Evolution, 2023, **77(4)**, 1117–1130
https://doi.org/10.1093/evolut/qpad027
Advance access publication 15 February 2023
Original Article

The rate of chromosomal inversion fixation in plant genomes is highly variable

Kaede Hirabayashi and Gregory L. Owens

Department of Biology, University of Victoria, Victoria, BC, Canada

Corresponding author: Department of Biology, University of Victoria, PO Box 1700, Station CSC, Victoria, BC, V8W 2Y2, Canada. Email: grego@uvic.ca

Abstract

Chromosomal inversions are theorized to play an important role in adaptation by preventing recombination, but testing this hypothesis requires an understanding of the rate of inversion fixation. Here, we use chromosome-level whole-genome assemblies for 32 genera of plants to ask how fast inversions accumulate and what factors affect this rate. We find that on average species accumulate 4–25 inversions per million generations, but this rate is highly variable, and we find no correlation between sequence divergence or repeat content and the number of inversions or the proportion of genome that was inverted and only a small correlation with chromosome size. We also find that inversion regions are depleted for genes and enriched for TEs compared to the genomic background. This suggests that idiosyncratic forces, like natural selection and demography, are controlling how fast inversions fix.

Keywords: inversion, structural variation, plant genome, transposable element, mutation rate, fixation rate

Introduction

The field of genomics has undergone a remarkable expansion in the last decade. With the rapid advances in long- and linked-read sequencing technologies, assembling a chromosome-resolved eukaryotic genome is no longer a fantasy (Pucker et al., 2022). While earlier draft genomes often covered the entire genome, they were in many small contigs which meant that genome structure was not resolved. Recent empirical work has highlighted that changes in genome structure can be critical for important evolutionary processes such as adaptation and speciation, and chromosome-resolved genome assemblies allow for these to be surveyed in an unbiased way for the first time (Mérot et al., 2020).

One type of structural variation, inversions, are particularly interesting because of their effect on meiosis and recombination. Initial investigations into inversions focused on underdominance effects (White, 1973). Due to how homologous alignment occurs during meiosis, in heterozygous individuals a recombination event within an inversion will lead to unbalanced gametes and a loss of fertility (Dobzhansky, 1933; White, 1978). This underdominance led to research into their role in reproductive isolation, as inversions in different orientation are often fixed between species (Rieseberg, 2001; Trickett & Butlin, 1994) and have been shown to directly cause hybrid sterility (reviewed in Zhang et al., 2021). It was also recognized that the recombination suppression abilities of inversions may link together favorable alleles and be relevant in the context of adaptation (Charlesworth & Charlesworth, 1973; Kirkpatrick & Barton, 2006; Ohta & Kojima, 1968; Sturtevant & Mather, 1938). For example, loci that are locally adaptive can benefit from inversions since they can be co-inherited as an adaptive multi-loci haplotype and protected from recombination with non-adaptive alleles (Kirkpatrick & Barton, 2006). Numerous non-model systems have been observed to have inversions with adaptive significance: sunflower (Todesco et al., 2020), monkeyflower (Lowry & Willis, 2010), Atlantic cod (Barth et al., 2017; Sodeland et al., 2022), stickleback (Jones et al., 2012), marine snail (Koch et al., 2022), fire ant (Wang et al., 2013; Purcell et al., 2014), honeybee (Wallberg et al., 2017), ruff (Küpper et al., 2016; Lamichhaney et al., 2016), and deer mice (Hager et al., 2022; Harringmeyer & Hoekstra, 2022) to name a few.

There are several molecular mechanisms of inversions, all of which are triggered by some form of DNA strand breaks during meiosis or in other situations (reviewed in detail by Burssed et al., 2022; Casals & Navarro, 2007). In non-allelic homologous recombination (NAHR), segments with a high degree of sequence similarity misalign during meiosis and recombine within the chromatid (intrachromatid) instead of between sister chromatids. This is known as ectopic recombination. When repeats are found in the opposite orientation, inversion of the region surrounded by the misaligned repeats can occur. This means that inverted repeats are necessary for inversions mutations with this mechanism. Non-homologous end-joining (NHEJ) inversions occur from random paired double-stranded breaks surrounding a segment of the genome, followed by 180° flipping of the detached region and a repair. This mechanism does not require any sequence homology around breakpoints and could occur during any stage of the cell cycle. The isochromatid mechanism is similar to NHEJ but requires staggered single-stranded breaks instead of double-stranded breaks (Ranz et al., 2007). The inversion through this mechanism would result in the creation of inverted sequence duplicates around the breakpoint

Received September 2, 2022; revisions received January 20, 2023; accepted February 9, 2023

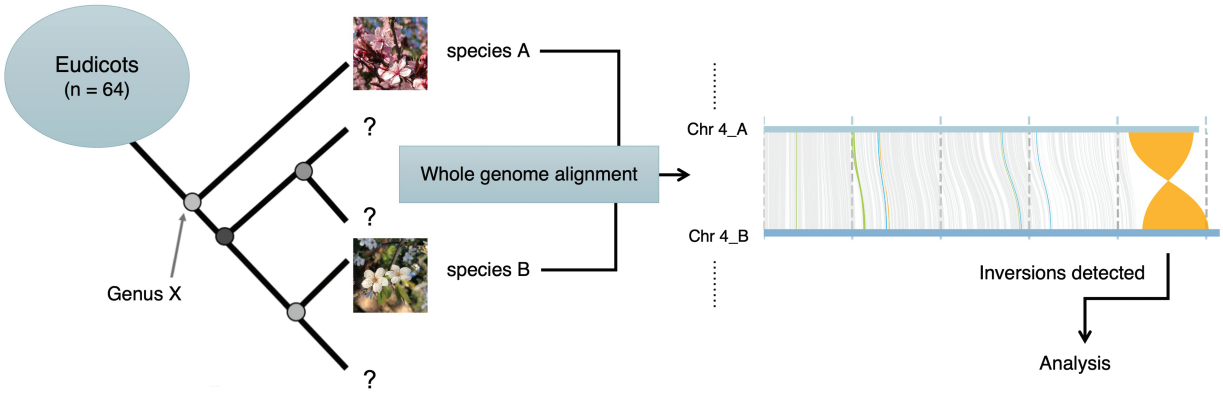

Figure 1. Overview of our comparative genomics framework. We aligned the whole genome for two different species in the same genus and identified inversions based on these alignments. The inversions identified were used in further analyses.

regions. Lastly, inversions that are especially small and found in a complex rearrangement pattern could arise from DNA replication machinery being disturbed and confused due to formation of DNA secondary structures (e.g., Cruciform, hairpin, triple helical DNA, quadruplex). The replication fork may be subjected to breakage or stalling and cause template switching event, called invasion. If the new position has switched orientations of strand (so say invasion occurred from template to lagging strand), then inversions of that segment would result. Additionally, large number of mobile elements (transposable elements or TEs) has been associated with recurrent chromosomal rearrangements, documented by the reuse of some of the breakpoints identified (Porubsky et al., 2022; Ranz et al., 2007). These regions could be prone to inversions and be the driving force of rearrangements.

Despite growing interest in chromosomal inversions for their adaptive potential and appreciation of their prevalence, there is a lack of knowledge on the rate of inversion fixation in plants. Most studies of the rate of chromosomal inversions have focused on *Drosophila* species (Bhutkar et al., 2008; Ranz et al., 2001). A literature review on plants has estimated the rate of inversion fixation to be around 15–30 inversions per million years based on the divergence time of a small number of species analyzed (Huang & Rieseberg, 2020). However, the method of inversion detection and divergence estimation cited in this review were variable and did not account for the size of the inversions. A recent theoretical work proposed that different mechanisms driving inversion fixation (e.g., local adaptation, drift) will lead to different size distributions of inversions (Connallon & Olito, 2022). This means that the size distribution of inversions tells us something about the evolutionary forces driving their fixation. Altogether, further studies with robust and consistent approaches are needed.

Leveraging the resource of publicly available genome assemblies, we conducted a comparative study to investigate how chromosomal inversions accumulate in eudicots. We expected to observe that inversions would accumulate over time at a consistent rate; therefore, more divergent species pairs should harbour more inversions. We also predicted that if there is more space in the genome for chromosomal rearrangement (i.e., larger genome size), would lead to more inversions. Alternatively, if fixation of inversions was dominated by selection, which can be highly variable, then rates will be idiosyncratic. We further investigated the genomic context of inversions, which gives us insight into their functional significance. Lastly, we asked if inversions were surrounded by inverted repeats, which tells about the relative rate of different inversion mechanisms.

Methodology

Data collection

The schematics for the overview of our study is found in Figure 1. Publicly available genomes within Eudicots were searched through the published plant genomes database (last screened in November 2021; https://www.plabipd.de/plant_genomes_pa.ep). We selected genera with at least two sequenced species, but did not specifically target sister species. Then individual genomes were screened for all of the following criteria: (1) same ploidy (diploid), (2) chromosome-resolved assembly, (3) same number of chromosomes, (4) de novo assembled and (5) valid NCBI BioProject number or similar. Assemblies published before 2018 were carefully assessed for quality. Those that used guidance from genetic mapping or reference assembly were excluded because such assemblies rely on alignment to the older reference genomes, which tend to be less accurate than current techniques. If the assemblies did not have an associated publication or did not provide adequate information on how the genome was sequenced, they were excluded from the study. For one species per genus, we downloaded coding sequences and gene annotations. In total, 64 chromosome-resolved genomes from 32 genera were included (Supplementary Tables 1 and 2). To explore biological factors that may affect the fixation of inversions, we surveyed the literature to characterize for each species pair the range overlap, evidence of hybridization, reproductive self-compatibility, domestication history, and generation time. Range overlap was categorized as a binary trait (yes/no) based on geographic distribution overlap. Evidence of hybridization was categorized into three groups: weak (no evidence of hybridization or evidence of sterile hybrid attempt), strong (evidence of hybridization/introgression in nature or lab or genes), unclear (lack of evidence to support either weak or strong). Self-compatibility was assessed by three categories: selfing (mainly reproduce by self-pollinating), mixed (can self or mate), outcrossing (cannot produce viable offspring from selfing). Domestication history was categorized by whether it has been domesticated or not, and the generation time was divided into either annuals or perennials. Two genome pairs were missing the gene annotation file and thus excluded from analyses involving gene locations. All species used and recorded information are in Supplementary Table 1.

Structural analysis

The collected 32 genome pairs were analyzed for structural variant detection using Synteny and Rearrangement Identifier v1.5 (SyRI) (Goel et al., 2019). To do this, first, genomes were assigned either reference or query based on assembly statistics (N50 and number of contigs) and availability of gene annotation file. "Reference" status was assigned to those with good contiguity and gene annotation file available. Next, only chromosomal sequences were extracted from genome assemblies and scaffolds were removed. Therefore, from here onward, the "total genome length" in our study refers to the total length of assembly that has been assigned to chromosomes. For *Luffa* and *Acer* genomes, the chromosomes in the query genome that had half synteny with one chromosome and the other half with another chromosome in the reference genome were removed from SyRI analysis. This is due to SyRI being unable to perform comparison between divergent genomes lacking one-to-one synteny. Specifically, in *Luffa*: CM029395, CM029402 (reference), CM022716, CM022722 (query) were excluded, in *Acer*: chr3, chr4 (reference), CM017761, CM017766 (query) were excluded. When running SyRI, default parameters were used and whole-genome base-to-base alignment was performed by minimap2 v2.17-r974-dirty (Li, 2018, 2021). When necessary, chromosomes were reverse complemented using Samtools v1.12 (Li et al., 2009) prior to SyRI. All the following statistical analyses were performed in R (tidyverse) (Wickham et al., 2019).

For our analyses, we identified inversions that differ in orientation between closely related species, but we are not able to determine which orientation is derived or ancestral. An inversion region in the reference species is a region that is in the opposite orientation in the query species, which is because either the reference or query species inverted. If we assume that the rate of inversion fixation is roughly equal between the two species—which may not be true—then roughly half of the inversion regions contain a derived inversion in the reference species and the other half contain the ancestral state, but have inverted in the query species.

Transposable element annotation

Transposable elements (TEs) were detected and annotated using Extensive De novo TE Annotator v2.0.0 (EDTA) (Ou et al., 2019) pipeline on the 30 reference genomes with gene annotations available. TEs are defined as repetitive, mobile elements and EDTA incorporates multiple individual repeat finders to annotate them (Ou et al., 2019); thus, in this work we are using TEs as a general proxy for repetitive elements. Due to resource limitation, the pipeline was performed chromosome-by-chromosome. The whole-genome fasta file was first divided into individual chromosome files. Then, the EDTA pipeline was performed separately on each chromosome as follows and the resulting TE library was later combined. In brief, candidate TE sequences were de novo identified using LTR-Finder (Xu & Wang, 2007; Ou & Jiang, 2019), LTRharvest (Ellinghaus et al., 2020), LTR_retriever (Ou & Jiang, 2018), generic repeat finder (Shi & Liang, 2019), and HelitronScanner (Xiong et al., 2014), respectively. Once individual TEs were identified, candidates were filtered (Zhang et al., 2019) and further refined by RepeatModeler2 (Flynn et al., 2020) according to EDTA pipeline default parameters. Finally, TE-free coding sequences retrieved from gene annotation files (feature name = "CDS") were aligned to the repeat library, and those overlapped with coding sequence (CDS) were excluded from the identified TE candidates. Total genomic TE content (%) was calculated by the total length of detected TE without overlaps divided by the total genome length.

Species divergence calculation and the rate of inversion accumulation

The principal advantage of SyRI over basic genome aligners is that SyRI identifies larger regions with consistent synteny. To do this, SyRI takes alignment blocks identified by the initial aligner, in this case minimap2, and identifies larger regions consisting of multiple consecutive one-to-one alignment blocks. We focused our analysis on regions which each represent a single structural variant (or contiguous syntenic region without any structural variation). While SyRI identifies several types of structural variation, we focused our analysis on inversions and syntenic regions only. To minimize noise from small amounts of data, detected regions smaller than 1 kbp were excluded from the dataset. To calculate the sequence divergence for each region (X), we used Equation 1. From SyRI output, for each block (i) in a region, we measured percent identity (xi) and reference length (Li).

$$X = 1 - \frac{\sum_{i=1}^{n} L_i x_i}{\sum_{i=1}^{n} L_i} \quad (1)$$

Previous work has found inversions haplotypes older than speciation events, so to explore the possibility of ancient segregating inversions differing between species we compared the sequence divergence of syntenic and inversion regions (Todesco et al., 2020). We first used two-sided *F*-statistics to test if variance in sequence divergence differed between inversion and syntenic regions for each genus. We then tested for differences in the level of sequence divergence using unpaired *t*-tests. We intend this as an exploration of general patterns, and not to identify exceptional species, so we did not correct these significance values for multiple testing.

To calculate the average sequence identity between species, we used the Equation 1 but, in this case, combining all syntenic blocks. With this formula, sequence divergence score of 0 indicates identical species, and the value increases as the genomes become more different. To convert sequence divergence into divergence time, we used the genome-wide substitution rate estimated from *Arabidopsis thaliana* (Exposito-Alonso et al., 2018), which estimated 2–5 × 10^{-9} substitutions per generation. Based on this rate, and that the divergence rate is equal for each species in the pair, each percentage point divergence represents a common ancestor 1–2.5 million generations ago. This estimate should be used with caution as substitution rates are known to vary and divergence between close relatives is affected by standing variation (Ho et al., 2011).

We next aimed to estimate the rate of accumulation of inversions for each genus. To do this, we divided the number of inversions in a size category by the average amount of syntenic sequence divergence in that genus. To convert this into generations, we divided that value by 2 × 10^{-9} and 5 × 10^{-9} to represent the range of possible mutation rates. When calculating rates, we excluded inversions with sizes < 1 kbp because SyRI is less reliable at detecting these small inversions.

To test if there is a consistent rate of inversion accumulation, we used a linear model testing the effect of sequence divergence on inversion number or proportion. Similarly, we tested for the

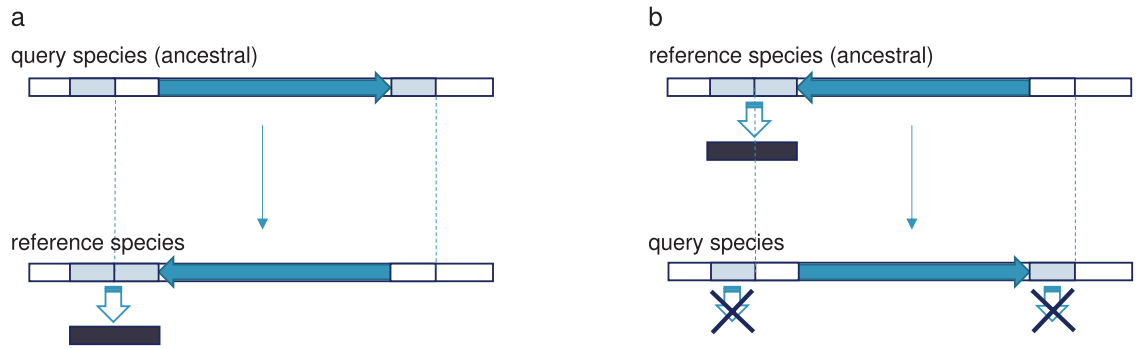

Figure 2. Two different interpretations of inversion consequences. (A) An inversion creates a new gene in the reference species. (B) An inversion disrupts an ancestral gene in the query species.

effect of genome size, chromosome size, genetic map length, and proportion of the genome in TEs on the number and proportion of inversions. We were also interested in whether features of the species, or the genome assembly affected the amount of inversions. Specifically, we tested the effect of generation time (annual vs. perennial), domestication status (domesticated vs. non-domesticated), evidence of hybridization (weak vs. strong), reproductive strategy (outcrossing vs. mixed mating vs selfing), assembler category (accurate vs. fast vs. short-read based), sequencing platform (Oxford Nanopore vs. PacBio vs. short-read) and physical mapping which include Hi-C and optical mapping (present vs. absent). For each of these, we used a one-way ANOVA. We did not include sequence divergence as a covariate in this test because it was not a significant predictor in our earlier analyses.

Quantifying CDS and TE in inverted regions vs. syntenic regions

We were interested if inversions contained different proportions of genomic elements. Bedtools v2.30.0 (Quinlan & Hall, 2010) intersect function was used to analyze how many features and base pairs of coding sequence (CDS; as identified by annotations) or TEs (as identified by EDTA) were found to overlap with the inverted regions in the genome. Inversion regions identified by SyRI were extracted using custom perl script and reorganized into a bed format using bash shell commands. The gene annotation file and TE library in gff3 were reformatted into bed format using gff2bed. Prior to using bedtools intersect, CDS and TE bed files were edited by bedtools merge function to concatenate overlapping features into a single feature, avoiding overrepresentation due to isoforms of the same gene sometimes present in the gene annotation file. The validity of data was confirmed by ensuring the proportion of CDS or TE per inverted region no larger than 1. In addition to inverted regions, the process was repeated with syntenic regions identified by SyRI. We used paired *t*-tests to ask if inversions were enriched for CDS or TEs when compared with syntenic regions. The total proportion of CDS or TE for all inversions or all syntenic regions for a species pair was used as a single datapoint, which means our unit of replication is species pair, not individual inversions.

Quantifying CDS and TE in inversion breakpoint regions

We were also interested in whether genomic elements differ at the breakpoints of inversions compared to the rest of the genome. Quantification of CDS/TE in inversion breakpoints were performed using Bedtools (Quinlan & Hall, 2010) intersect function. The number of CDS and TE in the inversion breakpoint regions were extracted as follows. First, inversion breakpoint regions (defined as 4 kbp regions surrounding the inversion; 4 kbp upstream of the inversion start point and 4 kbp downstream of the inversion end point) were identified. Five inversion breakpoint regions were within 4 kbp of the end of the chromosome, and were excluded from the dataset. A total of 11,225 breakpoint regions from 30 species pairs were used for analysis. To infer whether breakpoints are enriched with CDS/TE, baseline numbers for the whole genome are necessary. To determine the baseline number of CDS/TE in the genome, the total length of CDS/TE was computed from the merged bed file from previous step. Then, the genomic CDS/TE proportion was calculated by the total length of CDS/TE divided by the length of the reference genome. The resulting average proportion of CDS/TE in the 4 kbp breakpoint regions was compared to the genomic CDS/TE proportion.

Quantifying frequency of gene occurrence at inversion breakpoints

Chromosomal inversions can disrupt gene sequence if an inversion breakpoint occurs within the gene itself. Since we cannot identify whether the derived orientation occurred in the reference or query genome, we are focusing on genes identified in the reference species. If an inversion breakpoint falls within a gene sequence it has two possible meanings (Figure 2):

1. The orientation is derived in the reference species. The inversion either created the gene or modified its coding sequence.
2. The orientation is ancestral in the reference species. The inversion disrupted or modified the coding sequence of the gene in the query species.

Both models assume that the genes are largely shared between the reference and query species. For example, if a new gene appears only in the reference species genome, it cannot be disrupted by an inversion in the query genome.

We were interested in how often inversion breakpoints fell within genes. We used Bedtools intersect to count the number of breakpoints that occurred within the coding sequence or introns of a gene. To identify if breakpoints are less likely to occur within genes, we selected *n* random positions (*n* adjusted to 1,000 per chromosome for each genome) from

the syntenic portion of the genome using Bedtools (Quinlan & Hall, 2010) random function and used this as a baseline. Additionally, we also calculated the proportion of genes that contained at least one inversion breakpoint to identify how often inversions may be disrupting genes.

Breakpoints regions sequence similarity analysis

Some models for inversion mutations require or create segmental duplications at the breakpoints of the inversion. To explore this idea, the 10 kbp regions upstream and downstream of breakpoints described above were aligned to each other using BLAST v2.12.0 (Zhang et al., 2000). In some cases, breakpoints were within 10 kbp of the ends of chromosomes, resulting in smaller analysis regions. For these, we required there to be at least 2 kbp of sequence both upstream and downstream of the inversion. In cases where one reference sequence (breakpoint start region) had multiple BLAST hits to the query sequence (breakpoint end region), only the longest aligned hit was retained. Additionally, to search for potential repeats within the breakpoint regions themselves, we divided each 10 kbp breakpoint region into ten 1 kbp pieces and aligned against each other using BLAST, with default parameters. Self-alignment, meaning a 1 kbp piece aligning to itself, was removed in all cases. For each 10 kbp region, we characterized it as being repetitive if at least one of the 1 kbp pieces had a blast hit to a different 1 kbp piece within the region.

Results

Inversion accumulation is idiosyncratic

We identified a total of 6,140 inversions across our 32 comparisons, 5,298 of which were larger than 1,000 bp. In general, inversions tended to be small: 45.0% (2,766/6,140) of inversions were less than 10 kbp, 46.8% (2,873/6,140) were between 10 kbp and 1 Mbp and 8.2% (501/6,140) were greater than 1 Mbp (Figure 3A). The number of inversions varied between comparisons (Figure 3B), along with the proportion of the total genome length in inversions, with the lowest being 1.3% to the highest being 37.4% (Figure 3C). For each of our genera, we observed whether sequence identity in inversion regions differed from sequence identity in syntenic regions and found a statistically significant difference in variance in 14 out of 32 comparisons; however, the mean value of sequence identities in syntenic and inverted regions were mostly similar, except four genera (*Corymbia, Gossypium, Malus, and Raphanus*; Supplementary Figure 1, Supplementary Table 3). In syntenic regions, our species pair had on average 3% to 5% sequence divergence, supporting their close relationship (Supplementary Figures 2 and 3). Inside inversions, similar sequence divergence was observed on average.

We were interested in determining factors that affected the number of inversions and the rate of inversion accumulation. To investigate the evolutionary forces driving inversion accumulation, we plotted their size distributions by genus (Supplementary Figure 4). In a majority of genera, short inversions were most common (e.g., *Populus and Prunus*), but in some cases, we saw a peak at larger sizes (e.g., *Corymbia*) or two different peaks (e.g., *Cucumis*) (Figure 3D). Since sequence divergence is expected to increase over time and can be a proxy for divergence time, we expected a general positive correlation between the number of inversions and species divergence (with many potential caveats) (Drummond et al., 2006; Ho et al., 2011). We used a linear model to ask if species pair with higher sequence divergence had more inversions and found no significant relationship (p = .6) (Figure 3E). This was also not significant if we instead used the proportion of the genome that is inverted, rather than the number of inversions (p = .5) (Supplementary Figure 5). All else being equal, we expected that larger genomes should have more inversions. When tested using the entire genome length, we saw no significant relationship (p = .96) (Figure 3F), but when treating each chromosome separately, we found a slight positive relationship (n = 405; F value = 12.78; p = .00039***; linear model R^2 = 0.029; Figure 3G). We tested for a relationship between recombination rate (linkage map length/haploid genome size) and inversion number/proportion for those species pairs with physical maps available, but no significant relationship was observed (n = 27; p = .46, .215; Supplementary Figure 5), Since some mechanisms of genomic inversion require repetitive regions, we expected that species with a higher proportion of TEs would have more inversions, but again this relationship was not significant (p = .3) (Figure 3H).

Given that inversions are not accumulating in a clock-like manner, we used a one-way ANOVA to explore other factors controlling the number and genomic proportion of inversions. We found that the assembly method used, sequencing technology, whether species are annual or perennial, or whether domesticated or wild did not affect the number of inversions significantly (Table 1). But, evidence of hybridization with limited dataset (n = 11) did show a statistical significance, where and weakly hybridizing pairs have more inversions than strongly hybridizing pairs (p = .0143*), and reproductive strategy also did when it was calculated with the proportion of inverted genome instead of the number (p = .031*, n = 32). We found that selfing species had the highest proportion of the genome inverted while mixed and outcrossing species had a similar proportion (Supplementary Figure 6).

Although the rate of inversion accumulation relative to sequence divergence is not consistent in our dataset, the range of possible values may be a useful baseline for other studies. We present a fixation rate by dividing the number of inversions by the sequence divergence for different size categories of inversions (Figure 4). We find the smaller inversions are more common, with the exception of the smallest category (<1 kbp) which was omitted from the results. Further, based on an estimate of 1–2.5 million generations of divergence per 1% sequence divergence, we present a range of estimates of inversion accumulation by size (Table 2).

Inversions tend to occur in less functional regions

If inversions cause the deleterious disruption of gene sequence or expression, then we expect fixed inversions to contain less gene sequence when compared with syntenic regions. Indeed, we found this was the case (Figure 5A; t = –13.97, df = 29, p = 2.08e–14 ***). We also found the reverse pattern for TEs, where inversions were enriched compared to syntenic regions (Figure 5B; t = –3.93, df = 29, p = 4.77e–4 ***).

We next asked whether inversion breakpoints differed from the genome background. Although SyRI identifies breakpoints, limitations in the alignment of repetitive regions means that these may not be the exact breakpoint, therefore we selected 4 kbp regions downstream/upstream of identified breakpoint positions. These regions were compared against

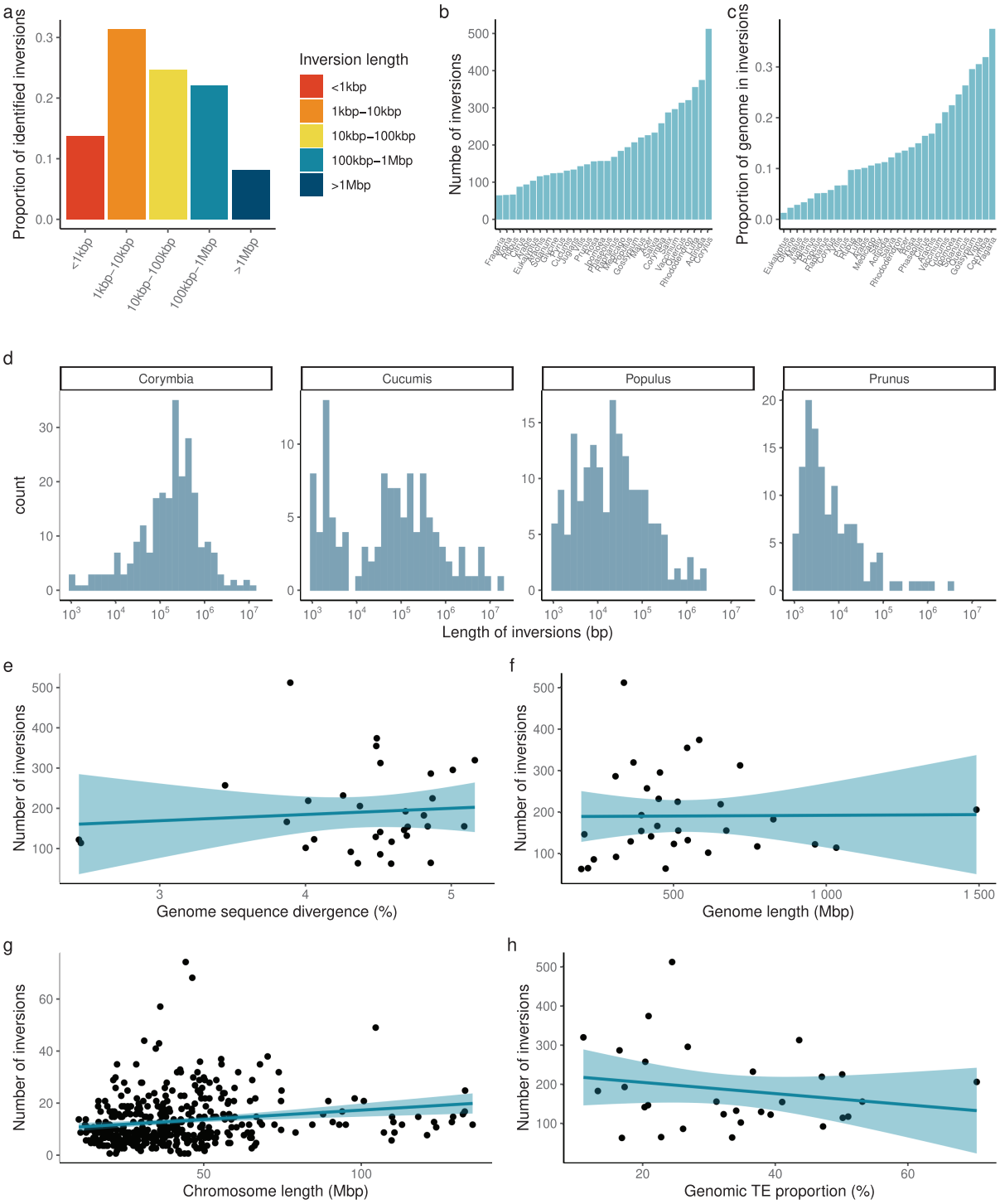

Figure 3. Summary of inversions between paired eudicot genomes in the same genus. (A) Size distribution of inversions by inversion length. Note that the number of small inversions (< 1 kbp) was probably underestimated due to detection sensitivity. (B) Number of inversions and (C) proportion of reference genome in inversion, for 32 species pairs used in the study. (D) Inversion length distribution by genus, showing representative patterns of predominantly large (Corymbia), predominant small (Populus and Prunus), or combination of the two (Cucumis). Number of inversions plotted against (E) percent sequence divergence between the 32 species pair (F value = 0.264, p = .611), (F) genome length of the reference species genome (n = 32; F value = 0.002, p = .961), (H) genomic TE proportion calculated by total length of TE/reference genome length (n = 30; F value = 1.139, p = .295). Each datapoint represents a paired species within the same genus. (G) Number of inversions per chromosome plotted against its length (n = 405; F value = 12.78; p = .00039 ***; linear model R^2 = 0.029).

the genomic CDS and TE content in reference genome. We found that there was a lower proportion of CDS in breakpoint regions (Figure 5C; p = .000107 ***), but no consistent trend for TEs (Figure 5D; p = .142).

To infer whether inversions are disrupting gene sequence, or alternatively involved in the creation of new genes, we counted the number inversion breakpoints (start/end) that fell within a gene (Figure 6A). Strikingly high occurrence of

Table 1. One-way ANOVA results.

Inversion measure	Number of inversions		Proportion of genome inverted	
	F value	Pr(>*F*)	*F* value	Pr(>*F*)
Annual vs. perennial	0.851	.437	0.313	.733
Domestication status	0.345	.711	0.23	.796
Evidence of hybridization[+]	9.157	.0143*	0.002	.970
Reproductive strategy	1.73	.163	2.942	.031 *
Assembler category (accurate, fast, short-read based)	1.223	.324	0.73	.579
Sequencing platform (ONT vs PacBio vs short-read only)	0.772	.52	1.504	.235
Physical mapping (Y/N)	1.679	.204	0.023	.977

Note. Tests performed for the number of inversions and proportion of reference genome in inverted orientation against the following seven factors: annual vs. perennial, domestication status, evidence of hybridization (weak/strong)[+], reproductive strategy (selfing/mixed/outcrossing), genome assembler category (accurate/fast or not resource intensive/short-read based), long-read sequencing platform (Oxford Nanopore/Pacific Bioscience/none or short-read only), whether physical mapping (i.e., Hi-C, BioNano optical map, long-range Chicago) is performed or not. The reproductive strategy and assembly methods were assigned to each genome assembly (reference and query) separately, then the paired category was used in the stats (n = 32).
* Statistically significant (p < .05) .
[+] Species pairs filtered by those with geographical overlap and non-domesticated status (N = 11).

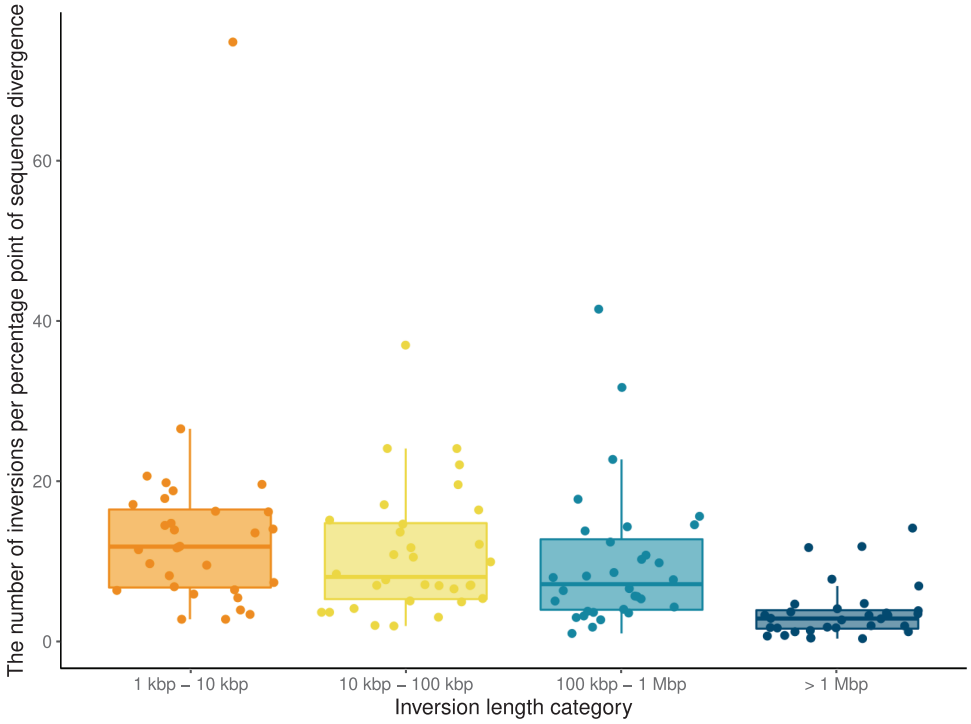

Figure 4. Rate of inversion fixation by inversion size. Rate of inversion occurrence calculated as number of inversions per percentage point of sequence divergence by different size categories (n = 32 for each category).

inversion breakpoints at a gene was observed in several genera (Figure 6A). Notably *Vigna, Ipomoea, Medicago, Solanum, Cucumis, Salix, and Phaseolus* had > 50% of breakpoints occurring within a gene. Despite this, the relatively small number of inversions compared to genes means that < 1% of genes contained an inversion breakpoint (Figure 6B).

Inversions are most often not surrounded by repeats

Lastly, to test the molecular mechanism of inversions, we specifically looked for duplicates or inverted repeats surrounding the inversion breakpoints. If inversions were created by ectopic recombination or the recombination between the regions on the same chromosome arm aligning in opposite orientation, then we would expect to see inverted repeats or segmental duplications around breakpoints. If inversions were resulted from NHEJ, then no trace of duplicates or inverted repeats would be expected, unless it is the isochromatid staggered breaks that would result in partial duplications. Among the 5,626 tested 10 kbp breakpoint regions pair, 1,160 (20.6%) regions pair resulted in at least one BLAST hit. Of these, 592 were in the forward orientation and 568 were in inverted orientation (Figure 7A). To see if the breakpoint region itself is repetitive, we broke down the 10 kbp regions into 1 kbp pieces and aligned them to each other. The total of 122,800 1kbp windows were created—out

Table 2. Rate of inversion accumulation by inversion length with respect to species divergence time.

Inversion length category	Mean number of inversions per 1% seq divergence (25th – 50th – 75th percentile)			Mean number of inversions per estimated species divergence time (million generations ago, 25th – 50th – 75th percentile)					
Conversion				1 mga/1% seq divergence			2.5 mga/1% seq divergence		
1 kbp—10 kbp	6.8	11.9	16.5	3.4	6.0	8.4	1.4	2.4	3.3
10 kbp—100 kbp	5.3	8.1	14.8	2.7	4.0	7.4	1.1	1.6	3.0
100 kbp—1 Mbp	4.0	7.2	12.8	2.0	3.6	6.4	0.8	1.4	2.6
> 1Mbp	1.7	2.9	4.0	0.8	1.5	3.0	0.3	0.6	0.8
All	17.8	30.1	48.1	8.9	15.1	25.2	3.6	6.0	9.7

Note. Species divergence time was estimated using *Arabidopsis thaliana* substitution rate (Exposito-Alonso et al., 2018).

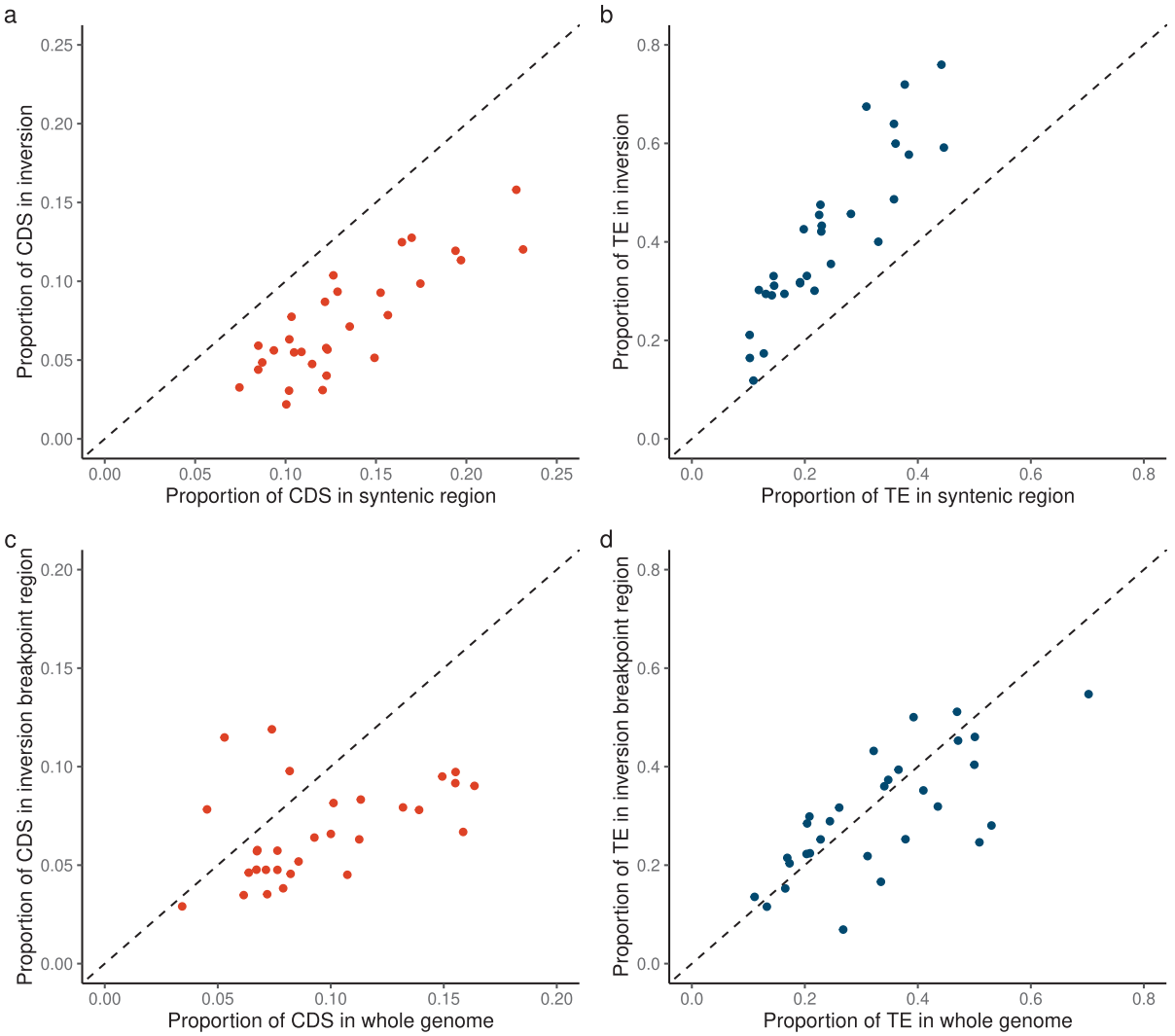

Figure 5. Genomic context of inversions and surrounding regions. The mean proportion of (A) coding sequence and (B) transposable elements inside syntenic region plotted against those inside inversions (n = 30; p = 2.08e−14***, 4.77e−4***, respectively). The mean proportion of (C) coding sequence and (D) transposable elements in 4 kbp breakpoint regions compared to the genomic CDS and TE proportion in the genome (n = 30; p = .000107 ***, .142, respectively). Each datapoint represents a species pair, and the dashed line shows the equal proportions of CDS/TE between the two compared region types.

of which 122,254 windows had valid coordinates on chromosomes—and tested for alignment by BLAST. We found that 2,379 (19.4%) of them showed at least one alignment to other non-overlapping 1 kbp windows within that breakpoint region, and most of these matches were aligned in the forward direction (~ 90%) (Figure 7B).

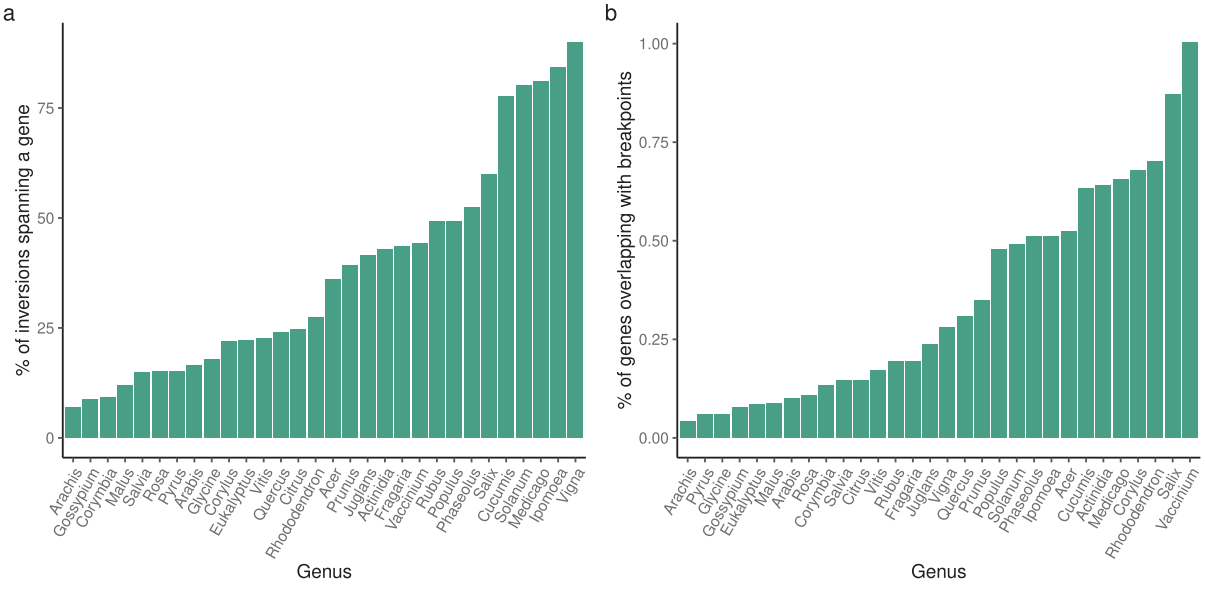

Figure 6. Inversion breakpoints and genes overlap. (A) The percentage of inversion breakpoints that overlapped with a gene by genus. (B) The percent of genes in reference genome that contained an inversion breakpoint.

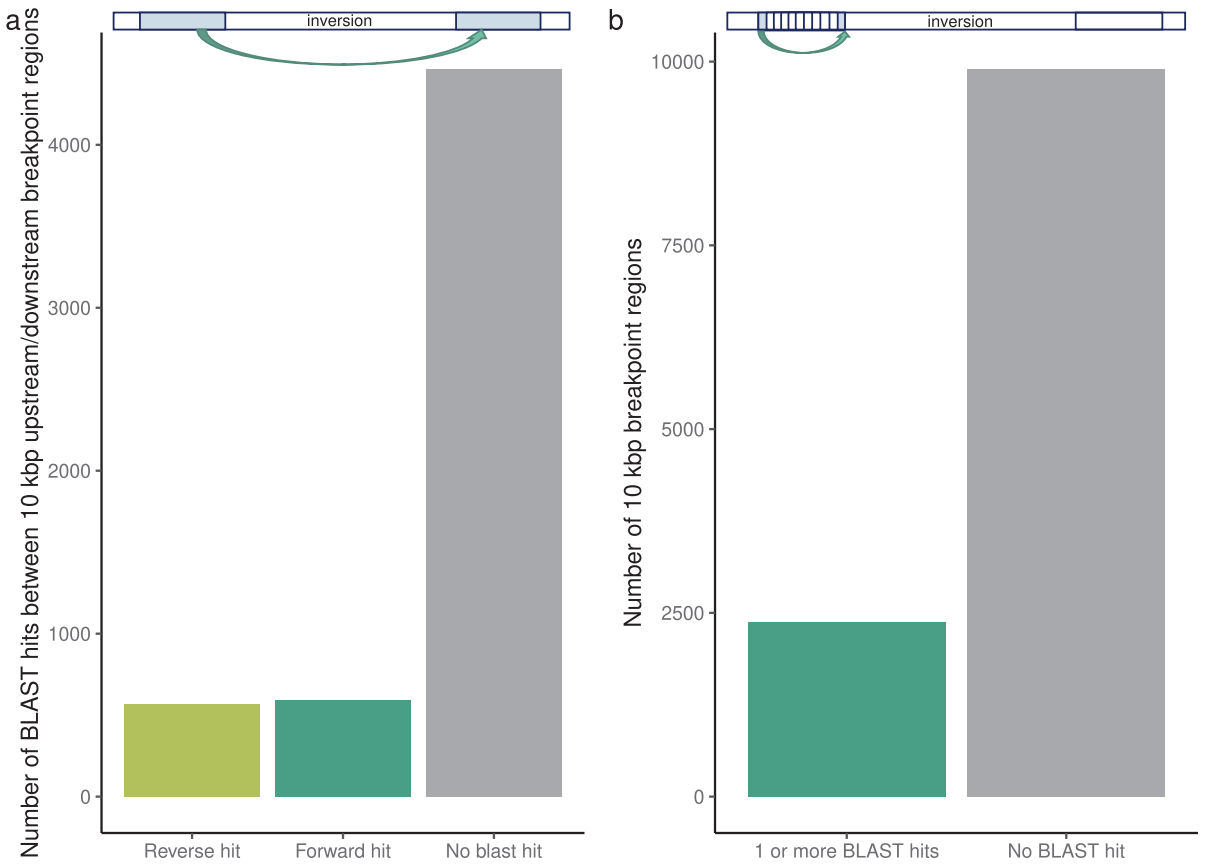

Figure 7. Rare presence of duplicates or inverted repeats surrounding breakpoints, but breakpoint regions can be repetitive. Summary of alignment between (A) the 10 kbp window surrounding the 5,626 inversions (i.e., breakpoint regions) and (B) among the 1 kbp ×10 windows within each side of inversions aligned to each other detected using BLAST.

Discussion

The rate of inversion accumulation

Several high-profile studies have highlighted the adaptive importance of inversions. For example, the prairie sunflower (*Helianthus petiolaris*) has adapted to the dune environment while continuing to exchange genes with non-dune neighbouring populations (Huang et al., 2020). Recent studies have shown that the alleles controlling dune adaptation are found within large inversion regions suggesting the possibility that the recombination suppression of inversions is playing a role in maintaining adaptations (Huang et al., 2020; Todesco et al., 2020). One challenge to this hypothesis is that the number and size of inversions is not known so null models are challenging to parameterize. For example, if inversions are ubiquitous across the genome, we expect adaptive variation to be found in them regardless of other features of inversions.

To measure the rate of inversion fixation, we aligned chromosome-level genome assemblies for two species within the same genus. These comparisons explore how fast inversions fix in the genome by comparing the sequence divergence between species with the number of inversion differences. We are treating the rate as a baseline in comparison to systems like *H. petiolaris* where there is evidence that inversions are positively selected, but we recognize that the inversions themselves may not be neutral and that each species has its own demographic and adaptive history. Additionally, our method is based on one representative genome per species and so inversions detected in our study represent both fixed differences between species as well as some segregating inversions.

Surprisingly, we find no correlation between sequence divergence, which is a proxy for coalescence time, and the number of inversions separating species (Figure 3E), suggesting that the rate of accumulation is dependent on factors not consistent between genera. Based on our data, the middle 50% of comparisons accumulated 17–48 inversions per 1% sequence divergence. Converting from sequence divergence to divergence time is fraught because it is affected by generation time, the mutation rate, and other factors which are likely to vary between our genera but it is helpful for scaling expectations (Ho et al., 2011; Wang et al., 2019). Based on substitution rates from *Arabidopsis*, we estimate 4–25 inversions per million generations which is within the range estimated by Huang et al. from a more limited dataset (Huang & Rieseberg, 2020).

Although we present a range of inversion accumulation rates, our primary finding is that this value is not consistent. There are several reasons—both technical and biological—that could explain the variation. Our inversion counts are based on whole-genome alignments which is much better at capturing small inversions than genetic mapping but are not without error. Artifactual inversions can be introduced during assembly or scaffolding, and idiosyncrasies in sequencing quality or depth and methodology may mean that different genomes have different rates of error. We looked for the effect of sequencing and assembly method and failed to find a significant relationship, but our test is underpowered for the number of categories based on our sample size. There are also limitations in our ability to detect inversions based on alignments. Overall, we find that smaller inversions are more common, except for the smallest category, which we excluded from analyses. These were excluded to prevent detectability from biasing our relative rate calculations.

Different taxa have diverse genomic properties. Some may have inherently high recombination rate or epigenetic patterns that allow them to harbour chromosomal rearrangements more easily (Henderson, 2012; Lloyd, 2022). Each species also has its own demographic and selection history. Since efforts to sequence plant genomes have focused on accessible and economically beneficial species, about half of the species used in this analysis are domesticated crops, which have likely undergone greater bottlenecks and distinct selection pressures compared to wild species. This could lead to greater fixation of deleterious inversions similar to mutational load seen in some domestic species (Bosse et al., 2019).

The rate of inversion fixation is expected to be related to the rate of inversion mutation. Logically, longer genomes should have more opportunity for double-stranded breaks and therefore more inversion mutations. When considering each genome as a replicate we do not see a significant relationship, although there is a relatively small but significant relationship when treating chromosomes as replicates (Figure 3G). Higher repeat content is also often associated with chromosomal rearrangements, and for some mechanisms of inversion generation, repeats are necessary. We therefore expected that genomes with more transposable elements would have more inversions, but again we do not see a significant relationship (Figure 3H). We observed a clear linear relationship between the genome size and TE content (data not shown) as expected, but the occurrence of inversions seemed to be correlated with neither. Since neutral genomic features seem to play a relatively small role, we suggest that selection and demography are key in inversion fixation.

Natural selection and inversions

A consistent rate of inversion accumulation makes an underlying assumption that the inversions themselves are neutral. If inversions are primarily fixed through positive selection, they likely represent a small fraction of total inversion mutations, and their count would be highly dependent on the distribution of fitness effects of new inversions. If inversions are instead primarily deleterious, then their fixation is dependent on the amount of genetic variation, the amount of selection on heterozygotes and homozygotes as well as the population structure of the species (Lande, 1984). In both cases, the fixation rate is going to vary due to species specific effects (e.g., demographic history) as well as inversion specific effects (e.g., inversion selection coefficient) which lead to inconsistent fixation rates.

Depending on the type of selection inversions are under, we expect different size distributions of inversions (Connallon & Olito, 2022). For inversions that are neutral, underdominant or beneficial, the size distribution should be biased toward short lengths. In contrast, locally adaptive inversions are expected to have predominantly intermediate sizes. When analyzed all together, our data shows that inversions are predominantly short, but individual genera show different patterns (Supplementary Figure 4). In some cases, we see distinct peaks of inversion number at intermediate sizes (such as *Corymbia and Quercus*) supporting local adaptation driving inversion accumulation. We also find genera with two different peaks of inversion sizes, suggesting that inversion accumulation is likely due to multiple forces. Unfortunately, we are unable to distinguish between neutral, underdominant and beneficial inversion scenarios as the distribution is

dependent on effective population size and the rate of deleterious mutations, factors which are unknown for our systems.

Another factor suggesting a role for selection in inversion fixation is the effect of mating strategy. In primarily selfing species, underdominant inversions are more likely to fix because of reduced heterozygosity (Charlesworth, 1992; Hedrick, 1981; Hoffmann & Rieseberg, 2008). We see this play out in our ANOVA of mating system which showed the highest amount of the genome in inversions for comparisons involving selfing species (Table 1; Supplementary Figure 6). This idea was inspired by the observation of unusually rapid speciation of selfing plants from sympatric outcrossing species with no potential geographical isolation history, minimal genetic/allelic differentiation but chromosomal differentiation (Gottlieb, 1973; Lewis & Raven, 1958). While theoretical support is strong (Charlesworth, 1992; Coyne & Orr, 2004), empirical data has been scarce and occasionally somewhat contradictory (Hoffmann & Rieseberg, 2008; Martin & Willis, 2010). Our result may serve as a starting point to encourage more empirical studies with modern technologies to explore the predominance of this theory.

Inversions can play a role in speciation by containing reproductive isolation alleles or by causing reproductive isolation itself through underdominance (Rieseberg, 2001). We explored this in our dataset by asking if evidence of hybridization affected the number or size of inversions. That being said, our species comparisons are not necessarily sister species or geographically close so they are not appropriate species to test models of speciation. When species pairs were filtered by non-domesticated plants pairs and potentially overlapping in geographical habitat to fit into this biological limit for hybridization, we saw that there tended to be more inversions in weakly hybridizing species pairs than those with strong evidence of hybridization, which might support the idea that inversions do somewhat play a role in reproductive isolation. However, our sample size is quite small (n = 11), so this should be interpreted with caution (Supplementary Figure 6). More data from known sister species and non-domesticated plants need to be incorporated to test the relationship between inversion and barriers to hybridization.

The location of inversions supports the hypothesis that they are often deleterious. We found that inversions and the breakpoint regions were depleted for coding sequence among the 32 genera (Figure 5A and C). While we cannot eliminate the possibility that this reflects underlying biases in the inversion mutation rate or the reduced recombination effect after-the-fact, we think it is more likely to represent selection against inversions that disrupt genes or gene expression. Similar to how newly inserted TEs tend to get purged if they were inserted in genic part of the genome (Quadrana et al., 2016), newly formed inversions that directly break the coding sequence should be negatively selected. Nevertheless, the relatively high proportion of inversions spanning a gene in some genera was unexpected, raising the question that inversions could also sometimes involve creation of new genes (e.g., Korneev & O'Shea, 2002). But, the relatively small number of inversions compared to genes means that < 1 % of genes were affected by inversions in our dataset (Figure 6), which suggest inversions are responsible for only modest differences in gene content between our species. Even without disrupting a coding sequence itself, inversions that capture several genes inside themselves would shuffle gene order, potentially leading to disrupted gene regulation and expression pattern. In *Drosophila*, gene order of essential genes involved in embryonic development are highly conserved among divergent species despite a high degree of chromosomal rearrangements (Bhutkar et al., 2008). Our observation partially supports their hypothesis that cross-species evolutionarily conserved regions may be protected from rearrangement (at least inversions) through strong expression correlation and purifying selection against variants.

We also found that inversions were enriched for TEs consistently in all 32 genera while breakpoints had only mixed evidence for TE enrichment (Figure 5B and D). TE enrichment around breakpoints is often associated as the driver of inversions—as discussed in the next section, whereas those inside inversions is more reasonably interpreted as the consequence of inversions. Reduced recombination allows for such selfish elements to accumulate over time by escaping natural selection. Empirical data has suggested this relationship to exist in a few insect species (Sniegowski & Charlesworth, 1994; Jay et al., 2021), but not in plants (Huang et al., 2022). Our result could also reflect the fact that there is a significant inverse correlation between the gene density and TE content along the chromosome, and therefore increased TE content could be a consequence of reduced gene density.

The mechanism of inversions

Different mechanisms of inversion mutations require or cause repeats around inversion breakpoints so examining repeats around inversions helps illuminate how they formed. We found that inversion breakpoints were not enriched for TEs compared to the genome-wide background. To more directly test whether inversions are surrounded by repeats, we looked for duplications across inversions. This means that a single sequence was found at both breakpoints for an inversion. We found that a relatively small proportion of inversions were surrounded by these duplications (Figure 7A). We also examined whether each breakpoint was repetitive within itself, and only found a comparable amount of repetition (Figure 7B). Taken at face value, the mixed evidence of repeats at inversion breakpoints suggests NHEJ created a majority of inversions, but we cannot rule out error in the determination of inversion breakpoints or the nucleotide sequence itself around breakpoints. A recent study in inversion polymorphisms among human genomes observed a contrasting result suggesting that inversions were predominantly formed through non-allelic homologous recombination (Porubsky et al., 2022). They employed multiple approaches to identify and characterize inversions among 82 haplotypes and showed that as high as 72% of the balanced inversions were flanked by segmental duplications or retrotransposons, a convincing support for non-allelic homologous recombination mechanism.

How do we explain the discrepancies in our results? Our analysis uses one-to-one mappings between divergent genomes to detect inversion regions, so if an inversion breakpoint were within a highly repetitive region—which it seems like it is to some extent (Figure 7B)—or otherwise unalignable regions we may have incorrectly identified breakpoints. Additionally, genome assemblies struggle resolving highly repetitive regions, such as centromeres (Naish et al., 2021). Although the genomes we included achieved chromosome-scale scaffolding, centromeric regions are especially dynamic and repetitive which means they are highly challenging to assemble correctly. Inversions can overlap with centromeres, and if breakpoints fell within centromeric

regions we may not accurately locate them (Harringmeyer & Hoekstra, 2022; Huang et al., 2022; Kirubakaran et al., 2020).

Depending on the mechanism, inversion mutations either create tandem duplicates at their breakpoints, or require them to be already present. With our current analysis, we cannot tell which is happening for each inversion, but future studies will be able as more chromosome-level genomes are produced. With multiple species genomes for a single genus, it will be possible to determine the derived and ancestral state for inversions and answer whether tandem duplications are a cause or consequence of inversions. Most current methods of genome comparison focus on pairwise comparisons (e.g., SyRI) but promising new methods using genomes graphs should allow for evolutionary analyses of genome structure across multiple species (Garrison & Guarracino, 2022).

Conclusion

Our results reject clock-like fixation of inversions in plants, and support early theoretical work that emphasized the central importance of selective and demographic effects on inversion fixation rates (Lande, 1984). What we do not know is what proportion of the inversions fixed under different selection scenarios, although comparisons between genera suggest in some cases local adaptation is a major player in inversion fixation (Connallon & Olito, 2022). Future projects should combine population level sampling with comparative genomics to determine how often newly fixed inversions show signs of selection. Additionally, for systems where de novo genome assemblies are available for multiple species across a genus, phylogenetic methods could polarize inversion mutations as well as reconstruct when inversions occurred in a phylogenetic context. This design could more robustly test the connection between mating system and inversion accumulation that is hinted at with our current analyses.

Supplementary material

Supplementary material is available online at *Evolution* (https://academic.oup.com/evolut/qpad027)

Data availability

All the codes used in this manuscript and data produced from this work are deposited in Github public repository: https://github.com/kaede0e/two-species-genome-comparison-pipeline. Additional supplementary data (detailed annotation files for TEs and SVs) is deposited in Figshare: https://doi.org/10.6084/m9.figshare.21624297.

Author contributions

K.H. compiled data and conducted analyses, K.H. and G.L.O. conceived of the project, wrote the manuscript, and created the figures.

Funding

This work was supported by an Natural Sciences and Engineering Research Council of Canada Discovery Grant to G.L.O. All images and schematics included are taken by or created by K.H.

Conflict of interest: The authors declare no conflicts of interest.

Acknowledgments

We would like to thank the creators of the genomes used in this work for making their resources publicly available.

References

Barth, J. M. I., Berg, P. R., Jonsson, P. R., Bonanomi, S., Corell, H., Hemmer-Hansen, J., Jakobsen, K. S., Johannesson, K., Jorde, P. E., Knutsen, H., Moksnes, P. O., Star, B., Stenseth, N. C., Svedäng, H., Jentoft, S., & André, C. (2017). Genome architecture enables local adaptation of Atlantic cod despite high connectivity. *Molecular Ecology*, 26(17), 4452–4466. https://doi.org/10.1111/mec.14207.

Bhutkar, A., Schaeffer, S. W., Russo, S. M., Xu, M., Smith, T. F., & Gelbart, W. M. (2008). Chromosomal rearrangement inferred from comparisons of 12 *Drosophila* genomes. *Genetics*, 179(3), 1657–1680. https://doi.org/10.1534/genetics.107.086108

Bosse, M., Megens, H. J., Derks, M. F. L., de Cara, M. R., & Groenen, M. A. M. (2019). Deleterious alleles in the context of domestication, inbreeding, and selection. *Evolutionary Applications*, 12(1), 6–17. https://doi.org/10.1111/eva.12691

Burssed, B., Zamariolli, M., Bellucco, F. T., & Melaragno, M. I. (2022). Mechanisms of structural chromosomal rearrangement formation. *Molecular Cytogenetics*, 15, 1–15.

Casals, F., & Navarro, A. (2007). Chromosomal evolution: Inversions: The chicken or the egg? *Heredity (Edinb)*, 99(5), 479–480. https://doi.org/10.1038/sj.hdy.6801046

Charlesworth, B. (1992). Evolutionary rates in partially self-fertilizing species. *American Naturalist*, 140(1), 126–148. https://doi.org/10.1086/285406.

Charlesworth, B., & Charlesworth, D. (1973). Selection of new inversions in multi-locus genetic systems. *Genetical Research*, 21(2), 167–183. https://doi.org/10.1017/s0016672300013343

Connallon, T., & Olito, C. (2022). Natural selection and the distribution of chromosomal inversion lengths. *Molecular Ecology*, 31(13), 3627–3641. https://doi.org/10.1111/mec.16091

Coyne, J. A., & Orr, A. H. (2004). Species: Reality and concepts. In Sinauer (Ed.), *Speciation* (pp. 8–54). Oxford University Press.

Dobzhansky, T. (1933). On the sterility of the interracial hybrids in *Drosophila pseudoobscura*. *Proceedings of the National Academy of Sciences*, 19, 397–403.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4(5), e88. https://doi.org/10.1371/journal.pbio.0040088.

Ellinghaus, D., Kurtz, S., & Willhoeft, U. (2020). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*, 9, 18. https://doi.org/10.1186/1471-2105-9-18

Exposito-Alonso, M., Becker, C., Schuenemann, V. J., Reiter, E., Setzer, C., Slovak, R., Brachi, B., Hagmann, J., Grimm, D. G., Chen, J., Busch, W., Bergelson, J., Ness, R. W., Krause, J., Burbano, H. A., & Weigel, D. (2018). The rate and potential relevance of new mutations in a colonizing plant lineage. *PLoS Genetics*, 14(2), e1007155. https://doi.org/10.1371/journal.pgen.1007155.

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, 117, 9451–9457.

Garrison, E., & Guarracino, A. (2022). Unbiased pangenome graphs. *bioRxiv*, 2022(02), 14.480413.

Goel, M., Sun, H., Jiao, W. B., & Schneeberger, K. (2019). SyRI: Finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology*, 20, 1–13.

Gottlieb, L. D. (1973). Genetic differentiation, sympatric speciation, and the origin of a diploid species of stephanomeria.

American Journal of Botany, 60(6), 545–553. https://doi.org/10.1002/j.1537-2197.1973.tb05956.x.

Hager, E., Harringmeyer, O., Wooldridge, T., Theingi, S., Gable, J., McFadden, S., Neugeboren, B., Turner, K., Jensen, J., & Hoekstra, H. (2022). A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. *Science (80-.)*, 337, 399–405.

Harringmeyer, O. S., & Hoekstra, H. E. (2022). Massive inversion polymorphisms shape the genomic landscape of deer mice. *bioRxiv*.

Hedrick, P. W. (1981). The Establishment of Chromosomal Variants. *Evolution (N. Y)*, 35, 322.

Henderson, I. R. (2012). Control of meiotic recombination frequency in plant genomes. *Current Opinion in Plant Biology*, 15(5), 556–561. https://doi.org/10.1016/j.pbi.2012.09.002

Ho, S., Lanfear, R., Bromham, L., Phillips, M., Soubrier, J., Rodrigo, A., & Cooper, A. (2011). Time-dependent rates of molecular evolution. *Molecular Ecology*, 20, 3087–3101.

Hoffmann, A. A., & Rieseberg, L. H. (2008). Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and Systematics*, 39, 21–42. https://doi.org/10.1146/annurev.ecolsys.39.110707.173532

Huang, K., Andrew, R. L., Owens, G. L., Ostevik, K. L., & Rieseberg, L. H. (2020). Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Molecular Ecology*, 29(14), 2535–2549. https://doi.org/10.1111/mec.15428

Huang, K., Ostevik, K. L., Elphinstone, C., Todesco, M., Bercovich, N., Owens, G. L., & Rieseberg, L. H. (2022). Mutation load in sunflower inversions is negatively correlated with inversion heterozygosity. *Molecular Biology and Evolution*, 39(5), msac101. https://doi.org/10.1093/molbev/msac101

Huang, K., & Rieseberg, L. H. (2020). Frequency, origins, and evolutionary role of chromosomal inversions in plants. *Frontiers in Plant Science*, 11, 1–13.

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. *Nature Genetics*, 53(3), 288–293. https://doi.org/10.1038/s41588-020-00771-1

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., & Kingsley, D. M.; Broad Institute Genome Sequencing Platform & Whole Genome Assembly TeamBroad Institute Genome Sequencing Platform & Whole Genome Assembly Team (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392), 55–61. https://doi.org/10.1038/nature10944

Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, 173(1), 419–434. https://doi.org/10.1534/genetics.105.047985

Kirubakaran, T. G., Andersen, O., Moser, M., Árnyasi, M., McGinnity, P., Lien, S., & Kent, M. (2020). A nanopore based chromosome-level assembly representing atlantic cod from the celtic sea. *G3 Genes|Genomes|Genetics*, 10(9), 2903–2910. https://doi.org/10.1534/g3.120.401423

Koch, E. L., Ravinet, M., Westram, A. M., Johannesson, K., & Butlin, R. K. (2022). Genetic architecture of repeated phenotypic divergence in Littorina saxatilis ecotype evolution. *Evolution (N. Y).*, 76, 2332–2346.

Korneev, S., & O'Shea, M. (2002). Evolution of nitric oxide synthase regulatory genes by DNA inversion. *Molecular Biology and Evolution*, 19(8), 1228–1233. https://doi.org/10.1093/oxfordjournals.molbev.a004183

Küpper, C., Stocks, M., Risse, J. E., dos Remedios, N., Farrell, L. L., McRae, S. B., Morgan, T. C., Karlionova, N., Pinchuk, P., Verkuil, Y. I., Kitaysky, A. S., Wingfield, J. C., Piersma, T., Zeng, K., Slate, J., Blaxter, M., Lank, D. B., & Burke, T. (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics*, 48(1), 79–83. https://doi.org/10.1038/ng.3443

Lamichhaney, S., Fan, G., Widemo, F., Gunnarsson. U., Thalmann, D. S., Hoeppner, M. P., Kerje, S., Gustafson, U., Shi, C., Zhang, H., Chen, W., Liang, X., Huang, L., Wang, J., Liang, E., Wu, Q., Lee, S. M.-Y., Xu, X., Höglund, J., Liu, X., and Andersson, L. (2016). Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax). *Nature Genetics*, 48, 84–88.

Lande, R. (1984). The expected fixation rate of chromosomal inversions. *Evolution (N. Y)*. 38:743–752.

Lewis, H., & Raven, P. H. (1958). Rapid evolution in Clarkia. *Evolution (N. Y)*, 12, 319–336.

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H. (2021). New strategies to improve minimap2 alignment accuracy. *Bioinformatics*, 37(23), 4572–4574. https://doi.org/10.1093/bioinformatics/btab705

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

Lloyd, A. (2022). Crossover patterning in plants. *Plant Reprod*, https://doi.org/10.1007/s00497-022-00445-4. Springer Berlin Heidelberg

Lowry, D. B., & Willis, J. H. (2010). A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biology*, 8, e1000500. https://doi.org/10.1371/journal.pbio.1000500

Martin, N. H., & Willis, J. H. (2010). Geographical variation in postzygotic isolation and its genetic basis within and between two Mimulus species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 2469–2478.

Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends in Ecology and Evolution*, 35(7), 561–572. https://doi.org/10.1016/j.tree.2020.03.002.

Naish, M., Alonge, M., Wlodzimierz, P., Tock, A. J., Abramson, B. W., Schmucker, A., Mandakova, T., Jamge, B., Lambing, C., Kuo, P., Yelina, N., Hartwick, N., Colt, K., Smith, L. M., Ton, J., Kakutani, T., Martienssen, R. A., Schneeberger, K., Lysak, M. A., & Henderson, I. R. (2021). The genetic and epigenetic landscape of the Arabidopsis centromeres. *Science, 374*(6569):eabi7489..

Ohta, T., & Kojima, K. I. (1968). Survival probabilities f new inversions in large populations. *Biometrics*, 24(3), 501–516.

Ou, S., & Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiology*, 176(2), 1410–1422. https://doi.org/10.1104/pp.17.01310

Ou, S., & Jiang, N. (2019). LTR_FINDER_parallel: parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. *Mob DNA*, 10, 48. https://doi.org/10.1186/s13100-019-0193-0

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, A. J., Lugo, C. S. B., Elliott, T. A., Ware, D., Peterson, T., Jiang, N., Hirsch, C. N., & Hufford, M. B. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology,* 20, 1–18.

Porubsky, D., Höps, W., Ashraf, H., Hsieh, P. H., Rodriguez-Martin, B., Yilmaz, F., Ebler, J., Hallast, P., Maria Maggiolini, F. A., Harvey, W. T., Henning, B., Audano, P. A., Gordon, D. S., Ebert, P., Hasenfeld, P., Benito, E., Zhu, Q., Lee, C., Antonacci, F., Steinrücken, M., Beck, C. R., Sanders, A. D., Marschall, T., Eichler, E. E., & Korbel, J. O. (2022). Recurrent inversion polymorphisms in humans associate with genetic instability and genomic disorders. *Cell*, 185, 1986–2005.e26.

Pucker, B., Irisarri, I., Vries, J. de, & Xu, B. (2022). Plant genome sequence assembly in the era of long reads: Progress, challenges and future directions. Quant. *Plant Biology*, 3, E5.

Purcell, J., Brelsford, A., Wurm, Y., Perrin, N., & Chapuisat, M. (2014). Convergent genetic architecture underlies social organization in ants. *Current Biology*, 24(22), 2728–2732. https://doi.org/10.1016/j.cub.2014.09.071

Quadrana, L., Bortolini Silveira, A., Mayhew, G. F., LeBlanc, C., Martienssen, R. A., Jeddeloh, J. A., & Colot, V. (2016). The Arabidopsis thaliana mobilome and its impact at the species level. *Elife*, 5, e15716. https://doi.org/10.7554/eLife.15716.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Ranz, J. M., Casals, F., & Ruiz, A. (2001). How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus *Drosophila*. *Genome Research*, 11(2), 230–239. https://doi.org/10.1101/gr.162901

Ranz, J. M., Maurin, D., Chan, Y. S., von Grotthuss, M., Hillier, L. W., Roote, J., Ashburner, M., & Bergman, C. M. (2007). Principles of genome evolution in the *Drosophila melanogaster* species group. *PLoS Biology*, 5(6), e152–e152. https://doi.org/10.1371/journal.pbio.0050152.

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution*, 16(7), 351–358. https://doi.org/10.1016/s0169-5347(01)02187-5

Shi, J., & Liang, C. (2019). Generic repeat finder: A high-sensitivity tool for genome-wide de novo repeat detection. *Plant Physiology*, 180(4), 1803–1815. https://doi.org/10.1104/pp.19.00386

Sniegowski, P. D., & Charlesworth, B. (1994). Transposable element numbers in cosmopolitan inversions from a natural population of *Drosophila melanogaster*. *Genetics*, 137(3), 815–827. https://doi.org/10.1093/genetics/137.3.815

Sodeland, M., Jentoft, S., Jorde, P. E., Mattingsdal, M., Albretsen, J., Kleiven, A. R., Synnes, A. E. W., Espeland, S. H., Olsen, E. M., Andrè, C., Stenseth, N. C., & Knutsen, H. (2022). Stabilizing selection on Atlantic cod supergenes through a millennium of extensive exploitation. *Proceedings of the National Academy of Sciences*, 119(8), e2114904119.

Sturtevant, A. H., & Mather, K. (1938). The interrelations of inversions, heterosis and recombination. *American Naturalist*, 72(742), 447–452. https://doi.org/10.1086/280797.

Todesco, M., Owens, G. L., Bercovich, N., Légaré, J. S., Soudi, S., Burge, D. O., Huang, K., Ostevik, K. L., Drummond, E. B. M., Imerovski, I., Lande, K., Pascual-Robles, M. A., Nanavati, M., Jahani, M., Cheung, W., Staton, S. E., Muños, S., Nielsen, R., Donovan, L. A., Burke, J. M., Yeaman, S. L., & Rieseberg, H. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, 584, 602–607.

Trickett, A. J., & Butlin, R. K. (1994). Recombination suppressors and the evolution of new species. *Heredity (Edinb)*, 73, 339–345.

Wallberg, A., Schöning, C., Webster, M. T., & Hasselmann, M. (2017). Two extended haplotype blocks are associated with adaptation to high altitude habitats in East African honey bees. *PLoS Genetics*, 13(5), e1006792. https://doi.org/10.1371/journal.pgen.1006792.

Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y. C., Shoemaker, D., & Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature*, 493(7434), 664–668. https://doi.org/10.1038/nature11832

Wang, L., Ji, Y., Hu, Y., Hu, H., Jia, X., Jiang, M., Zhang, X., Zhao, L., Zhang, Y., Jia, Y., Qin, C., Yu, L., Huang, J., Yang, S., Hurst, L. D., & Tian, D. (2019). The architecture of intra-organism mutation rate variation in plants. *PLoS Biology*, 17(4), e3000191. https://doi.org/10.1371/journal.pbio.3000191.

White, M. J. D. (1973). *Animal cytology and evolution*. Cambridge University Press.

White, M. J. D. (1978). *Modes of speciation*. W.H. Freeman & Company.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Yutani, A. H., Henry, L., Hester, J., M, K. A., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., & Hiroaki, K. W. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4, 1686.

Xiong, W., He, L., Lai, J., Dooner, H., & Du, C. (2014). HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 10263–10268.

Xu, Z., & Wang, H. (2007). LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, 35, W265–W268. https://doi.org/10.1093/nar/gkm286

Zhang, L., Reifová, R., Halenková, Z., & Gompert, Z. (2021). How important are structural variants for speciation? *Genes (Basel)*, 12(7), 1084.

Zhang, R. G., Wang, Z, Ou, S., & Li, G. Y. (2019). TEsorter: Lineage-level classification of transposable elements using conserved protein domains. *bioRxiv* 800177.

Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, 7(1–2), 203–214. https://doi.org/10.1089/10665270050081478.