fitness. On the other hand, significantly greater load was observed for inversions in populations polymorphic for a given inversion compared to populations monomorphic for one of the arrangements, suggesting that the local state of inversion polymorphism affects deleterious load. These seemingly contradictory results can be explained by the low frequency of inversion heterozygotes in wild sunflower populations, apparently due to divergent selection and associated geographic structure. Inversions contributing to local adaptation represent ideal recombination modifiers, acting to facilitate adaptive divergence with gene flow, while largely escaping the accumulation of deleterious mutations.

Key words: inversion, recombination rate, deleterious mutation, transposable element, *Helianthus*, centromeres.

Introduction

Chromosomal inversions are thought to play an important role in adaptation and speciation. In one class of models, inversions contribute to adaptive evolution by reducing gene flow over large genomic regions, while maintaining advantageous epistatic interactions and/or favorable genotypic combinations at loci affecting adaptation to different environments (Charlesworth and Charlesworth 1973; Kirkpatrick and Barton 2006; Feder et al. 2011; Charlesworth and Barton 2018). Other models suggest that inversions can facilitate speciation by increasing effective linkage between genes that contribute to local adaptation and those causing assortative mating (Trickett and Butlin 1994; Ortiz-Barrientos et al. 2016) or by enabling the accumulation of intrinsic genetic incompatibilities (Noor et al. 2001; Navarro and Barton 2003).

These recombination suppression models offer a means for resolving the widely recognized antagonism between divergent natural selection and recombination (Felsenstein 1974, 1981). However, genetic recombination between

homologous chromosomes is essential for removing deleterious mutations from the genome (Carvalho 2003; Keightley and Otto 2006). In genomic regions of low recombination, the efficiency of selection in eliminating slightly deleterious mutations is expected to weaken because deleterious mutations might hitchhike with genes that are under positive selection (i.e., Hill-Robertson effects; Hill and Robertson 1966; Morrell et al. 2012). Population genomic analyses of *Drosophila* and other animal species have provided evidence for this evolutionary advantage of genetic recombination (Bachtrog 2003; Kaiser and Charlesworth 2009; Wang et al. 2013; Tuttle et al. 2016; Yoshida et al. 2020). Likewise, studies of a number of crops have reported a negative correlation between mutational load and recombination rate (Lu et al. 2006; Renaut and Rieseberg 2015; Lozano et al. 2021), consistent with a critical role for recombination in plants as well.

In inversions, effective recombination between the arrangements is reduced through selection against recombinant gametes in heterokaryotypes or via crossover interference mechanisms (Fuller et al. 2019; Huang and

Open Access

Rieseberg 2020). Such suppression of recombination promotes independent evolution of each arrangement and may result in a decrease in effective population size within each arrangement and, thus, lower the efficacy of purifying selection in removing deleterious mutations in inversions. Theoretical work indicates that both arrangements of inversions accumulate deleterious load at higher rates compared to regions without inversions (Navarro et al. 2000; Faria et al. 2019; Berdan et al. 2021), while this can be mitigated by gene conversion or, rarely, double crossovers (Berdan et al. 2021). Recessive deleterious mutations restricted to only one inversion arrangement may result in higher survival rate in heterokaryotypes than in homokaryotypes. Such associative overdominance favors intermediate-to-large inversions that persist as balanced polymorphisms (Ohta 1971; Faria et al. 2019; Connallon and Olito 2021; Berdan et al. 2021), and has been showed to be an important force for maintaining polymorphism in a number of inversions in Insecta, such as *Drosophila pseudoobscura* (Crumpacker and Salceda 1969; Yang et al. 2002), seaweed flies *Coelopa frigida* (Butlin and Day 1985), and *Heliconius* butterflies (Jay et al. 2021).

Inversion polymorphism can also establish through other mechanisms. An inversion arrangement might spread to fixation by selection in some local populations while being absent or lost in others. Inversions in this case should have a lower heterozygote frequency than those established via associative overdominance and be less prone to the accumulation of deleterious alleles due to the large local effective population size of each arrangement and increased occurrence of recombination (Faria et al. 2019). However, few theoretical or empirical studies discuss deleterious mutations in inversions that are involved in local adaption and/or fixed locally. Emerging data have identified a great number of inversions that may establish in the scenario of local adaptation (Lowry and Willis 2010; Lee et al. 2017; Coughlan and Willis 2019; Huang et al. 2020). However, we are unaware of studies of patterns of deleterious mutations in these inversions.

While deleterious load is typically detected by the study of protein coding genes in population genomic studies (Renaut and Rieseberg 2015; Hahn 2019; Chen et al. 2020), transposable elements (TEs), which represent another potentially abundant source of deleterious mutations, are also predicted to be more abundant in chromosome regions with reduced levels of recombination, both because selection acting against detrimental TE insertion in functional sequences is weaker in regions of low recombination and because TEs are less likely to be involved in unequal recombination between elements located on homologous chromosomes (Langley et al. 1988; Ma and Bennetzen 2006; Kent et al. 2017). In *Drosophila*, TE abundance is strongly associated with the local recombination rate (Langley et al. 1988; Bartolome et al. 2002). In plants, recombination rates are significantly associated with long terminal repeat retrotransposon (LTR-RT) abundance in rice (Tian et al. 2009), soybean

(Tian et al. 2012), and wheat (Daron et al. 2014), as well as frequencies of different classes of TEs in *Eucalyptus* (Gion et al. 2016). Because chromosomal inversions can reduce recombination, it has also been proposed that TEs should accumulate in inverted regions (Montgomery et al. 1987; Sniegowski and Charlesworth 1994). Such an increase in TE copy number within inversions has been reported in *Drosophila* (Eanes et al. 1992; Sniegowski and Charlesworth 1994) as well as in *Heliconius* (Jay et al. 2021). However, we are unaware of such reports in plants.

Sunflowers are an excellent system to study the effects of inversions on the accumulation of deleterious mutations. First, large non-recombining haplotype blocks (haploblocks) have been found to play a critical role in ecotype formation in wild sunflowers, many of which were found to be linked to polymorphic chromosomal inversions (Todesco et al. 2020). The haploblocks displayed varying levels of homozygosity across populations, but most are associated with phenotypic or environmental variation and have geographic distributions suggestive of local adaptation (Todesco et al. 2020). This provides an opportunity to explore how inversion genotype frequency affects mutation accumulation in inversions. Second, an increase in nonsynonymous substitutions and alternative stop codons has previously been reported in regions of reduced recombination in cultivated sunflower (Renaut and Rieseberg 2015), suggesting that it should be possible to detect an effect of inversions on protein evolution in sunflower species, if it exists. Third, more than three quarters of the sunflower genome is comprised of TEs, of which 77% are LTR-RTs (Staton et al. 2012; Badouin et al. 2017). Sunflower LTR-RTs are predominantly of recent origin and exhibit transcriptional activity among wild species (Cavallini et al. 2010; Kawakami et al. 2011; Renaut et al. 2014). However, whether and how the accumulation pattern of TEs is shaped by inversions has not previously been explored in sunflowers.

In this study, we investigated the distribution of TEs, specifically LTR-RTs, and coding sequence evolution across the sunflower genome, as well as within nine inversions. We focused on the impact of recombination and inversions on TE density and the accumulation of deleterious mutations in protein coding genes, quantified as elevated nonsynonymous diversity and increased incidence of nonsense mutations. We further looked for a signal of elevated genetic load in inversions by testing for overdominance of phenotypic traits. In addition, we mapped the location of centromeres in the sunflower genome both to assess their impact on recombination rates and to determine if inversions included centromeres (pericentric) or not (paracentric). We showed that variation in recombination rate along the sunflower genome plays a crucial role in the accumulation of TEs and in modulating the efficacy of selection in removing deleterious mutations. However, we found that inversions, which displayed higher than expected homozygosity across populations, had little impact on deleterious load.

Results

SNP and Inversion Datasets

In this study, we used the whole-genome sequencing data of three annual sunflower species (*H. annuus*, *H. argophyllus*, *H. petiolaris*) from a previous study (Todesco et al. 2020) and generated whole-genome sequencing data for an additional 38 samples from two dune systems in *H. petiolaris*. We conducted variant calling on a new high-quality reference genome for *H. annuus* (Ha412HOv2.0), which used Hi-C (Marie-Nelly et al. 2014) for contig and scaffold ordering and was shown to have improved quality (Todesco et al. 2020). A total of 71 populations and 719 samples for *H. annuus*, 30 populations and 299 samples for *H. argophyllus*, and 43 populations and 427 samples for *H. petiolaris* were included in this study (see Materials and Methods). After calling variants and filtering, we obtained 15,452,562, 8,706,003, and 8,797,015 bi-allelic single nucleotide polymorphisms (SNPs) for *H. annuus*, *H. argophyllus*, and *H. petiolaris*, respectively. This is approximately double the number SNPs included in our previous datasets, despite application of similar filtering standards (Todesco et al. 2020).

We selected a subset of haploblocks from (Todesco et al. 2020) for which there was solid evidence of an underlying inversion (table 1). Most of these inversions have also been directly linked to ecological adaptation. For *H. annuus*, we selected inversions ann01.01 and ann05.01, the structure of which are well-characterized by comparative genomic analysis, plus inversions ann13.01 and ann15.01, which were associated with flowering time and climate continentality and are the largest inversions found in sunflowers (Todesco et al. 2020). For *H. argophyllus*, inversion arg10.01 was chosen because it was shown to be associated with the formation of the early flowering ecotype on the barrier islands of Texas (Moyers 2015; Todesco et al. 2020). For *H. petiolaris*, we selected four inversions (pet05.01, pet09.01, pet11.01, pet14.01), which were found to be associated with adaptive traits (e.g., seed size) in well-characterized dune ecotypes of the species (Huang et al. 2020; Todesco et al. 2020).

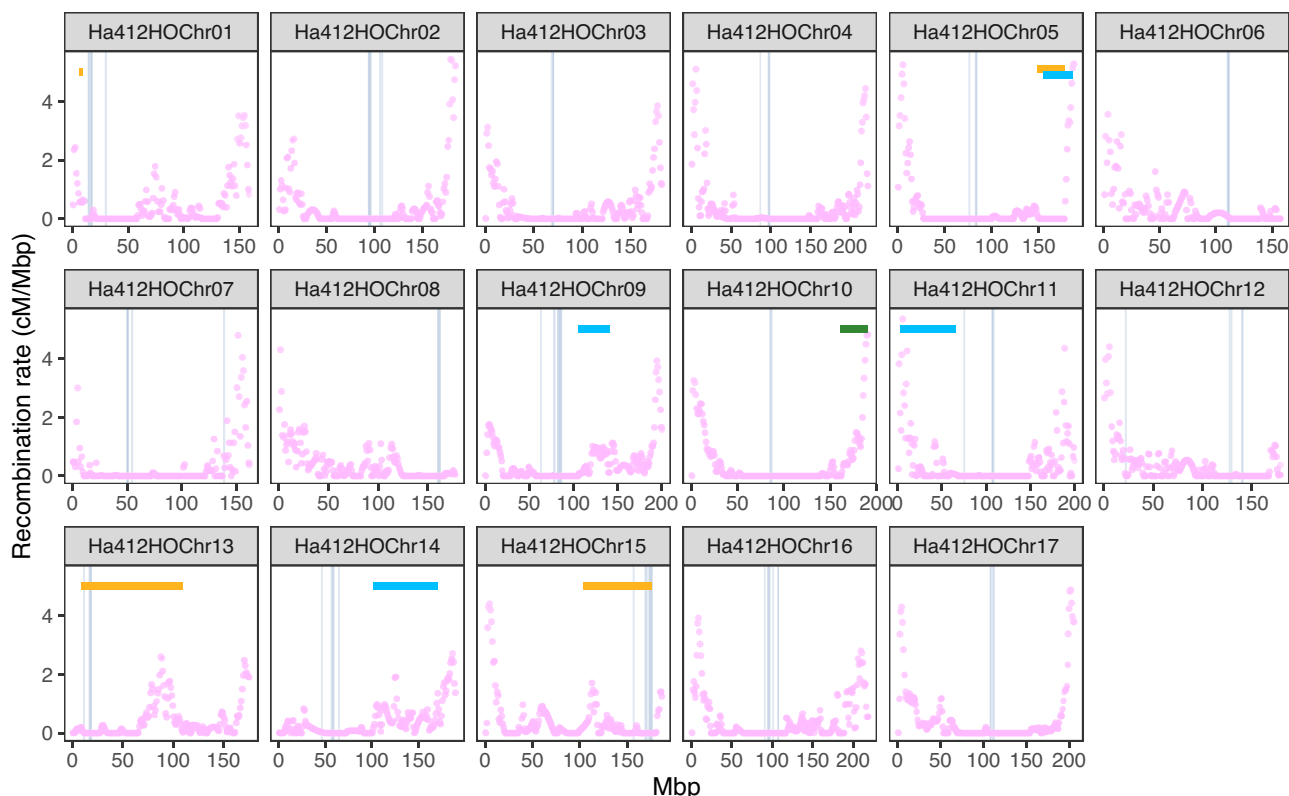## Genomic Recombination Rate and Centromeres

To explore the effect of recombination on molecular evolution across the genome, we made use an integrated genetic map for cultivated sunflower (Todesco et al. 2020) and remapped the markers to our current reference genome. This map integrated genetic maps from seven crosses to produce an average recombination rate for the species. Analysis of patterns of recombination rate variation across the genome of Ha412HOv2.0 using the integrated genetic map showed that most of the chromosomes have large regions of low recombination around central regions and high recombination toward the distal ends. Some chromosomes, such as Ha412HOChr06, have regions of low recombination near one terminal end (fig. 1). Since the recombination rate is relatively constant among annual sunflower species (Barb et al. 2014), we used the recombination map for *H. annuus* to estimate patterns of recombination in all three species in our study.

We mapped four sequences previously found to be targeted by sunflower centromere-specific histone H3 (LC075744-LC075747; Nagaki et al. 2015) to the reference genome to determine the position of centromeres. Among the four sequences, the LINE-like sequence LC075745 was mapped to a single centralized region on most chromosomes, except for chromosomes 7 and 12 where the sequence was also mapped to another small region (fig. 1; supplementary fig. S1, Supplementary Material online). The other three tandem repeat/retrotransposon sequences showed less well-defined patterns. Most of the hits of LC075744 were located on un-anchored scaffolds. Other hits were found on 7 of 17 chromosomes, the majority of which appeared in the vicinity of corresponding hits for LC075745. Sequence LC075746 was mapped to wider regions across the chromosomes but mostly appeared densely around the region identified by LC075745. Sequence LC075747 was found throughout the genome (supplementary fig. S1, Supplementary Material online). From this, we infer that the centromere locations were characterized by the regions delineated by LC075745. As expected, all the centromeres were located within a region of extremely low recombination, although the centromere of chromosome 1 was on the edge of the region; chromosomes 6, 8, and 13 are acrocentric, which fit the distribution of recombination rate on these chromosomes (fig. 1). Among the inversions, ann13.01 on chromosome 13 and ann15.01 on chromosome 15 are pericentric while the others are paracentric (fig. 1).

Table 1. Codes and Basic Features of the Inversions in the Analyses.

Code	Species	Chromosome	Size (Mbp)	Minor Allele Frequency
ann01.01	*Helianthus annuus*	Ha412HOChr01	3.5	0.220
ann05.01	*H. annuus*	Ha412HOChr05	29.3	0.446
ann13.01	*H. annuus*	Ha412HOChr13	100.8	0.474
ann15.01	*H. annuus*	Ha412HOChr15	72.9	0.389
arg10.01	*H. argophyllus*	Ha412HOChr10	29.5	0.192
pet05.01	*H. petiolaris*	Ha412HOChr05	31.2	0.206
pet09.01	*H. petiolaris*	Ha412HOChr09	35.3	0.241
pet11.01	*H. petiolaris*	Ha412HOChr11	62.2	0.372
pet14.01	*H. petiolaris*	Ha412HOChr14	69.4	0.077

Fig. 1. Recombination rate across the genome of Ha412HOv2.0. The recombination rate was calculated for 500 kbp windows based on an integrated genetic map of cultivated sunflower (*Helianthus annuus*). Positions of the centromere-specific sequence LC075745 are indicated by grey vertical bars. Positions of the inversions used in this study are indicated by horizontal bars (orange = *H. annuus*, green = *H. argophyllus*, light blue = *H. petiolaris*).

Inversions Suppress Recombination Between Arrangements

Using the newly generated SNP dataset, we examined patterns of linkage disequilibrium (LD) across our samples and sequence differentiation between arrangements to confirm the effect of recombination suppression for the inversions in sunflowers. When we compared LD between samples from populations that are polymorphic for the inversion and those from populations that contain only the homozygous genotype for the major arrangement, all inversions were characterized by a region of high LD among samples from polymorphic populations, while this pattern was not found among samples from populations with homozygous genotypes, except that a small region of high LD persisted in the centromeric regions in paracentric inversions ann13.01 and ann15.01 (supplementary fig. S2, Supplementary Material online). These results are consistent with the role of inversions in altering recombination in heterozygotes, while recombination in homozygotes remains unaffected.

In most of the inversions, differentiation between arrangements is consistently high across the whole region relative to the low F_{ST} outside the rearranged regions (supplementary fig. S3, Supplementary Material online), except for the abrupt gaps in the middle of pet05.10 and pet09.01. These result from the structural differences between *H. petiolaris* and the reference genome, which is

for cultivated sunflower (*H. annuus* var. *macrocarpus*; Ostevik et al. 2020). In the two largest and pericentric inversions, ann13.01 and ann15.01, F_{ST} is more heterogeneous with a gradual decline in the middle of the inverted region, probably due to increased double recombination events and gene conversion between arrangements away from the breakpoints as reported in previous studies (Navarro et al. 1997; Guerrero et al. 2012).

TE Abundance

To explore the effect of inversions on TE abundance, we made use of the three reference genomes of cultivated sunflowers (Ha412HOv2.0, XRQv2, PSC8) that have been found to contain different arrangements of two inversions found in the wild populations (Todesco et al. 2020): Ha412HOv2.0 and XRQv2 have the same arrangement for the inversion on chromosome 1 (ann01.01), while PSC8 has the alternate; XRQv2 and PSC8 possess the same arrangement for the inversion on chromosome 5 (ann05.01), while Ha412HOv2.0 has the alternate. To generate comparable datasets, we conducted de-novo TE annotation for the three reference genomes using the program EDTA, which is a sophisticated pipeline that creates raw TE libraries using various structure-based programs and filters out false discoveries in raw TE candidates to generate a high-quality non-redundant TE
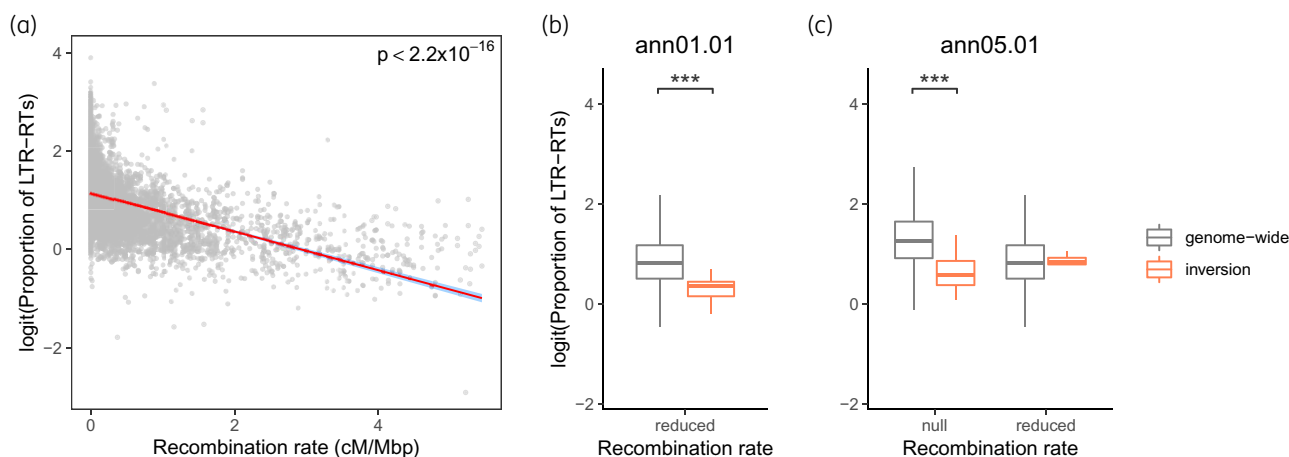
FIG. 2. Recombination rate and density of LTR-RTs. (*a*) Genome-wide correlation LTR-RT density and recombination rate in Ha412HOv2.0. The red line denotes the best-fit linear regression line with 95% confidence intervals shaded in blue. (*b*) Comparison of LTR-RT density within inversion ann01.01 to genome-wide. (*c*) Comparison of the LTR-RT density within inversion ann05.01 to genome-wide. The density of LTR-RTs was summarized by calculating the proportion of LTR-RT sequences in sliding windows of 500 kbp. Windows in each recombination rate category (reduced: 0.01–2 cM/Mbp, null: <0.01 cM/Mbp) were compared separately. Asterisks denote significance in independent *t*-test: ***P < 0.001.

library (Ou et al. 2019). Using the EDTA pipeline to annotate the three genome assemblies, we confirmed that the genomes of cultivated sunflower are composed of a high proportion of TEs (83.02% in Ha412HOv2.0, 83.33% in XRQv2, 83.21% in PSC8), with 71–72% of these genomes being LTR-RTs (71.50% in Ha412HOv2.0, 71.20% in XRQv2, 71.29% in PSC8). In agreement with previous studies of the cultivated sunflower genome (Staton et al. 2012), there is a major bias in TE composition towards *Gypsy* (Ha412HOv2.0 42.04%, XRQv2 41.93%, PSC8 41.63%) and *Copia* elements (Ha412HOv2.0 11.12%, XRQv2 12.04%, PSC8 12.52%). Class II TEs (DNA transposons) were much lower in abundance relative to LTR-RTs, comprising about 12% of each genome (Ha412HOv2.0 11.50%, XRQv2 12.14%, PSC8 11.92%).

The genomic distributions of LTR-RTs in the three assemblies are similar to those previously reported for the first reference genome for cultivated sunflower (Badouin et al. 2017; supplementary fig. S4, Supplementary Material online). We found a strong negative correlation between density of LTR-RTs and recombination rates ($P < 2.2 \times 10^{-16}$; fig. 2*a*). The results were similar when considering only *Gypsy* or *Copia* elements (not shown). Our integrated genetic map has not been aligned to the genomes of XRQv2 and PSC8, but the similar distributions of LTR-RTs across the three genomes imply a comparably strong negative correlation between recombination rate and TE abundance.

We divided 500 kbp windows of the genome into three different categories based on the recombination rate ("high": >2 cM/Mbp, "reduced": 0.01–2 cM/Mbp, "null": <0.01 cM/Mbp), and compared windows of each category in an inversion to those of the same category from across the genome, to control for the effect of genomic context on TE density (Bartolome et al. 2002). We note that although neighboring genomic windows are not completely

independent, linkage typically declines within about 10 kbp in sunflowers (Todesco et al. 2020). For the two inversions that are polymorphic between the reference genomes, the recombination rate of most windows in ann01.01 fell within the "reduced" category, whereas for ann05.01, both "reduced" and "null" windows were found. Recall that these recombination rates are derived from genetic mapping populations that are homozygous for the inversions. Thus, this is the background recombination rate for these windows rather than the recombination rate in the inversion heterozygotes. When compared to windows of a similar recombination rate across the genome, ann01.01 exhibited significantly lower density of LTR-RTs ($P = 9.17 \times 10^{-4}$; fig. 2*b*). Likewise, for ann05.01, windows with a "null" recombination rate contained fewer LTR-RTs than windows with a similar recombination rate elsewhere in the genome ($P = 6.62 \times 10^{-22}$), but no significant difference was found between windows of the "reduced" recombination rate category across the genome and those within the inversion ($P = 0.557$; fig. 2*c*). LTR-RT densities in forward and reverse inversion arrangements from the genome assemblies did not show visible differences and were all lower than the genomic average (supplementary fig. S5, Supplementary Material online).

Protein Evolution

We selected 20 samples from *H. annuus*, *H. argophyllus*, and *H. petiolaris*, respectively, and summarized zero-fold and four-fold diversity across the whole genome to estimate the genetic load of these samples. Across the three species, π_0/π_4 shows a negative correlation with levels of genetic diversity (mean of π_0 and π_4; supplementary fig. S6, Supplementary Material online). *Helianthus argophyllus* samples have the lowest levels of genetic diversity among the species. However, they show disproportionately higher

levels of diversity at zero-fold sites (supplementary fig. S6, Supplementary Material online). *Helianthus annuus* has similar but slightly higher levels of genetic diversity and lower levels of π_0/π_4 compared to *H. petiolaris*. To test the significance of this inverse relationship while controlling for spurious signal caused by mathematical correlation between π_0/π_4 and genetic diversity, both π_0 and π_4 were scaled by dividing by the standard deviation of each, and the association between the average of the scaled values and the ratio between them were then tested (Irwin, 2018). This resulted in a statistically significant inverse relationship (Pearson's $r = -0.7116$, $P = 1.848 \times 10^{-10}$).

In all three species, the ratio of nonsynonymous and synonymous mutations as estimated by window-based π_0/π_4 is similar to that seen in other taxa (Yang and Gaut 2011) with a small proportion of the values greater than 2 (supplementary fig. S7, Supplementary Material online). Mean π_0/π_4 ranges from 0.3–0.5, with the highest value found in *H. argophyllus* and the lowest in *H. petiolaris* (*H. annuus*: 0.3787; *H. argophyllus*: 0.4248; *H. petiolaris*: 0.3124). Across the genome, π_0/π_4 is negatively correlated with recombination rate in all three species, but the correlation is weak in *H. petiolaris* (*H. annuus*: Pearson's $r = -0.1162$, $P < 2.2 \times 10^{-16}$; *H. argophyllus*: Pearson's $r = -0.1752$,
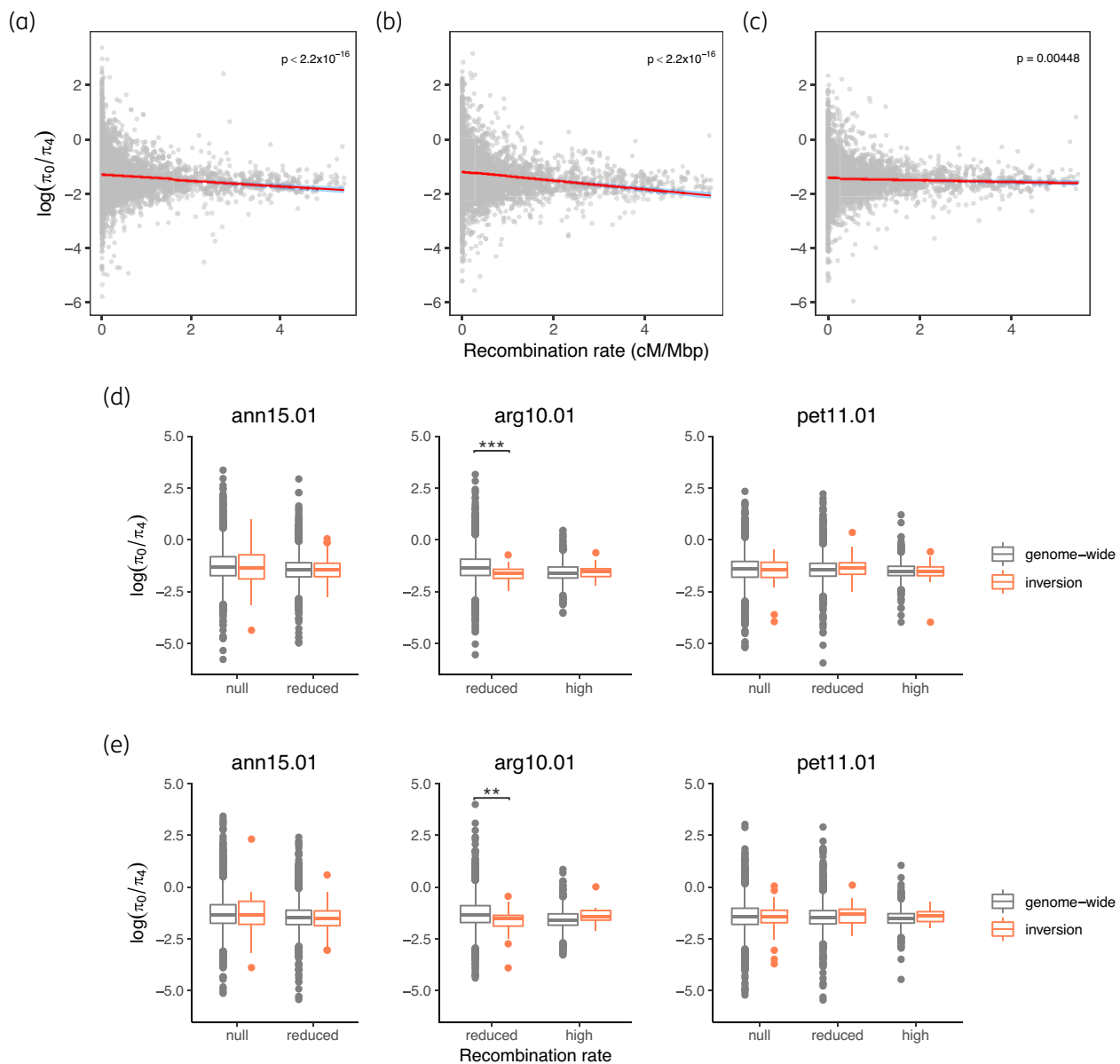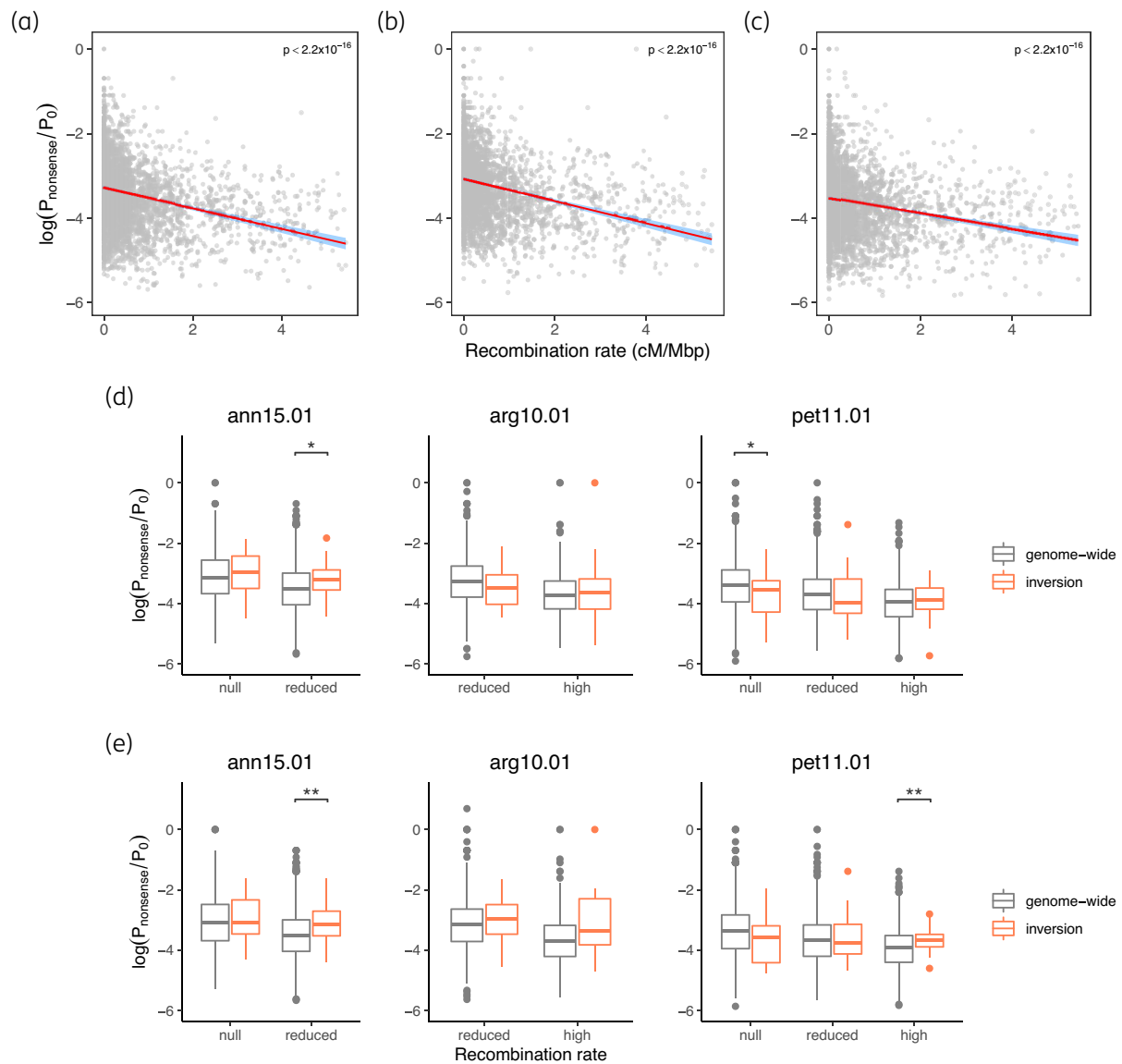

Fig. 3. Recombination rate and π_0/π_4. Genome-wide correlation of π_0/π_4 and recombination rate in (*a*) *Helianthus annuus*, (*b*) *H. argophyllus*, and (*c*) *H. petiolaris*. The red lines denote the best-fit linear regression line with the 95% confidence intervals shaded in blue. The values within inversions were compared to genome-wide regions in the same recombination rate category using (*d*) all samples or (*e*) samples homozygous for the rarer arrangement. Windows of each recombination rate category (high: >2 cM/Mbp, reduced: 0.01–2 cM/Mbp, null: <0.01 cM/Mbp) were compared separately. Asterisks denote significance in independent *t*-test: **0.01 > *P* > 0.001; ***P < 0.001. Results for all inversions are presented in supplementary figs. S8 and S9, Supplementary Material online.

$P < 2.2 \times 10^{-16}$; *H. petiolaris*: Pearson's $r = -0.03836$, $P = 0.00448$; fig. 3a–c).

We compared π_0/π_4 for the inversions to genome-wide measures in a similar way to that in the above analysis of TE density. Contrary to what was expected, most of the nine inversions that we examined showed no difference in deleterious load ($P > 0.1$) when compared to the genomic background in each recombination rate category. In the few comparisons where significant ($P < 0.01$) results were obtained, such as for windows with "reduced" recombination rate in inversion arg10.01, the inversions have lower π_0/π_4 values, indicating fewer nonsynonymous

substitutions in those windows than the genome-wide average for regions with similar recombination rates (fig. 3d; supplementary fig. S8, Supplementary Material online). This pattern remained unchanged when only samples homozygous for the minor arrangement were included (fig. 3e; supplementary fig. S9, Supplementary Material online).

When looking at the ratio of nonsense mutations versus nonsynonymous mutations ($P_{nonsense}/P_0$), we found a stronger negative correlation with recombination rates in all species (*H. annuus*: Pearson's $r = -0.2770$, $P < 2.2 \times 10^{-16}$; *H. argophyllus*: Pearson's $r = -0.3104$,

FIG. 4. Recombination rate and proportion of nonsense mutations. Genome-wide ratio of nonsense mutations and zero-fold mutations ($P_{nonsense}/P_0$) was compared to recombination rate in (*a*) *Helianthus annuus*, (*b*) *H. argophyllus*, and (*c*) *H. petiolaris*. The red lines denote the best-fit linear regression line with the 95% confidence intervals shaded in blue. The values within inversions were compared to genome-wide regions in the same recombination rate category using (*d*) all samples or (*e*) samples homozygous for the rarer arrangement. Windows of each recombination rate category (high: >2 cM/Mbp, reduced: 0.01–2 cM/Mbp, null: <0.01 cM/Mbp) were compared separately. Asterisks denote significance in independent *t*-test: *0.05 > P > 0.01; **0.01 > P > 0.001. Results of all inversions are presented in supplementary figs. S10 and S11, Supplementary Material online.
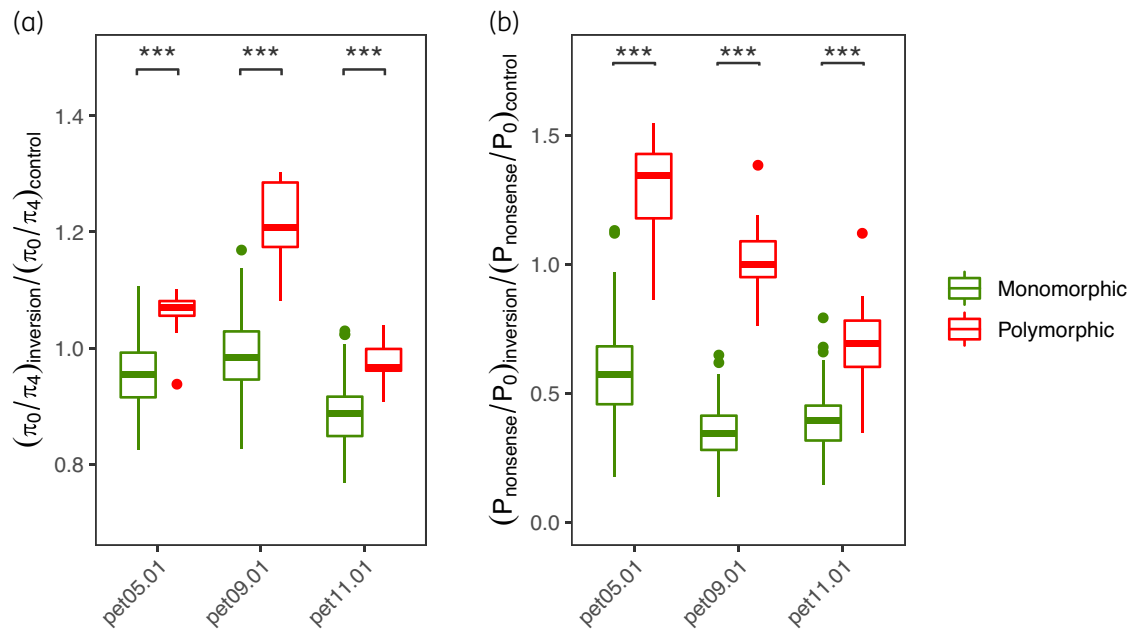
FIG. 5. Comparisons of (a) π_0/π_4 and (b) $P_{nonsense}/P_0$ between monomorphic populations and polymorphic populations for inversions pet05.01, pet09.01, and pet11.01. The statistics were calculated for samples with the same homozygous genotype from populations that were monomorphic and polymorphic for an inversion and normalized to account for differences in demography. Asterisks denote significance in independent t-test: ***$P < 0.001$. Distribution maps of inversion genotypes in populations used in the analyses are presented in supplementary fig. S12, Supplementary Material online.

$P < 2.2 \times 10^{-16}$; *H. petiolaris*: Pearson's $r = -0.2183$, $P < 2.2 \times 10^{-16}$; fig. 4a–c) than for π_0/π_4. Similar to the results with π_0/π_4, however, there was no excess in their proportion within inversions when compared with the genomewide average for windows with a similar recombination rate (fig. 4d; supplementary fig. S10, Supplementary Material online). One exception was in windows of the "reduced" recombination rate category in ann15.01 ($P = 0.0104$; fig. 4d), although the increase in the "null" recombination rate category in the same inversion was not significant. When only minor arrangements were analyzed, the "reduced" category in ann15.01 continued to have greater $P_{nonsense}/P_0$ than the genomic average (fig. 4e). Such a significant increase was also found in the "high" category in inversion pet11.01 and in pet14.01, while the $P_{nonsense}/P_0$ values did not differ significantly for other categories and inversions (fig. 4e; supplementary fig. S11, Supplementary Material online).

To more precisely estimate how genotype frequency of inversions affects the accumulation of deleterious load, we also compared π_0/π_4 and $P_{nonsense}/P_0$ between samples from populations that contain only one inversion arrangement and samples from populations that are polymorphic. For three inversions pet05.01, pet09.01, and pet11.01, which are mostly monomorphic in populations in one subspecies and polymorphic across populations in the other subspecies, homozygous samples from populations with both arrangements showed strikingly higher π_0/π_4 and $P_{nonsense}/P_0$ compared to samples of the same genotype in populations with only one arrangement, when

controlled for differences in demography (fig. 5; supplementary fig. S12, Supplementary Material online).

Establishment of the Inversion Polymorphism

Polymorphic inversions can arise from divergent selection or balancing selection (Faria et al. 2019), leading to different predictions regarding fitness of genotypes and, by inference, genotype frequency. In the circumstance of divergent selection, arrangements capturing different sets of locally adapted alleles should be favored in populations with different environmental conditions, and heterokaryotypes are likely to suffer reduced fitness if the alleles are not completely dominant. A balanced polymorphism can establish through favorable epistatic interactions among alleles at different loci in an inversion (Charlesworth and Charlesworth 1973), or through associative overdominance caused by complementation of recessive deleterious alleles by dominant alleles from the alternate arrangement. In the latter case, individuals heterozygous for an inversion should possess higher survival rates and occur at higher frequency than expected under Hardy-Weinberg equilibrium.

To test whether associative overdominance contributed to the maintenance of inversion polymorphism (Berdan et al. 2021; Connallon and Olito 2021), we used the extensive records of developmental, morphological and physiological traits for the samples from a previous common garden experiment (Todesco et al. 2020) to see if the inversions show overdominance in any phenotypes

FIG. 6. Excess of homozygosity of inversions. *U*-scores of both inversions and SNPs were calculated across whole species and scaled by number of samples. A positive *U*-score indicates excess of homozygotes, while a negative value indicates overrepresentation of heterozygotes. One individual was randomly chosen from each population and permuted for 100 times to generate sample sets. A hundred SNPs in the same 0.1 allele frequency bin as the inversion and in regions >100 kb from all inversions were randomly chosen as a control. Asterisks denote significance in independent *t*-test: ***$P < 0.001$.
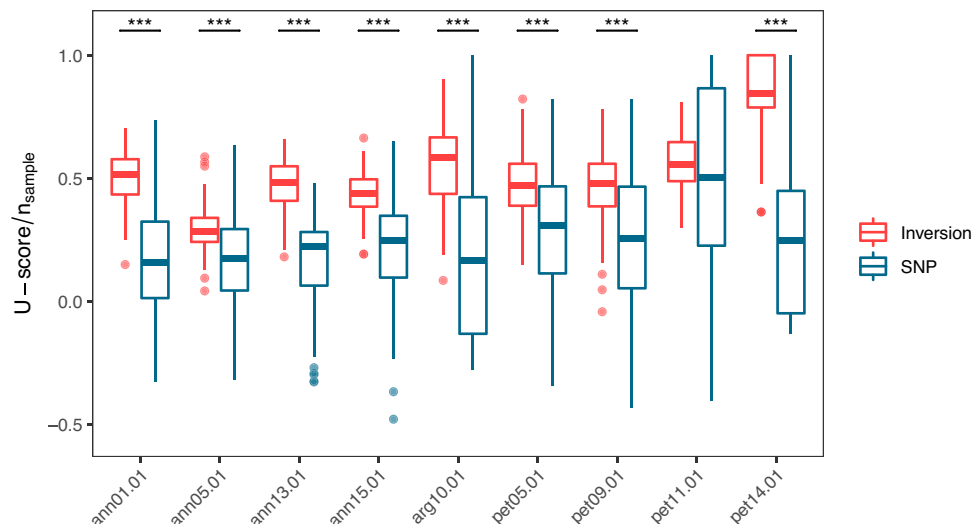
that may relate to fitness. Among the 87 traits across three species, many inversion genotypes did not differ in phenotype. In several cases, however, inversion heterozygotes fell between the homozygous genotypes, such as in ann01.01 for the diameter of the inflorescence disk and in pet05.01 for leaf area, or displayed a similar phenotype to one of the homozygotes, such as in ann13.01 for phyllary width and in pet11.01 for total leaf number. The only cases where heterozygotes have a significantly ($P < 0.01$) more extreme phenotype were in inversion ann15.01 for phyllary width and phyllary diameter (supplementary fig. S13, Supplementary Material online).

The presence of recessive deleterious alleles unique to one inversion arrangement will result in reduced fitness in homozygotes for the inversion and increase the frequency of heterozygotes. However, in our examination of genotype frequencies, we found no excess of inversion heterozygotes across the three species. Instead, when analyzed species-wide, there was an excess of homozygotes (*U*-score > 0) for all of the inversions that we examined, including two inversions that lack explicit evidence of selection (ann01.01 and ann05.01). All except one (pet11.01) of the inversions showed significantly higher ($P < 0.001$) homozygosity than the genomic background level (fig. 6). At the population level, *U*-scores were highly variable, but almost none were significantly different from Hardy-Weinberg expectations. Two exceptions did occur, where populations show a significant excess of homozygotes (supplementary fig. S14, Supplementary Material online).

Discussion

Using three genome assemblies of cultivated sunflower and newly generated SNP datasets in three wild sunflower species, we analyzed genome-wide patterns of recombination, TE abundance and protein evolution. We found that all of the statistics that we used regarding deleterious mutations, especially $P_{nonsense}/P_0$, displayed a significant negative correlation with recombination rate across the genome in all three species. These results are consistent with previous reports of enriched deleterious mutations in genomic regions with reduced levels of recombination (Lu et al. 2006; Renaut and Rieseberg 2015; Lozano et al. 2021). Furthermore, we found a negative correlation of genome-wide π_0/π_4 with effective population size in the three species. Reductions in population size and inbreeding will lower effective rates of recombination and reduce the efficacy of natural selection in removing nonadaptive mutations (Charlesworth et al. 1993; Lynch et al. 1993). As a consequence, deleterious mutations should be enriched in taxa with reduced population size, such as asexual species (Hollister et al. 2015), domesticated crops (Lu et al. 2006; Renaut and Rieseberg 2015), and taxa with a history of population bottlenecks (Marsden et al. 2016; Yang et al. 2018; Chen et al. 2020). Our results from the analyses of sunflower genomes are broadly consistent with this earlier work and further support the use of π_0/π_4 for measuring the load of deleterious mutations.

Besides protein sequences, we also found an enrichment of LTR-RTs in regions of low recombination in sunflowers. Negative associations between TEs and recombination rates have been commonly observed in eukaryotic genomes (Bartolome et al. 2002; Tian et al. 2009; Gion et al. 2016), although the factors driving these correlations remain poorly understood (Kent et al. 2017). One way in which recombination can affect TE accumulation is through the action of Hill-Robertson interference, assuming that the deleterious effects of TEs mainly result from insertions into genes or regulatory sequences. However, simulations under this model showed that TEs accumulate only in regions of extremely low recombination and when TEs do not excise (Dolgin and Charlesworth 2008). The deleterious effects of TEs can also be caused by ectopic recombination between non-homologous element copies. The ectopic recombination rate is assumed to closely

follow the meiotic recombination rate (Kent et al. 2017). Therefore, TEs in actively recombining regions are more likely to generate deleterious chromosomal rearrangements and thus be eliminated by selection. Unequal homologous recombination has been shown to be an important mechanism for the removal of LTR-RTs in sunflowers (Staton et al. 2012), and variation in its incidence across the genome may account for the distribution observed here. Another possible reason for the negative association with recombination is that the insertion of TEs leads to a direct modification of rates of recombination (Kent et al. 2017). However, recombination rate in sunflower species is relatively stable despite difference in TE content (Barb et al. 2014), so the large-scale recombination landscape estimated from genetic maps appears to be largely independent of TEs. Therefore, while variation in effective recombination rates undoubtedly affects the pattern of TE abundance in sunflowers, identification of the specific mechanisms driving TE accumulation in sunflower genomes require further investigation.

While all statistics, including LTR-RT density, π_0/π_4 and $P_{nonsense}/P_0$, captured the effects of population size and recombination rate variation across the genome, none of the nine locally-adapted inversions that we examined displayed the predicted increase in TE abundance or deleterious amino acid substitutions when compared to other regions of the genome with a similar background recombination rate, except for increases in $P_{nonsense}/P_0$ ratio in parts of three inversions (ann15.01, pet11.01 and pet14.01).

The finding of an increased $P_{nonsense}/P_0$ ratio in parts of the ann15.01, pet11.01 and pet14.01 inversions implies that it is a more sensitive indicator of deleterious load than LTR-RT density or π_0/π_4. This observation is consistent with the stronger correlation of the former with recombination rate (fig. 4a–c). On the other hand, these inversions are generally large in size and large inversions are more likely to carry deleterious mutations during their establishment (Santos 1986; Connallon and Olito 2021). In ann15.01 and pet11.01, the increase in $P_{nonsense}/P_0$ was found in windows with high or moderate background recombination rate, but not in windows of the "null" recombination rate category within the same inversions, possibly suggesting that the impact of recombination suppression may be stronger in those regions. It is noteworthy that for pet14.01 an impact of recombination suppression is only seen in the minor rearrangement, which has the lowest minor allele frequency among inversions examined (table 1). Thus, allele frequency likely plays a critical role in affecting genetic load in these inversions.

Consistent with observations of minimal deleterious load in sunflower inversions, we found little evidence of overdominance of these inversions in the 87 traits examined. There was a hint of heterosis in ann15.01, one of the largest inversions, for phyllary width and diameter. This is consistent with the elevated $P_{nonsense}/P_0$ observed in this inversion and with a possible link between inversion lengths and deleterious mutations (Connallon and Olito

2021). For many important traits, the inversions were dominant or incompletely dominant, which does not confer an advantage on heterozygotes. One case of underdominance was found in pet05.01 for internode length, but whether this confers advantage or disadvantage on heterozygotes is unknown. Nevertheless, we did not perform comprehensive association analyses that account for population structure so the differences in phenotypes that we observed may be caused by other genetic variation. Furthermore, we only tested a limited set of phenotypes, and we assumed that extrema in a single trait might affect survival while certain combinations of traits might be the target of natural selection. Common garden experiments in native habitats are needed to investigate survival rate of different inversion genotypes to confirm associative overdominance (e.g., Goebl et al. 2020).

Analyses of genotype frequencies of inversions also failed to support the overdominance model. Across the ranges of species, we found an excess of homozygotes for all inversions, which is contrary to expectations for heterosis. A previous examination of inversion genotypes in a dune system in *H. petiolaris* also provided evidence of a heterozygosity deficit for inversions pet05.01, pet09.01 and pet11.01 (Huang et al. 2020). However, population structure may also cause deviations from Hardy-Weinberg equilibrium, so this result should be viewed as an overall summary of genotypic proportions within species, rather than a strict test of differential mating or survivorship among genotypes within populations. Nevertheless, when compared to SNPs with similar allele frequencies from the rest of the genome, the inversions displayed higher homozygosity, consistent with the presence of selection for inversion homozygotes and/or against heterozygotes, or greater environmental selection. For local populations where gene flow should be unrestricted, most did not have *U*-scores significantly different from 0, although the sample size for each population was too small for a valid test (supplementary tables S2–S4, Supplementary Material online). In addition, the samples used were collected from seeds on mature plants, and thus reflected post-mating population frequencies rather than those of living plants, so the impact of viability selection could not have been detected (Goebl et al. 2020).

Although the cause for the excess of inversion homozygotes within sunflower species is not fully resolved, multiple lines of evidence have demonstrated that many inversions in sunflowers are under divergent ecological selection. For example, previous studies have associated inversions with numerous environmental variables and locally adapted traits (Huang et al. 2020; Todesco et al. 2020). For instance, in *H. annuus*, inversion ann13.01 was associated with flowering time and climate continentality; in *H. argophyllus*, inversion arg10.01 is enriched on the barrier islands of Texas and was associated with temperature and relative humidity; pet05.01, pet09.01, pet11.01, and pet14.01 were all found to contribute to dune adaptation in *H. petiolaris* (Huang et al. 2020; Todesco et al. 2020). Hybrids between dune and nondune ecotypes of

H. petiolaris, which are enriched with different inversion arrangements, showed reduced fitness in either of the local environments in reciprocal transplant experiments (Ostevik et al. 2016), and Goebl et al. (2020) showed that several of these inversions were under divergent natural selection. Such divergent selection could account, in part, for the excess of homozygous genotypes, especially at locations where selection is strong and likely explains the contrary results in sunflower inversions and those in Insecta, many of which establish via associative overdominance (Butlin and Day 1985; Eanes et al. 1992; Sniegowski and Charlesworth 1994; Jay et al. 2021).

Assortative mating may also play a role in maintaining the high homozygosity of inversion arrangement. Such assortative mating can occur via ecogeographic isolation since, as discussed above, sunflower inversions contribute to local adaptation. Conspecific pollen precedence was found between dune and non-dune ecotypes in *H. petiolaris* (Ostevik et al. 2016), although whether the underlying loci are located in inversions is still unknown. However, theoretical studies suggest that this is likely the case since inversions offer a means to establish LD between loci for local adaption and assortative mating (Trickett and Butlin 1994; Servedio 2009; Huang and Rieseberg 2020).

Our analyses of the deleterious load among populations with differences in genotype composition further confirmed that levels of deleterious load in inversions vary with heterozygote frequencies. In populations where inversions are polymorphic, inversions exhibited higher deleterious load compared to samples from populations where only one arrangement is found (fig. 5). These striking differences between polymorphic and monomorphic populations confirm that recombination suppression in inversion heterozygotes, to some extent, has detrimental consequences in sunflower and that high homozygosity of inversions largely averts such cost. We also expect increased TE abundance in populations with a higher frequency of inversion heterozygotes, but because of the difficulty in precisely identifying TEs within inversions using short reads data, we could not test this hypothesis in the present study.

In addition to high homozygosity of inversions, gene flux can also mitigate the accumulation of deleterious mutations (Berdan et al. 2021). Our examination of the pattern of F_{ST} between karyotypes suggest that while inversions allow independent evolution of chromosomes with different arrangements, there is a visible level of gene flux between arrangements due to double recombination or gene conversion, especially in the largest inversion ann13.01 (supplementary fig. S3, Supplementary Material online). Such flux may further limit deleterious mutation accumulation in the inversions, but precise measurement of gene flux between arrangements requires genetic mapping in specific crosses (Crown et al. 2018; Korunes and Noor 2019).

Taken together, our results suggest that the large inversions we have discovered to be segregating in sunflower populations are generally less prone to deleterious mutation accumulation than expected. This is mainly due to the high frequency of inversion homozygotes in natural populations, but gene flux may contribute as well. High homozygosity of inversion arrangements should be common for locally adapted inversions, especially when the environmental differences are correlated with geography and vagility is limited. Theoretical studies of the role of recessive deleterious mutations in inversions normally assume mutation-selection equilibria (Nei et al. 1967; Ohta 1971; Santos 1986; Berdan et al. 2021). Analyses of deleterious mutation accumulation in the common scenario of local adaptation are required for a better understanding of inversion fates.

Recombination increases the rate of adaptive evolution by bringing beneficial mutations together and decoupling them from deleterious mutations, while the suppression of recombination can facilitate adaptive divergence in the face of gene flow by holding advantageous alleles together. Thus, recombination modifiers that simultaneously permit high levels of recombination within populations, while reducing such recombination between ecologically divergent populations should be favored. Our analyses in sunflowers suggest that inversions can offer such versatility; recombination can occur among chromosomes of the same arrangement, but not between arrangements, thereby permitting divergence with gene flow, whilst largely averting the accumulation of deleterious mutations. This contrasts with other genomic features known to suppress meiotic recombination, such as telomeric and heterochromatic regions, which suppress recombination universally. Conversely, several other kinds of structural variants may affect recombination in a similar way as inversions, thus providing an evolutionary advantage. For example, hemizygosity resulting from insertion and deletion polymorphisms also precludes recombination between chromosomes with different haplotypes, but not within those possessing the same haplotype (He and Dooner 2009; Schwander et al. 2014; Lawrence et al. 2017). However, compared to inversions, because deleterious mutations would not be masked in hemizygous genotypes, hemizygosity may follow a different evolutionary trajectory.

Materials and Methods

Samples and Variant Calling

Most of the sequences analyzed in the present study were generated by Todesco et al. (2020). In brief, samples were grown from seeds collected from wild populations across the native range of each species. Tissue from young leaves was collected from all individual plants, and genomic DNA was extracted from leaf tissue to prepare paired-end whole-genome shotgun Illumina libraries. After treatment with duplex-specific nuclease to reduce the representation of repetitive sequences, the libraries were sequenced on HiSeq2500, HiSeq4000 and HiSeqX instruments to produce paired-end, 150-bp reads. Illumina adapters and poor quality reads were clipped using Trimmomatic

v0.36 (Bolger et al. 2014) and reads shorter than 36 bp were discarded. These processes resulted in an average 6.34-fold coverage of gene space (Todesco et al. 2020).

To increase sampling of rare inversion arrangements, we collected and sequenced additional samples from two dune systems in *H. petiolaris*, where a number of inversions are enriched in the dune ecotypes (Todesco et al. 2020). Specifically, in 2008, we visited Great Sand Dunes National Park and Preserve in Colorado, USA, and collected seeds from 11 dune and 9 nondune subpopulations; in 2011, we collected seeds from 8 dune and 10 nondune subpopulations in Monahans Sandhills State Park in Texas, USA (supplementary table S1, Supplementary Material online). To avoid sequencing recent immigrants in each subpopulation, we selected one maternal family from each subpopulation that had seed sizes within the expected range of that subpopulation (i.e., larger seeds for dune subpopulations and smaller seeds for non-dune subpopulations). We germinated seeds in the lab and collected tissue from young leaves from one individual per maternal family (i.e., one individual per population). High molecular weight DNA was extracted from leaf tissue using a modified CTAB protocol (Murray and Thompson 1980; Zeng et al. 2002). TruSeq gDNA libraries for each individual were prepared and sequenced at Genome Quebec Innovation Center (Montreal, Canada). Each library was split across two lanes of Illumina HiSeq 2000 to produce 100-bp paired-end reads.

All the samples were aligned to a newly released reference genome for *H. annuus* (Ha412Hov2.0), which used Hi-C (Marie-Nelly et al. 2014) for contig and scaffold ordering and was shown to have improved quality (Todesco et al. 2020). The filtered reads were aligned to the new reference genome using NextGenMap v0.5.3 (Sedlazeck et al. 2013) and resulting BAM files were concatenated, sorted, and duplicate-marked using samtools v0.1.19 (Li et al. 2009). Libraries sequenced in multiple lanes were merged with sambamba v0.6.6 (Tarasov et al. 2015), and PCR duplicates were remarked.

Variant calling was performed with the Genome Analysis Tool Kit v 4.1.4.1 (GATK; McKenna et al. 2010). To reduce computational time in the variant calling process, we excluded genomic regions that contain TEs, which represent more than ¾ of the sunflower genome (Badouin et al. 2017), as well as small unplaced chromosome contigs, chloroplast, and mitochondria. For each sample, a GVCF file was produced with the GATK "HaplotypeCaller" with the parameter "–heterozygosity 0.01". After individual variant calling, all samples from each species were jointly genotyped using GATK's "GenomicsDBImport" and "GenotypeGVCFs". The step was run over 1 Mbp regions of the genome for parallel computation, and the raw VCF chunks were then gathered by chromosome using "GatherVcfs".

In order to remove low-quality variants, we followed GATK best practices and conducted Variant Quality Score Recalibration (VQSR) to filter the raw VCFs. The 20 samples with the highest sequencing depth in each species were selected to produce a "gold set" using the following parameters: mapping quality > 50.0, missing rate $< 10\%$, $-1.0 <$ strand odds ratio < 1.0, minor allele frequency > 0.25, excess heterozygosity < 10.0, $-1.0 <$ BaseQRankSum < 1.0, depth within one standard deviation from the mean and ExcessHet z-score > -4.5. The raw set of all variants were first filtered to remove sites with extremely heterozygosity (ExcessHet z-score < -4.5) and the gold set was then applied against this filtered set of variants to produce recalibration models for SNPs and indels using "VariantRecalibrator." The 90% tranche for each species was selected based on these recalibration models using "ApplyVQSR." An additional filter was applied for each species to retain only bi-allelic SNPs with minor allele frequency > 0.01 and genotyping rate $> 50\%$.

Estimation of Recombination Rate and Localization of Centromeres

We remapped the markers of an integrated genetic map for cultivated sunflower (Badouin et al. 2017) to our current reference genome Ha412Hov2.0 using BWA MEM (Li 2013) with default settings. A cubic smoothing spline was fit to the remapped marker positions on each chromosome and outliers with residual > 1 were removed. We manually inverted the first 10 Mbp of chromosome 12 in the genetic map that was likely due to assembly error, and we fit monotonic increasing genetic coordinates at a resolution of 1 Mbp using a smoothing spline model. Recombination rate was calculated for each 500 kbp nonoverlapping window based on the physical and genetic distances on the transposed genetic map.

To determine the location of sunflower centromeres, we made use of the sequences previously found to be targeted by sunflower centromere-specific histone H3 (Nagaki et al. 2015). Four main sequences from (Nagaki et al. 2015) were downloaded from the Nucleotide database of National Center for Biotechnology Information with the following accession: LC075744, LC075745, LC075746, and LC075747. These sequences were then queried against the Ha412Hov2.0 genome using BLASTN (https://blast.ncbi.nlm.nih.gov/) with an E-value of 1×10^{-5}.

LD and Sequence Differentiation

To estimate the extent of recombination suppression caused by the inversions, we calculated LD between SNPs across the chromosomes where the inversions reside. We removed SNPs with minor allele frequencies $< 10\%$, thinned the variants to one per 5 kbp and calculated pairwise LD (R^2) values between SNPs using PLINK v1.9 (Purcell et al. 2007; Chang et al. 2015). For each inversion, we calculated LD for all samples from populations that are polymorphic for the inversion arrangements with a minor allele frequency > 0.1 and for samples from populations that contain only the homozygous genotype for the major arrangement (Todesco et al. 2020). Values of SNPs were

grouped into 500 kbp windows and the average R^2 value was calculated for each window pair.

To provide an additional assessment of the effect of inversions in suppressing recombination between arrangements, we calculated Weir and Cockerham's F_{ST} (Weir 1996) in 1-Mbp sliding windows with a step size of 200 kbp and 100 kbp sliding windows with a step size of 20 kbp, respectively, between samples homozygous for each arrangement for each inversion using VCFtools (Danecek et al. 2011) with the SNPs generated on the new reference genome.

TE Abundance

We conducted de-novo TE annotation for the three reference genomes of cultivated sunflowers using the program EDTA (Ou et al. 2019). The program was run with the default settings "−sensitive 0 −evaluate 0" and parameter "−anno 1" to perform whole-genome TE annotation on the 17 chromosomes of each genome assembly, and known coding sequences of sunflower from the gene annotation of the Ha412HOv2.0 genome (https://sunflowergenome.org/) were provided for TE filtration. We extracted LTR-RTs from the EDTA-generated annotations for downstream analyses because these elements are known to be transcriptionally active in sunflower species (Cavallini et al. 2010; Kawakami et al. 2011; Renaut et al. 2014).

We summarized the density of LTR-RTs by calculating the proportion of structurally intact and fragmented elements (including solo LTRs and truncated LTR-RTs) in sliding windows of 500 kbp across the genome. The TE densities were then compared with the recombination rates calculated on the Ha412HOv2.0 genome. Because recombination rates are highly heterogeneous across the genome, the accumulation of TEs within an inversion likely depends in part on the local recombination environment. To control for the effect of genomic context on TE density, we divided the windows into three different categories based on the recombination rate: 1) high recombination rate for windows with recombination rate > 2 cM/Mbp, 2) reduced recombination rate regions with recombination rate between 0.01 and 2 cM/Mbp, and null recombination regions with recombination rate < 0.01 cM/Mbp 3) (Bartolome et al. 2002). For each inversion, windows of each category were compared to those of the same category from across the genome. TE proportions were logit-transformed and compared using Welch two sample t-tests. For the inversions on the other two reference genomes without direct recombination rate estimation, we permuted the windows within each inversion 1,000 times and compared the averages to those summarized from the same number of windows across the genome using the same method. Genotypes and locations of ann01.01 and ann05.01 on each reference genome were extracted based on previous alignments of the genome assemblies (Todesco et al. 2020).

Protein Evolution

In addition to TEs, we examined if inversions show a greater accumulation of deleterious mutations in protein coding genes relative to background levels. To detect signatures of relaxed negative selection, we explored the pattern of protein evolution using the coding sequences of all protein-coding genes from a recently generated gene annotation for the Ha412HOv2.0 reference genome. In brief, the annotation was produced using the program EuGene (Foissac et al. 2008) and the predicted gene sequences were filtered by removing genes that were aligned to a known TE with > 90% identity and over 80% of its length (https://sunflowergenome.org/). The remaining gene set contains 58,234 genes, 53,800 of which are protein-coding genes. We further removed the coding sequences that do not begin with the regular "ATG" start codons or end with "TAG/TGA/TAA" stop codons or those whose lengths are not a multiple of three, leaving a total of 53,302 genes which were used for all downstream analyses.

We first calculated the ratio of zero-fold to four-fold diversity π_0/π_4 (Marsden et al. 2016; Yang et al. 2018; Chen et al. 2020), which, as an approximation of π_N/π_S, estimates the effect on selection on protein sequences (Hollister et al. 2015; Hahn 2019). The zero and four-fold degenerate sites were identified by iterating across all four possible bases at each site along the coding sequences and comparing the resulting amino acids. Sites where the four different bases resulted in four different amino acids were classified as zero-fold degenerate and those with no changes in amino acids were defined as four-fold degenerate. The diversity at zero-fold or four-fold degenerate sites was calculated as:

$$\pi = \sum_{j=1}^{S} h_j$$

where S is the number of segregating sites and h_j is the heterozygosity for the jth segregating site defined as:

$$h = \frac{n}{n-1}\left(1 - \sum p_i^2\right)$$

where n is the number of haplotype sequences in the sample and p_i is the sample frequency at the ith allele (Tajima 1989; Hahn 2019). The ratio of π_0 and π_4 was then calculated in sliding windows of 500 kbp using samples from all populations in each of the three species, or samples homozygous for the minor arrangement. The ratios were then compared to recombination rates and the values for the inversions were compared to genome-wide measures in a similar way to that in the above analysis of TE density, except that the ratios were log-transformed and recombination rate categories with less than 10 windows in an inversion were removed due to low statistical power.

To more precisely estimate the effect of inversion genotype frequency on the accumulation of deleterious load,

for inversions pet05.01, pet09.01, and pet11.01, which are mostly monomorphic in populations in one subspecies and polymorphic in populations in the other (supplementary fig. S12, Supplementary Material online), we chose samples from populations that were homozygous for one arrangement and those of the same homozygous genotype from populations polymorphic with a minor allele frequency > 0.2 and compared π_0/π_4 between these samples. To control for differences in demography, a region of the same size as the inversions on chromosome 4, where no inversions have been identified in sunflowers, was randomly chosen and the statistics were calculated with the same sets of samples. In addition, we randomly selected 20 samples from each species and calculated π_0/π_4 across the whole genome to estimate the deleterious load among species.

Because nonsense mutations that cause premature termination of protein synthesis are normally highly deleterious and subjective to purifying selection (Chu and Wei 2019), we further examined stop codon mutations using the same protein-coding gene set. We annotated the VCFs using the program snpEff v 5.0c (Cingolani et al. 2012) and extracted "nonsense" mutations (alternate stop codon) for each species. The deleterious load was represented as the ratio of the number of nonsense mutations $P_{nonsense}$ and the number of mutations at zero-fold sites P_0, which was used as an approximation of nonsynonymous mutations (Renaut and Rieseberg 2015). We calculated this statistic in sliding windows of 500 kbp and compared the inversions to genome-wide measures in the same way as for π_0/π_4. We also compared $P_{nonsense}/P_0$ between monomorphic and polymorphic populations in the same way as for π_0/π_4.

Overdominance of Inversions

We used the phenotypic data of the samples collected in a previous common garden experiment (Todesco et al. 2020) to test whether inversions show overdominance in the wild population. In total, we tested 86, 30 and 69 traits for *H. annuus*, *H. argophyllus*, and *H. petiolaris*, respectively (supplementary fig. S13, Supplementary Material online). The trait measurements were normalized and the averages of inversion heterozygotes were compared to those of each homozygous genotype. Significance was assessed using two-sample t-test if values of heterozygotes were greater or lower than both homozygous groups.

Genotype Frequency

We used the genotype information of the samples from (Todesco et al. 2020) to calculate allele frequency and genotype frequency of the inversions (supplementary tables S2–S4, Supplementary Material online) and tested for Hardy-Weinberg equilibrium. We calculated U-score for each inversion using the full-enumeration method of "hwx.test" in the R package "HWxtest" v1.1.9 (Engels 2009). A positive U-score indicates excess of homozygotes while a negative value indicates overrepresentation of heterozygotes. We calculated this score using all samples across the species range as well as for each population. For the species-wide calculation, we randomly selected 1 individual from each population for each inversion and repeated the process for 100 times to estimate U-score, and we randomly chose 100 bi-allelic SNPs in the same 0.1 allele frequency bin as the inversion and in regions >100 kb away from all inversions to account for possible population structure.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Authors' Contributions

K.H. and L.H.R. conceived the study; M.T. and N.B. contributed sequence data; K.L.O. collected and produced sequence data for new samples of *Helianthus petiolaris*; K.H. performed SNP calling, TE annotation, and analyzed the data; C.E. mapped the centromeres; G.L.O. helped with analyses; K.H. and L.H.R. wrote the paper; and all authors approved the final manuscript.

Data Availability

The new sequence data generated in this study have been deposited in the Sequence Read Archive database (accession number PRJNA322345; https://www.ncbi.nlm.nih.gov/bioproject/PRJNA322345). Other raw sequence data used in this study are stored under accessions PRJNA532579, PRJNA398560, and PRJNA564337. The Ha412HOv2.0, XRQv2, PSC8 genome assemblies are available at https://sunflowergenome.org/ and https://heliagene.org/. The newly generated SNP datasets are available at http://www.helianthome.org/. The trait data used in the overdominance analysis are available at https://easygwas.ethz.ch/gwas/myhistory/public/20/, https://easygwas.ethz.ch/gwas/myhistory/public/21/, https://easygwas.ethz.ch/gwas/myhistory/public/22/, and https://easygwas.ethz.ch/gwas/myhistory/public/23/. Codes associated with this paper and locations of the centromere-specific sequences can be found on GitHub: https://github.com/hkchi/DelMut_inv/.

References

Bachtrog D. 2003. Adaptation shapes patterns of genome evolution on sexual and asexual chromosomes in *Drosophila*. *Nat Genet*. **34**(2):215–219.

Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, Lelandais-Briere C, Owens GL, Carrere S, Mayjonade B, *et al.* 2017. The sunflower genome provides insights into oil metabolism, flowering and asterid evolution. *Nature* **546**(7656):148–152.

Barb JG, Bowers JE, Renaut S, Rey JI, Knapp SJ, Rieseberg LH, Burke JM. 2014. Chromosomal evolution and patterns of introgression in *Helianthus*. *Genetics* **197**(3):969–979.

Bartolome C, Maside X, Charlesworth B. 2002. On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. *Mol Biol Evol*. **19**(6):926–937.

Berdan EL, Blanckaert A, Butlin RK, Bank C. 2021. Deleterious mutation accumulation and the long-term fate of chromosomal inversions. *PLoS Genet*. **17**(3):e1009411.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* **30**(15): 2114–2120.

Butlin RK, Day TH. 1985. Genie and karyotypic selection on an inversion polymorphism in the seaweed fly, *Coelopa frigida*. *Heredity* **54**(2):267–274.

Carvalho AB. 2003. The advantages of recombination. *Nat Genet*. **34**(2):128–129.

Cavallini A, Natali L, Zuccolo A, Giordani Y, Jurman I, Ferrillo V, Vitacolonna N, Sarri V, Cattonaro F, Ceccarelli M, *et al.* 2010. Analysis of transposons and repeat composition of the sunflower (*Helianthus annuus* L.) genome. *Theor Appl Genet*. **120**(3): 491–508.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**(1):7.

Charlesworth B, Barton NH. 2018. The spread of an inversion with migration and selection. *Genetics* **208**(1):377–382.

Charlesworth B, Charlesworth D. 1973. Selection of new inversions in multi-locus genetic systems. *Genet Res*. **21**(2):167–183.

Charlesworth D, Morgan MT, Charlesworth B. 1993. Mutation accumulation in finite populations. *J Hered*. **84**(5):321–325.

Chen Z, Ai F, Zhang J, Ma X, Yang W, Wang W, Su Y, Wang M, Yang Y, Mao K, *et al.* 2020. Survival in the tropics despite isolation, inbreeding and asexual reproduction: insights from the genome of the world's southernmost poplar (*Populus ilicifolia*). *Plant J*. **103**(1):430–442.

Chu D, Wei L. 2019. Nonsynonymous, synonymous and nonsense mutations in human cancer-related genes undergo stronger purifying selections than expectation. *BMC Cancer* **19**(1):1–12.

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: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**(2):80–92.

Connallon T, Olito C. 2021. Natural selection and the distribution of chromosomal inversion lengths. *Mol Ecol*. **00**:1–15.

Coughlan JM, Willis JH. 2019. Dissecting the role of a large chromosomal inversion in life history divergence throughout the *Mimulus guttatus* species complex. *Mol Ecol*. **28**(6): 1343–1357.

Crown KN, Miller DE, Sekelsky J, Hawley RS. 2018. Local inversion heterozygosity alters recombination throughout the genome. *Curr Biol*. **28**(18):2984–2990.

Crumpacker DW, Salceda VM. 1969. Chromosomal polymorphism and genetic load in *Drosophila pseudoobscura*. *Genetics* **61**(4): 859–873.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, Depristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**(15): 2156–2158.

Daron J, Glover N, Pingault L, Theil S, Jamilloux V, Paux E, Barbe V, Mangenot S, Alberti A, Wincker P, *et al.* 2014. Organization and evolution of transposable elements along the bread wheat chromosome 3B. *Genome Biol*. **15**(12):1–15.

Dolgin ES, Charlesworth B. 2008. The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics* **178**(4):2169–2177.

Eanes WF, Wesley C, Charlesworth B. 1992. Accumulation of P elements in minority inversions in natural populations of *Drosophila melanogaster*. *Genet Res*. **59**(1):1–9.

Engels WR. 2009. Exact tests for Hardy-Weinberg proportions. *Genetics* **183**(4):1431–1441

Faria R, Johannesson K, Butlin RK, Westram AM. 2019. Evolving inversions. *Trend Ecol Evol*. **34**(3):239–248.

Feder JL, Gejji R, Powell TH, Nosil P. 2011. Adaptive chromosomal divergence driven by mixed geographic mode of evolution. *Evolution* **65**(8):2157–2170.

Felsenstein J. 1974. The evolutionary advantage of recombination. *Genetics* **78**:737–756.

Felsenstein J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* **35**(1):124–138.

Foissac S, Gouzy J, Rombauts S, Mathe C, Amselem J, Sterck L, de Peer YV, Rouze P, Schiex T. 2008. Genome annotation in plants and fungi: EuGene as a model platform. *Curr Bioinform*. **3**(2):87–97.

Fuller ZL, Koury SA, Phadnis N, Schaeffer SW. 2019. How chromosomal rearrangements shape adaptation and speciation: case studies in *Drosophila pseudoobscura* and its sibling species *Drosophila persimilis*. *Mol Ecol*. **28**(6):1283–1301.

Gion JM, Hudson CJ, Lesur I, Vaillancourt RE, Potts BM, Freeman JS. 2016. Genome-wide variation in recombination rate in *Eucalyptus*. *BMC Genomics* **17**(1):1–12.

Goebl AM, Kane NC, Doak DF, Rieseberg LH, Ostevik KL. 2020. Contrasting selection at multiple life stages maintains divergent adaptation between sunflower ecotypes. *bioRxiv*, 2020.08.08. 242503. https://www.biorxiv.org/content/10.1101/2020.08.08. 242503v1 (accessed 10 August 2020).

Guerrero RF, Rousset F, Kirkpatrick M. 2012. Coalescent patterns for chromosomal inversions in divergent populations. *Phil Trans R Soc B*. **367**(1587):430–438.

Hahn MW. 2019. *Molecular population genetics*. Sunderland (MA): Sinauer Associates.

He L, Dooner HK. 2009. Haplotype structure strongly affects recombination in a maize genetic interval polymorphic for *Helitron* and retrotransposon insertions. *Proc Natl Acad Sci USA*. **106**(21): 8410–8416.

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

Hollister JD, Greiner S, Wang W, Wang J, Zhang Y, Wong GKS, Wright SI, Johnson MT. 2015. Recurrent loss of sex is associated with accumulation of deleterious mutations in *Oenothera*. *Mol Biol Evol*. **32**(4):896–905.

Huang K, Andrew RL, Owens GL, Ostevik KL, Rieseberg LH. 2020. Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Mol Ecol*. **29**(14):2535–2549.

Huang K, Rieseberg LH. 2020. Frequency, origins, and evolutionary role of chromosomal inversions in plants. *FrontPlant Sci*. **11**:296.

Irwin DE. 2018. Sex chromosomes and speciation in birds and other ZW systems. *Mol Ecol*. **27**(19):3831–3851.

Jay P, Chouteau M, Whibley A, Bastide H, Parrinello H, Llaurens V, Joron M. 2021. Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nat Genet*. **53**(3):288–293.

Kaiser VB, Charlesworth B. 2009. The effects of deleterious mutations on evolution in non-recombining genomes. *Trends Genet*. **25**(1): 9–12.

Kawakami T, Dhakal P, Katterhenry AN, Heatherington CA, Ungerer MC. 2011. Transposable element proliferation and genome expansion are rare in contemporary sunflower hybrid populations

despite widespread transcriptional activity of LTR retrotransposons. *Genome Biol Evol.* **3**:156–167.

Keightley PD, Otto SP. 2006. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443**(7107):89–92.

Kent TV, Uzunovic J, Wright SI. 2017. Coevolution between transposable elements and recombination. *Phil Trans R Soc B.* **372**(1736): 20160458.

Kirkpatrick M, Barton N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* **173**(1):419–434.

Korunes KL, Noor MA. 2019. Pervasive gene conversion in chromosomal inversion heterozygotes. *Mol Ecol.* **28**(6):1302–1315.

Langley CH, Montgomery E, Hudson R, Kaplan N, Charlesworth B. 1988. On the role of unequal exchange in the containment of transposable element copy number. *Genet Res.* **52**(3):223–235.

Lawrence EJ, Griffin CH, Henderson IR. 2017. Modification of meiotic recombination by natural variation in plants. *J Exp Bot.* **68**(20): 5471–5483.

Lee CR, Wang B, Mojica JP, Mandakova T, Prasad KV, Goicoechea JL, Perera N, Hellsten U, Hundley HN, Johnson J, et al. 2017. Young inversion with multiple linked QTLs under selection in a hybrid zone. *Nat Ecol Evol.* **1**(5):0119.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv.* 1303.3997. See https://arxiv.org/abs/1303.3997 (accessed 26 May 2013).

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009 The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**:2078–2079.

Lowry DB, Willis JH. 2010. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol.* **8**(9): e1000500.

Lozano R, Gazave E, Dos Santos JP, Stetter MG, Valluru R, Bandillo N, Fernandes SB, Brown PJ, Shakoor N, Mockler TC, et al. 2021. Comparative evolutionary genetics of deleterious load in sorghum and maize. *Nat Plants.* **7**(1):17–24.

Lu J, Tang T, Tang H, Huang J, Shi S, Wu CI. 2006. The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends Genet.* **22**(3): 126–131.

Lynch M, Bürger R, Butcher D, Gabriel W. 1993. The mutational meltdown in asexual populations. *J Hered.* **84**(5):339–344.

Ma J, Bennetzen JL. 2006. Recombination, rearrangement, reshuffling, and divergence in a centromeric region of rice. *Proc Natl Acad Sci USA.* **103**(2):383–388.

Marie-Nelly H, Marbouty M, Cournac A, Flot JF, Liti G, Parodi DP, Syan S, Guillén N, Margeot A, Zimmer C, et al. 2014. High-quality genome (re)assembly using chromosomal contact data. *Nat Commun.* **5**(1):1–10.

Marsden CD, Ortega-Del Vecchyo D, O'Brien DP, Taylor JF, Ramirez O, Vila C, Marques-Bonet T, Schnabel RD, Wayne RK, et al. 2016. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc Natl Acad Sci USA.* **113**(1):152–157.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**(9):1297–1303.

Montgomery E, Charlesworth B, Langley CH. 1987. A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. *Genet Res.* **49**(1):31–41.

Morrell PL, Buckler ES, Ross-Ibarra J. 2012. Crop genomics: advances and applications. *Nat Rev Genet.* **13**(2):85–96.

Moyers BT. 2015. The landscape of divergence in silverleaf sunflowers [Doctoral dissertation]. [Vancouver (BC)]: University of British Columbia.

Murray MG, Thompson WF. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **8**(19):4321–4326.

Nagaki K, Tanaka K, Yamaji N, Kobayashi H, Murata M. 2015. Sunflower centromeres consist of a centromere-specific LINE and a chromosome-specific tandem repeat. *Front Plant Sci.* **6**:912.

Navarro A, Barbadilla A, Ruiz A. 2000. Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in *Drosophila*. *Genetics* **155**(2):685–698.

Navarro A, Barton NH. 2003. Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution* **57**(3):447–459.

Navarro A, Betran E, Barbadilla A, Ruiz A. 1997. Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. *Genetics* **146**(2):695–709.

Nei M, Kojima KI, Schaffer HE. 1967. Frequency changes of new inversions in populations under mutation-selection equilibria. *Genetics* **57**(4):741–750.

Noor MA, Grams KL, Bertucci LA, Reiland J. 2001. Chromosomal inversions and the reproductive isolation of species. *P Natl Acad Sci.* **98**(21):12084–12088.

Ohta T. 1971. Associative overdominance caused by linked detrimental mutations. *Genet Res.* **18**(3):277–86.

Ortiz-Barrientos D, Engelstädter J, Rieseberg LH. 2016. Recombination rate evolution and the origin of species. *Trend Ecol Evol.* **31**(3):226–236.

Ostevik KL, Andrew RL, Otto SP, Rieseberg LH. 2016. Multiple reproductive barriers separate recently diverged sunflower ecotypes. *Evolution* **70**(10):2322–2335.

Ostevik KL, Samuk K, Rieseberg LH. 2020. Ancestral reconstruction of karyotypes reveals an exceptional rate of nonrandom chromosomal evolution in sunflower. *Genetics* **214**(4): 1031–1045.

Ou S, Su W, Liao Y, Chougule K, Agda JR, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, et al. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**(1):1–18.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* **81**(3):559–575.

Renaut S, Rieseberg LH. 2015. The accumulation of deleterious mutations as a consequence of domestication and improvement in sunflowers and other compositae crops. *Mol Biol Evol.* **32**(9): 2273–2283.

Renaut S, Rowe HC, Ungerer MC, Rieseberg LH. 2014. Genomics of homoploid hybrid speciation: diversity and transcriptional activity of long terminal repeat retrotransposons in hybrid sunflowers. *Phil Trans R Soc B.* **369**(1648):20130345.

Santos M. 1986. The role of genic selection in the establishment of inversion polymorphism in *Drosophila subobscura*. *Genetica* **69**(1):35–45.

Schwander T, Libbrecht R, Keller L. 2014. Supergenes and complex phenotypes. *Curr Biol.* **24**(7):R288–R294.

Sedlazeck FJ, Rescheneder P, Von Haeseler A. 2013. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* **29**(21):2790–2791.

Servedio MR. 2009. The role of linkage disequilibrium in the evolution of premating isolation. *Heredity* **102**(1):51–56.

Sniegowski PD, Charlesworth B. 1994. Transposable element numbers in cosmopolitan inversions from a natural population of *Drosophila melanogaster*. *Genetics* **137**(3):815–827.

Staton SE, Bakken BH, Blackman BK, Chapman MA, Kane NC, Tang S, Ungerer MC, Knapp SJ, Rieseberg LH, Burke JM. 2012. The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements. *Plant J.* **72**(1):142–153.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**(3):585–595.

Tarasov A, Vilella AJ, Cuppen E, Nijman IL, Prins P. 2015. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**(12): 2032–2034.

Tian Z, Rizzon C, Du J, Zhu L, Bennetzen JL, Jackson SA, Gaut BS, Ma J. 2009. Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? *Genome Res.* **19**(12):2221–2230.

Tian Z, Zhao M, She M, Du J, Cannon SB, Liu X, Xu X, Qi X, Li MW, Lam HM, *et al.* 2012. Genome-wide characterization of nonreference transposons reveals evolutionary propensities of transposons in soybean. *Plant Cell* **24**(11):4422–4436.

Todesco M, Owens GL, Bercovich N, Légaré JS, Soudi S, Burge DO, Huang K, Ostevik KL, Drummond EBM, Imerovski I, *et al.* 2020. Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* **584**(7822):602–607.

Trickett AJ, Butlin RK. 1994. Recombination suppressors and the evolution of new species. *Heredity* **73**(4):339–345.

Tuttle EM, Bergland AO, Korody ML, Brewer MS, Newhouse DJ, Minx P, Stager M, Betuel A, Cheviron ZA, Warren WC, *et al.* 2016. Divergence and functional degradation of a sex chromosome-like supergene. *Curr Biol.* **26**(3):344–350

Wang J, Wurm Y, Nipitwattanaphon M, Riba-Grognuz O, Huang YC, Shoemaker D, Keller L. 2013. A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**(7434): 664–668.

Weir BS. 1996. *Genetic data analysis II*. Sunderland (MA): Sinauer Associates.

Yang L, Gaut BS. 2011. Factors that contribute to variation in evolutionary rate among Arabidopsis genes. *Arabidopsis* **28**(8):2359–2369.

Yang YY, Lin FJ, Chang HY. 2002. Comparison of recessive lethal accumulation in inversion-bearing and inversion-free chromosomes in *Drosophila*. *Zool Stud.* **41**(3):271–82.

Yang Y, Ma T, Wang Z, Lu Z, Li Y, Fu C, Chen X, Zhao M, Olson MS, Liu J. 2018. Genomic effects of population collapse in a critically endangered ironwood tree *Ostrya rehderiana*. *Nat Commun.* **9**(1):1–9.

Yoshida K, Ravinet M, Makino T, Toyoda A, Kokita T, Mori S, Kitano J. 2020. Accumulation of deleterious mutations in landlocked threespine stickleback populations. *Genome Biol Evol.* **12**(4): 479–492.

Zeng J, Zou YP, Bai JY, Zheng HS. 2002. Preparation of total DNA from "recalcitrant plant taxa". *Acta Bot Sin.* **44**(6):694–697.