1 ## TITLE:

2 The rate of inversion fixation in plant genomes is highly variable.

3 ## AUTHORS:

4 Kaede Hirabayashi[1], Gregory L. Owens[1*]

5

6 [1] Department of Biology, University of Victoria, Victoria, BC, Canada

7 *Corresponding author, grego@uvic.ca

8 ## ABSTRACT:

9 Chromosomal inversions are theorized to play an important role in adaptation by preventing recombination, but
10 testing this hypothesis requires an understanding of the rate of inversion fixation. Here we use chromosome-level
11 whole genome assemblies for 32 genera of plants to ask how fast inversions accumulate and what factors affect this
12 rate. We find that on average species accumulate 4 to 28 inversions per million generations, but this rate is highly
13 variable, and we find no correlation between sequence divergence or repeat content and the number of inversions
14 and only a small correlation with chromosome size. We also find that inversion regions are depleted for genes and
15 enriched for TEs compared to the genomic background. This suggests that idiosyncratic forces, like natural selection
16 and demography, are controlling how fast inversions fix.

17 ## INTRODUCTION:

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

25

26 One type of structural variation, inversions, are particularly interesting because of their effect on meiosis and
27 recombination. Initial investigations into inversions focused on underdominance effects (White 1973). Due to how
28 homologous alignment occurs during meiosis, in heterozygous individuals a recombination event within an inversion
29 will lead to unbalanced gametes and a loss of fertility (Dobzhansky 1933; White 1978). This underdominance led to
30 research into their role in reproductive isolation, as inversions in different orientation are often fixed between
31 species (Trickett and Butlin 1994; Rieseberg 2001) and have been shown to directly cause hybrid sterility (reviewed
32 in L. Zhang, Reifová, Halenková, & Gompert, 2021). It was also recognized that the recombination suppression
33 abilities of inversions may link together favourable alleles and be relevant in the context of adaptation (Sturtevant
34 and Mather 1938; Ohta and Kojima 1968; Charlesworth and Charlesworth 1973; Kirkpatrick and Barton 2006). For
35 example, loci that are locally adaptive can benefit from inversions since they can be co-inherited as an adaptive

36 multi-loci haplotype and protected from recombination with non-adaptive alleles (Kirkpatrick and Barton 2006).
37 Numerous non-model systems have been observed to have inversions with adaptive significance: sunflower
38 (Todesco et al. 2020), monkeyflower (Lowry and Willis 2010), rainbow trout (Leitwein et al. 2017), Atlantic cod (Barth
39 et al. 2017; Sodeland et al. 2022), stickleback (Jones et al. 2012), deer mice (Hager et al. 2022; Harringmeyer and
40 Hoekstra 2022).

41

42 There are several molecular mechanisms of inversions, all of which are triggered by some form of DNA strand breaks
43 during meiosis or in other situations (reviewed in detail by Burssed, Zamariolli, Bellucco, & Melaragno, 2022; Casals
44 & Navarro, 2007). In non-allelic homologous recombination (NAHR), segments with a high degree of sequence
45 similarity misalign during meiosis and recombine within the chromatid (intrachromatid) instead of between sister
46 chromatids. This is known as ectopic recombination. When repeats are found in the opposite orientation, inversion
47 of the region surrounded by the misaligned repeats can occur. This means that inverted repeats are necessary for
48 inversions mutations with this mechanism. Non-homologous end-joining (NHEJ) inversions occur from random
49 paired double-stranded breaks surrounding a segment of the genome, followed by 180˚ flipping of the detached
50 region and a repair. This mechanism does not require any sequence homology around breakpoints and could occur
51 during any stage of the cell cycle. The isochromatid mechanism is similar to NHEJ but requires staggered single-
52 stranded breaks instead of double-stranded breaks (Ranz et al. 2007). The inversion through this mechanism would
53 result in the creation of inverted sequence duplicates around the breakpoint regions. Lastly, inversions that are
54 especially small and found in a complex rearrangement pattern could arise from DNA replication machinery being
55 disturbed and confused due to formation of DNA secondary structures (e.g., Cruciform, hairpin, triple helical DNA,
56 quadruplex). The replication fork may be subjected to breakage or stalling and cause template switching event,
57 called invasion. If the new position has switched orientations of strand (so say invasion occurred from template to
58 lagging strand), then inversions of that segment would result. Additionally, large number of mobile elements
59 (transposable elements or TEs) has been associated with recurrent chromosomal rearrangements, documented by
60 the reuse of some of the breakpoints identified (Ranz et al. 2007). These regions could be prone to inversions and
61 be the driving force of rearrangements.

62

63 Despite growing interest of chromosomal inversions in adaptive potential and appreciation of their prevalence, there
64 is a lack of knowledge on the rate of inversion fixation between species in plants. Most studies of the rate of
65 chromosomal inversions have focused on *Drosophila* species (Ranz et al. 2001; Bhutkar et al. 2008). A literature
66 review on plants has estimated the rate of inversion fixation to be around 15-30 inversions per million years based
67 on the divergence time of a small number of species analyzed (Huang and Rieseberg 2020). However, the method
68 of inversion detection and divergence estimation cited in this review were variable and did not account for the size
69 of the inversions. Thus, further study with robust and consistent approach is needed.

70

71 Leveraging the resource for publicly available genome assemblies, we conducted a comparative study to investigate
72 how chromosomal inversions accumulate in eudicots. We expected to observe that inversions would accumulate
73 over time at a consistent rate; therefore, more divergent species pairs should harbour more inversions. We also
74 predicted that if there is more space in the genome for chromosomal rearrangement (i.e., larger genome size), would
75 lead to more inversions. Alternatively, if fixation of inversions was dominated by selection, then rates will be
76 idiosyncratic. We further investigated the genomic context of inversions, which gives us insight into their functional

77 significance. Lastly, we asked if inversions were surrounded by inverted repeats, which tells about the relative rate
78 of different inversion mechanisms.

METHODOLOGY:

DATA COLLECTION

81 Publicly available genomes within Eudicots were searched through the published plant genomes database (last
82 screened in November 2021; https://www.plabipd.de/plant_genomes_pa.ep). We selected genera with at least two
83 sequenced species. Then individual genomes were screened for all of the following criteria: 1) same ploidy (diploid),
84 2) chromosome-resolved assembly, 3) same number of chromosomes, 4) *de novo* assembled and 5) valid NCBI
85 BioProject number or similar. Assemblies published before 2018 were carefully assessed for quality. Those that used
86 guidance from genetic mapping or reference assembly were excluded. If the assemblies did not have an associated
87 publication or provided no adequate information on how exactly the genome was sequenced, they were excluded
88 from the study. For one species per genus, we downloaded coding sequences and gene annotations. In total, 64
89 chromosome-resolved genomes from 32 genera were included (Supplementary Table 1). To explore biological
90 factors that may affect the fixation of inversions, we surveyed the literature to identify whether the species have
91 evidence of hybridization and their reproductive self-compatibility. Evidence of hybridization was categorized into
92 three groups: weak (no evidence of hybridization or evidence of sterile hybrid attempt), strong (evidence of
93 hybridization/introgression in nature or lab or genes), unclear (lack of evidence to support either weak or strong).
94 Self-compatibility was assessed by three categories: selfing (mainly reproduce by self-pollinating), mixed (can self or
95 mate), self-incompatible (cannot produce viable offspring from selfing). Two genome pairs were missing the gene
96 annotation file and thus excluded from analyses involving gene locations. All species used and recorded information
97 are in Supplementary Table 1.

98

STRUCTURAL ANALYSIS

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

113

114 For our analyses, we identified inversions that differ in orientation between closely related species, but we are not
115 able to determine which orientation is derived or ancestral. An inversion region in the reference species is a region

116 that is in the opposite orientation in the query species, which is because either the reference or query species
117 inverted. If we assume that the rate of inversion fixation is roughly equal between the two species – which may not
118 be true – then roughly half of the inversion regions contain a derived inversion in the reference species and the other
119 half contain the ancestral state, but have inverted in the query species.

120

TRANSPOSABLE ELEMENT ANNOTATION

121
122 Transposable elements (TEs) were detected and annotated using Extensive De novo TE Annotator v2.0.0 (EDTA) (Ou
123 et al. 2019) pipeline on the 30 reference genomes with gene annotations available. Due to resource limitation, the
124 pipeline was performed by chromosome-by-chromosome. The whole genome fasta file was first divided into
125 individual chromosome files. Then the EDTA pipeline was performed separately on each chromosome as follows and
126 the resulting TE library was later combined. In brief, candidate TE sequences were *de novo* identified using LTR-
127 Finder (Xu and Wang 2007; Ou and Jiang 2019), LTRharvest (Ellinghaus et al. 2020), LTR_retriever (Ou and Jiang 2018),
128 generic repeat finder (Shi and Liang 2019), and HelitronScanner (Xiong et al. 2014) respectively. Once individual TEs
129 were identified, candidates were filtered (Zhang et al. 2019) and further refined by RepeatModeler (Flynn et al. 2020)
130 according to EDTA pipeline default parameters. Finally, TE-free coding sequences retrieved from gene annotation
131 files (feature name = 'CDS') were aligned to the repeat library, and those overlapped with coding sequence (CDS)
132 were excluded from the identified TE candidates. Total genomic TE content (%) was calculated by the total length of
133 detected TE divided by the total genome length.

134

SPECIES DIVERGENCE CALCULATION

135
136 The principal advantage of SyRI over basic genome aligners is that SyRI identifies larger regions with consistent
137 synteny. To do this, SyRI takes alignment blocks identified by the initial aligner, in this case minimap2, and identifies
138 larger regions consisting of multiple consecutive one-to-one alignment blocks. We focused our analysis on regions
139 which each represent a single structural variant (or contiguous syntenic region without any structural variation).
140 While SyRI identifies several types of structural variation, we focused our analysis on inversions and syntenic regions
141 only. To minimize noise from small amounts of data, detected regions smaller than 1 kbp were excluded from the
142 dataset. To calculate the sequence divergence for each region (X) we used equation 1. From SyRI output, for each
143 block (i) in a region, we measured percent identity (x_i) and reference length (L_i).

$$\text{Equation 1: } X = 1 - \frac{\sum_{i=1}^{n} L_i x_i}{\sum_{i=1}^{n} L_i}$$

145 To calculate the average sequence identity between species, we used the equation 1 but, in this case, combining all
146 syntenic blocks. With this formula, sequence divergence score of 0 indicates identical species, and the value
147 increases as the genomes become more different.

148

149 To convert sequence divergence into divergence time we used the genome-wide substitution rate estimated from
150 *Arabidopsis thaliana* (Exposito-Alonso et al. 2018), which estimated 2 to 5 x 10^{-9} substitutions per year. We note that
151 *Arabidopsis* is an annual plant, while the species studied here have varying generation times which may affect
152 substitution rates, but estimates of mutation rates from somatic tissue of plants found a range of values which

153   encompassed the substitution rate used here (Wang et al. 2019). Based on this rate, and that the divergence rate is

154   equal for each species in the pair, each percentage point divergence represents 1 to 2.5 million years of separation.

155   This estimate should be used with caution as substitution rates are known to vary and divergence between close

156   relatives is affected by standing variation (Ho et al. 2011). We present our rates of inversion accumulation in years

157   of evolution. This means that a species pair that had a common ancestor 1 million years ago, is separated by 2 million

158   years of evolution accounting for evolution in both species.

159

160   QUANTIFYING CDS AND TE IN INVERTED REGIONS VS. SYNTENIC REGIONS

161   We were interested if inversions contained different proportions of genomic elements. Bedtools v2.30.0 (Quinlan

162   and Hall 2010) intersect function was used to analyze how many features and base pairs of coding sequence (CDS;

163   as identified by annotations) or TEs (as identified by EDTA) were found to overlap with the inverted regions in the

164   genome. Inversion regions identified by SyRI were extracted using custom perl script and reorganized into a bed

165   format using Linux commands. The gene annotation file and TE library in gff3 were reformatted into bed format

166   using gff2bed. Prior to using bedtools intersect, CDS and TE bed files were edited by bedtools merge function to

167   concatenate overlapping features into a single feature, avoiding overrepresentation due to isoforms of the same

168   gene sometimes present in the gene annotation file. The validity of data was confirmed by ensuring the proportion

169   of CDS or TE per inverted region no larger than 1. In addition to inverted regions, the process was repeated with

170   syntenic regions identified by SyRI.
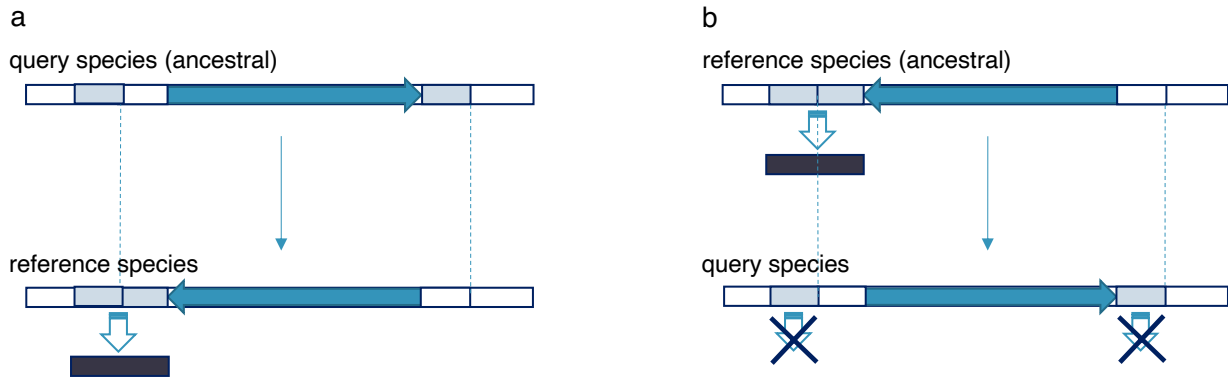
171

172   QUANTIFYING CDS AND TE IN INVERSION BREAKPOINT REGIONS

173   We were also interested in whether genomic elements differ at the breakpoints of inversions compared to the rest

174   of the genome. Quantification of CDS/TE in inversion breakpoints were performed using Bedtools (Quinlan and Hall

175   2010) intersect function. The number of CDS and TE in the inversion breakpoint regions were extracted as follows.

176   First, inversion breakpoint regions (defined as 4 kbp regions downstream or upstream of breakpoints start or end

177   respectively) were identified. A small number of inversion breakpoint regions were within 4 kbp of the end of the

178   chromosome, and were excluded from the dataset. A total of 11,225 breakpoint regions from 30 genus pairs were

179   used for analysis. To infer whether breakpoints are enriched with CDS/TE, baseline numbers for the whole genome

180   are necessary. To determine the baseline number of CDS/TE in the genome, the total length of CDS/TE was computed

181   from the merged bed file from previous step. Then the genomic CDS/TE proportion was calculated by the total length

182   of CDS/TE divided by the length of the reference genome. The resulting average proportion of CDS/TE in the 4 kbp

183   breakpoint regions was compared to the genomic CDS/TE proportion.

184

185   QUANTIFYING FREQUENCY OF GENE OCCURRENCE AT INVERSION BREAKPOINTS

186   Chromosomal inversions can disrupt gene sequence if an inversion breakpoint occurs within the gene itself. Since

187   we cannot identify whether the derived orientation occurred in the reference or query genome, we are focusing on

188   genes identified in the reference species. If an inversion breakpoint falls within a gene sequence it has two possible

189   meanings (Figure 1):

Figure 1: 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.

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 1000 per chromosome for each genome) from the syntenic portion of the genome using Bedtools (Quinlan and 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.

RESULTS:

INVERSION ACCUMULATION IS IDIOSYNCRATIC

217   We identified a total of 6,140 inversions across our 32 comparisons, 5,298 of which were larger than 1000 bp. In
218   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)
219   were between 10 kbp and 1 Mbp and 8.2% (501/6,140) were greater than 1 Mbp (Figure 2a). The number of
220   inversions varied between comparisons (Figure 2b), but in general represented a relatively small proportion of the
221   total genome length, from 1.3% to 37.4% (Figure 2c). For each of our genera, we observed whether sequence identity
222   in inversion regions differed from sequence identity in syntenic regions and found no significant difference for all 32
223   comparisons (Supplementary Figure 1). In syntenic regions, our species had on average 3% to 5% sequence
224   divergence, supporting their close relationship (Supplementary Figure 2).

225

226   We were interested in determining factors that affected the number of inversions and the rate of inversion
227   accumulation. Sequence divergence is expected to increase over time and can be a proxy for divergence time (with
228   many potential caveats) (Drummond et al. 2006; Ho et al. 2011). We used a linear model to ask if species with higher
229   sequence divergence had more inversions and found no significant relationship (p = 0.6) (Figure 2d). This was also
230   not significant if we instead used the proportion of the genome that is inverted, rather than the number of inversions
231   (p = 0.5) (Supplementary Figure 3). All else being equal, we expected that larger genomes should have more
232   inversions. When tested using the entire genome length, we saw no significant relationship (p = 0.96) (Figure 2e),
233   but when treating each chromosome separately we found a slight positive relationship (n = 405; F value = 12.78; p =
234   0.00039 ***; linear model R^2= 0.029) (Figure 2f). Since inversion origin is often associated with DNA repeats, we
235   expected that species with a higher proportion of TEs would have more inversions, but again this relationship was
236   not significant (p = 0.3) (Figure 2g).

237

238   Given that inversions are not accumulating in a clock-like manner, we used a one-way ANOVA to explore other
239   factors controlling the number of inversions. We found that the assembly method used, sequencing technology and
240   whether the species had evidence of hybridization did not affect the number of inversions significantly (Table 1), but
241   reproductive strategy did when it was calculated with the proportion of inverted genome instead of the number (p
242   = 0.031 *). We found that selfing species had the highest proportion of genome inverted, while mixed and self-
243   incompatible species had a similar proportion (Supplementary Figure 4).

244

7
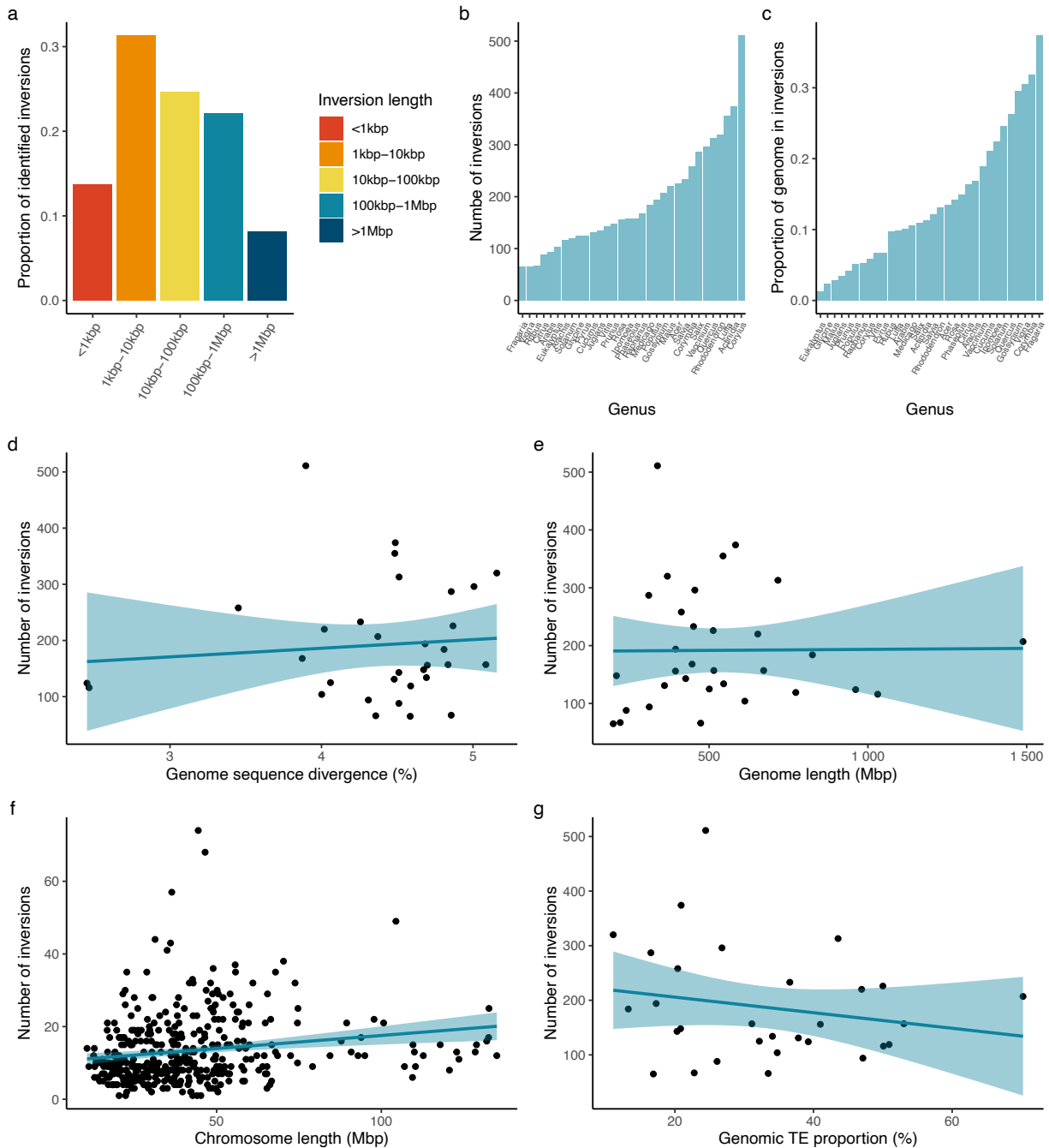
Figure 2: Summary of inversions between paired eudicot genomes in the same genus. a) Size distribution of inversions by inversion length. **b**) Number of inversions and **c**) proportion of reference genome in inversion, for 32 species pairs used in the study. Number of inversions plotted against **d**) percent sequence divergence between the 32 species pair (F value = 0.264, p = 0.611), **e**) genome length of the reference species genome (n = 32; F value = 0.002, p = 0.961), **g**) genomic TE proportion calculated by total length of TE / reference genome length (n = 30; F value = 1.139, p = 0.295). Each datapoint represents a paired species within the same genus. **f**) Number of inversions per chromosome plotted against its length (n = 405; F value = 12.78; p = 0.00039 ***; linear model R^2 = 0.029).

254 **Table 1: One-way ANOVA results.** Tests performed for the number of inversions and proportion of reference
255 genome in inverted orientation against the following five factors: evidence of hybridization (weak/strong),
256 reproductive strategy (selfing/mixed/self-incompatible), genome assembler category (accurate/fast or not resource
257 intensive/short-read based), long-read sequencing platform (Oxford Nanopore/Pacific Bioscience/none or short-
258 read only), whether physical mapping (i.e., Hi-C, BioNano optical map, long-range Chicago) is performed or not. The
259 reproductive strategy and assembly methods were assigned to each genome assembly (reference and query)
260 separately, then the paired category was used in the stats (n = 32).

Inversion measure	Number of inversions		Proportion of genome inverted	
	F value	Pr(>F)	F value	Pr(>F)
Evidence of hybridization	0.317	0.578	0.001	0.973
Reproductive strategy	1.73	0.163	2.942	0.031 *
Assembler category (accurate, fast, short-read based)	1.223	0.324	0.73	0.579
Sequencing platform (ONT vs PacBio vs short-read only)	0.772	0.52	1.504	0.235
Physical mapping (Y/N)	1.679	0.204	0.023	0.977

261 *** STATISTICALLY SIGNIFICANT (P <0.05)**

262

263 Although the rate of inversion accumulation relative to sequence divergence is not consistent in our dataset, the
264 range of possible values may be a useful baseline for other studies. We present a fixation rate by dividing the number
265 of inversions by the sequence divergence for different size categories of inversions (Figure 3). We find the smaller
266 inversions are more common, with the exception of the smallest category. Further, based on an estimate of 1 to 2.5
267 million years of divergence per 1% sequence divergence, we present a range of estimates of inversion accumulation
268 by size (Table 2).

269

9
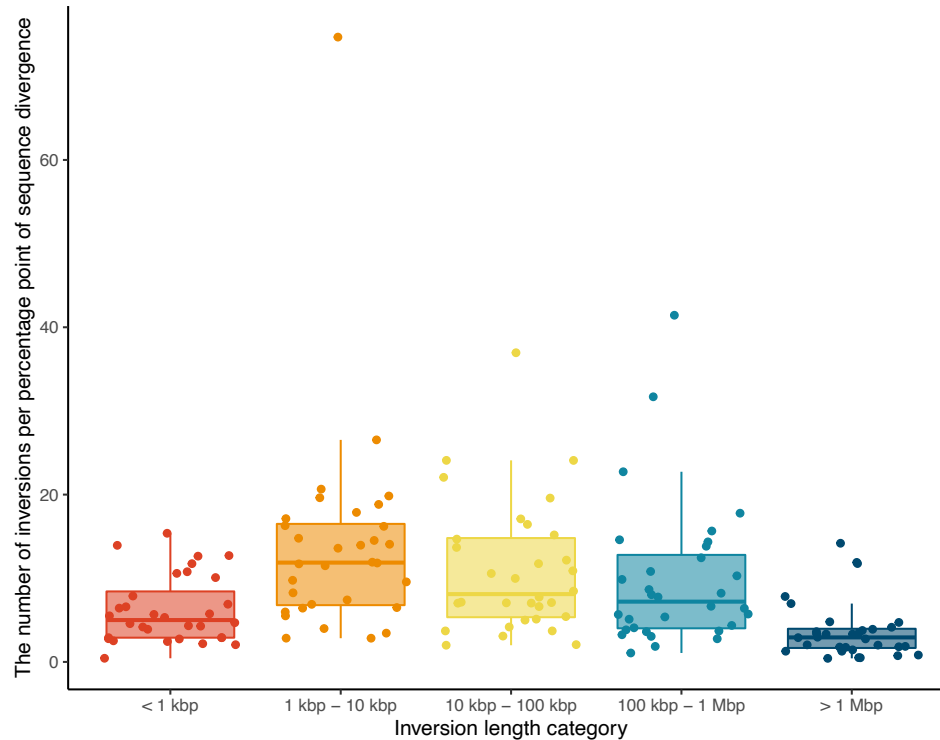
Figure 3: Rate of inversion fixation by inversion size. Rate of inversion occurrence calculated as number of inversions per percentage point of sequence divergence by different size categories (n = 32 for each category).

Table 2: Rate of inversion accumulation by inversion length with respect to species divergence time. Species divergence time was estimated using *Arabidopsis thaliana* substitution rate (Exposito-Alonso et al. 2018).

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 years ago, 25th – **50th** – 75th percentile)					
Conversion				1 mya / 1% seq divergence			2.5 mya / 1% seq divergence		
< 1 kbp	2.9	**5.0**	8.4	1.5	**2.5**	4.2	0.6	**1.0**	1.7
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	20.7	**35.1**	56.5	10.3	**17.5**	28.2	4.1	**7.0**	11.3

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 to syntenic regions. Indeed, we found this was the case (Figure 4a; t = -

10

279   13.97, df = 29, p = 2.08e-14 ***). We also found the reverse pattern for TEs, where inversions were enriched
280   compared to syntenic regions (Figure 4b; t = -3.93, df = 29, p = 4.77e-4 ***).

281
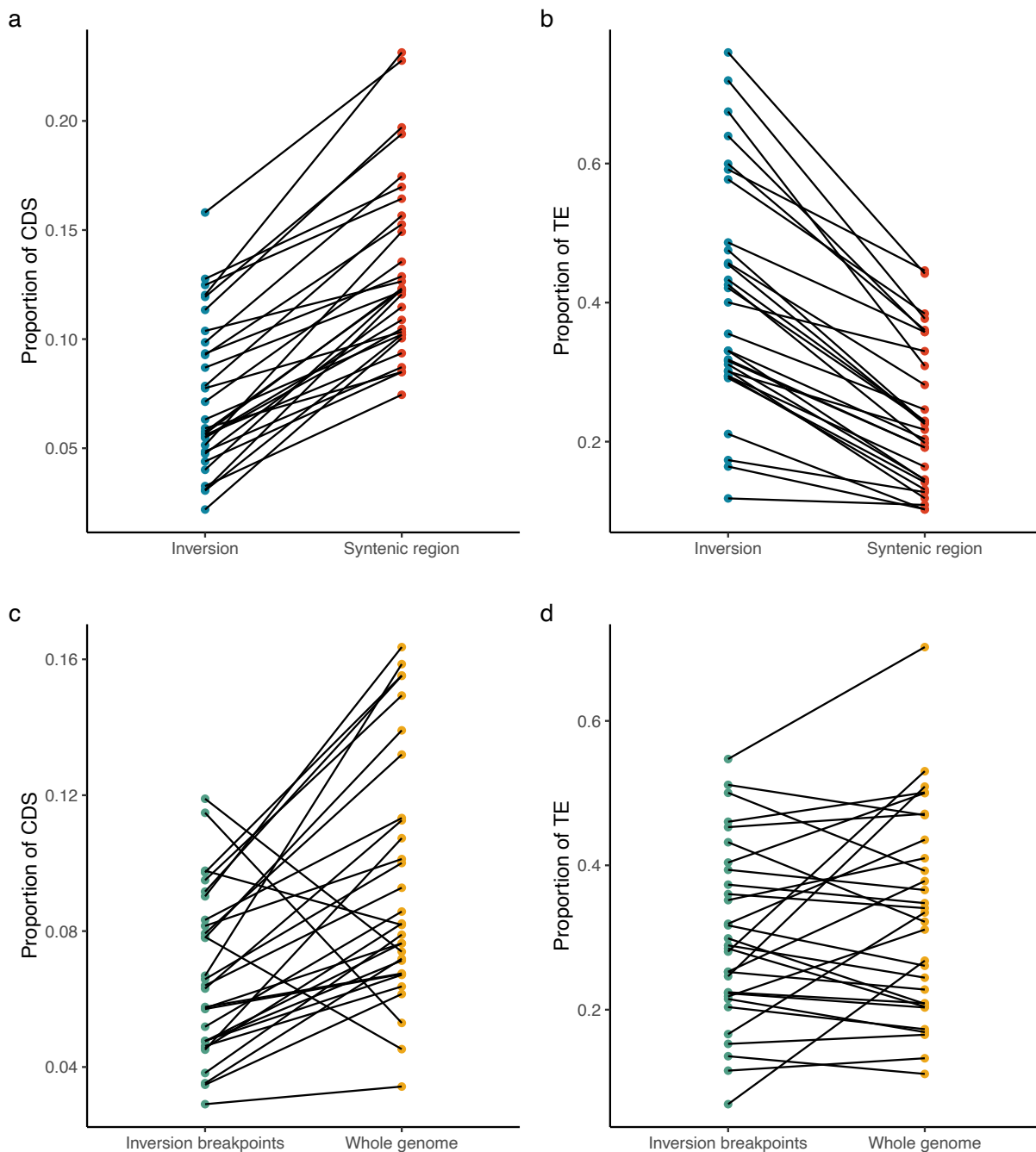
Figure 4: Genomic context of inversions and surrounding regions. The mean proportion of **a**) coding sequence and **b**) transposable elements inside inversion (blue) or syntenic region (red) (n = 30; p = 2.08e-14***, 4.77e-4***, respectively). The mean proportion of **c**) coding sequence and **d**) transposable elements in 4 kbp breakpoint regions (green) compared to the genomic CDS and TE proportion (yellow) in the genome (n = 30; p = 0.000107 ***, 0.142, respectively). Each datapoint represents a genus pair and the same species pair is connected by the line.

287

288   We next asked whether inversion breakpoints differed from the genome background. Although SyRI identifies
289   breakpoints, limitations in the alignment of repetitive regions means that these may not be the exact breakpoint,
290   therefore we selected 4 kbp regions downstream/upstream of identified breakpoint positions. These regions were
291   compared against the genomic CDS and TE content in reference genome. We found that there was a lower
292   proportion of CDS in breakpoint regions (Figure 4c; *p* = 0.000107 ***), but no consistent trend for TEs (Figure 4d; *p*
293   = 0.142).

294

295   To infer whether inversions are disrupting gene sequence, or alternatively involved in the creation of new genes, we
296   counted the number inversion breakpoints (start/end) that fell within a gene (Figure 5a). Strikingly high occurrence
297   of inversion breakpoints at a gene was observed in several genera (Figure 5a). Notably *Vigna*, *Ipomoea*, *Medicago*,
298   *Solanum*, *Cucumis*, *Salix*, and *Phaseolus* had > 50% of breakpoints occurring within a gene. Despite this, the relatively
299   small number of inversions compared to genes means that < 1% of genes contained an inversion breakpoint (Figure
300   5b).

301
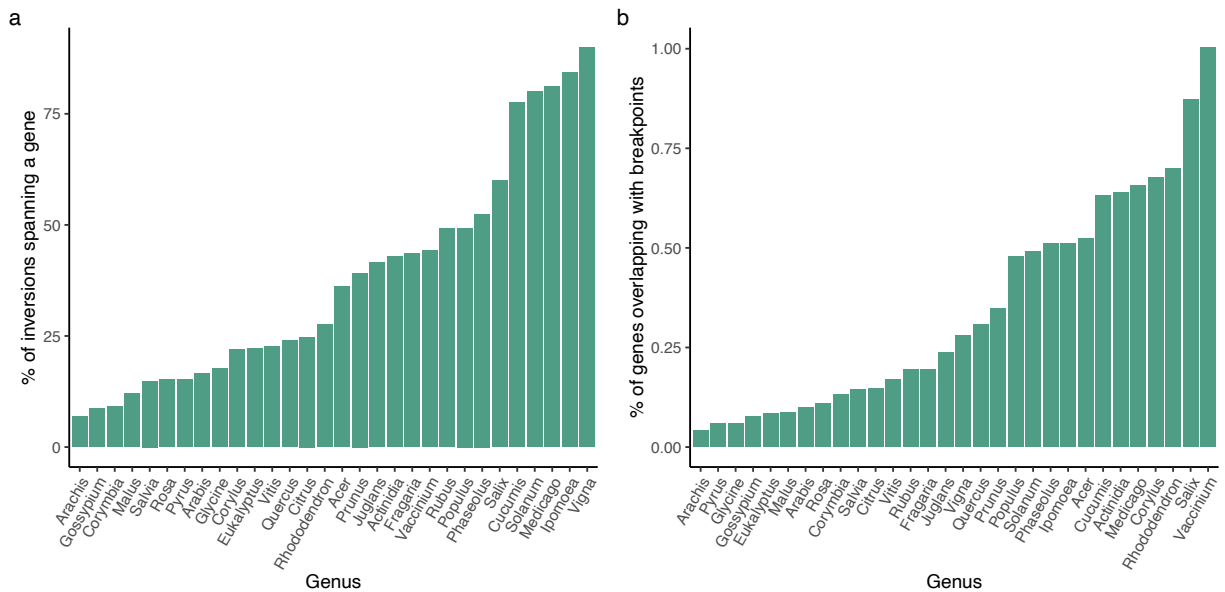
302

303   **Figure 5: Inversion breakpoints and genes overlap. a**) The percentage of inversion breakpoints that overlapped with
304   a gene by genus. **b**) The percent of genes in reference genome that contained an inversion breakpoint.

305

306   INVERSIONS ARE MOST OFTEN NOT SURROUNDED BY REPEATS

307   Lastly, to test the molecular mechanism of inversions, we specifically looked for duplicates or inverted repeats
308   surrounding the inversion breakpoints. Among the 5,626 tested 10 kbp breakpoint regions pair, 1,160 (20.6%)

309   regions pair resulted in at least one BLAST hit. Of these, 592 were in the forward orientation and 568 were in inverted
310   orientation (Figure 6).

311
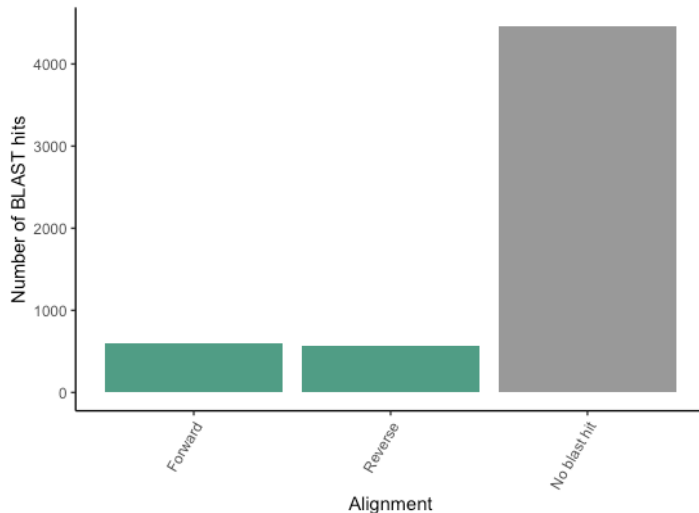
312   **Figure 6: Rare presence of duplicates or inverted repeats near breakpoints.** Summary of alignment between the 10
313   kbp window surrounding the 5,626 inversions detected using BLAST.

DISCUSSION:

THE RATE OF INVERSION ACCUMULATION

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

324

325   To measure the rate of inversion fixation, we aligned chromosome-level genome assemblies for two species within
326   the same genus. These comparisons explore how fast inversions fix in the genome by comparing the sequence
327   divergence between species with the number of inversion differences. We are treating the rate as a baseline in
328   comparison to systems like *H. petiolaris* where there is evidence that inversions are positively selected, but we
329   recognize that the inversions themselves may not be neutral and that each species has its own demographic and
330   adaptive history. Surprisingly, we find no correlation between sequence divergence, which is a proxy for coalescence
331   time, and the number of inversions separating species (Figure 2d), suggesting that the rate of accumulation is
332   dependent on factors not consistent between genera. Based on our data, the middle 50% of comparisons
333   accumulated 20 to 57 inversions per 1% sequence divergence. Converting from sequence divergence to divergence

13

334   time is fraught because it is affected by generation time, the mutation rate, and other factors which are likely to vary

335   between our genera but it is helpful for scaling expectations (Ho et al. 2011; Wang et al. 2019). Based on substitution

336   rates from *Arabidopsis*, we estimate 4 to 28 inversions per million years which is within the range estimated by

337   Huang et al. from a more limited dataset (Huang and Rieseberg 2020).

338

339   Although we present a range of inversion accumulation rates, our primary finding is that this value is not consistent.

340   There are several reasons — both technical and biological — that could explain the variation. Our inversion counts

341   are based on whole genome alignments which is much better at capturing small inversions than genetic mapping

342   but are not without error. Artifactual inversions can be introduced during assembly or scaffolding. We looked for

343   the effect of sequencing and assembly method and failed to find a significant relationship, but our test is

344   underpowered for the number of categories based on our sample size. There are also limitations in our ability to

345   detect inversions based on alignments. We find that smaller inversions are more common, except for the smallest

346   category (Figure 2a), which may reflect the limits of detectability or a trade-off between the high sensitivity to detect

347   smaller inversions while minimizing the erroneous inverted alignments, especially in repetitive regions.

348

349   Different taxa have diverse genomic properties. Some may have inherently high recombination rate or epigenetic

350   patterns that allow them to harbour chromosomal rearrangements more easily (Henderson 2012; Lloyd 2022). Each

351   species also has its own demographic and selection history. Several of the species used in this analysis are

352   domesticated crops, which have likely undergone greater bottlenecks and distinct selection pressures compared to

353   wild species. This could lead to greater fixation of deleterious inversions similar to mutational load seen in some

354   domestic species (Bosse et al. 2019).

355

356   The rate of inversion fixation is expected to be related to the rate of inversion mutation. Logically, longer genomes

357   should have more opportunity for double-stranded breaks and therefore more inversion mutations. When

358   considering each genome as a replicate we do not see a significant relationship, although there is a relatively small

359   but significant relationship when treating chromosomes as replicates (Figure 2f). Higher repeat content is also often

360   associated with chromosomal rearrangements, and for some mechanisms of inversion generation, repeats are

361   necessary. We therefore expected that genomes with more transposable elements would have more inversions, but

362   again we do not see a significant relationship (Figure 2g). We observed a clear linear relationship between the

363   genome size and TE content (data not shown) as expected, but the occurrence of inversions seemed to be correlated

364   with neither. Since neutral genomic features seem to play a relatively small role, we suggest that selection and

365   demography are key in inversion fixation.

366

367   Nᴀᴛᴜʀᴀʟ ꜱᴇʟᴇᴄᴛɪᴏɴ ᴀɴᴅ ɪɴᴠᴇʀꜱɪᴏɴꜱ

368   A consistent rate of inversion accumulation makes an underlying assumption that the inversions themselves are

369   neutral. If inversions are primarily fixed through positive selection, they likely represent a small fraction of total

370   inversion mutations, and their count would be highly dependent on the distribution of fitness effects of new

371   inversions. If inversions are instead primarily deleterious, then their fixation is dependent on the amount of genetic

372   variation, the amount of selection on heterozygotes and homozygotes as well as the population structure of the

373 species (Lande 1984). In both cases, the fixation rate is going to vary due to species specific effects (e.g., demographic
374 history) as well as inversion specific effects (e.g., inversion selection coefficient) which lead to inconsistent fixation
375 rates.

376

377 New inversions are expected to be underdominant because of disruptions to meiosis, and this should be stronger in
378 larger inversions (White 1978; Rieseberg 2001). For primarily selfing species, this should be less of a barrier because
379 of reduced heterozygosity (Hedrick 1981; Hoffmann and Rieseberg 2008). We see this play out in our ANOVA of
380 mating system which showed the highest amount of the genome in inversions for comparisons involving selfing
381 species (Supplementary Figure 4).

382

383 Inversions can play a role in speciation by containing reproductive isolation alleles or by causing reproductive
384 isolation itself through underdominance (Rieseberg 2001). We explored this in our dataset by asking if evidence of
385 hybridization affected the number or size of inversions and found no support. That being said, our species
386 comparisons are not necessarily sister species and do not necessarily have any geographic overlap, so they are not
387 appropriate species to test models of speciation.

388

389 The location of inversions supports the hypothesis that they are often deleterious. We found that inversions were
390 depleted for coding sequence and enriched for TEs consistently among the 32 genera (Figure 4). While we cannot
391 eliminate the possibility that this reflects underlying biases in the inversion mutation rate or the reduced
392 recombination effect after-the-fact, we think it is more likely to represent selection against inversions that disrupt
393 genes or gene expression. Despite inversions often occurring in genes, the relatively small number of inversions
394 compared to genes means that < 1 % of genes were affected by inversions in our dataset (Figure 5).

395

396 THE MECHANISM OF INVERSIONS

397 We found a relatively small proportion of inversions contained duplications near inversion breakpoints (Figure 6).
398 Taken at face value, this suggests non-homologous end-joining created a vast majority of inversions, but we cannot
399 rule out error in the determination of inversion breakpoints. The current analysis uses one-to-one mappings to
400 detect inversion regions, so if an inversion breakpoint were within a highly repetitive region, we may have
401 underestimated the inversion size. Additionally, genome assemblies struggle resolving highly repetitive regions, such
402 as centromeres (Naish et al. 2021). Although the genomes we included achieved chromosome-scale scaffolding,
403 centromeric regions are especially dynamic and repetitive which means they are highly challenging to assemble
404 correctly. Inversions can overlap with centromeres, and if breakpoints fell within centromeric regions we may not
405 accurately locate them (Kirubakaran et al. 2020; Harringmeyer and Hoekstra 2022; Huang et al. 2022).

406

407 Depending on the mechanism, inversion mutations either create tandem duplicates at their breakpoints, or require
408 them to be already present. With our current analysis, we cannot tell which is happening for each inversion, but
409 future studies will be able as more chromosome-level genomes are produced. With multiple species genomes for a
410 single genus, it will be possible to determine the derived and ancestral state for inversions and answer whether

411 tandem duplications are a cause or consequence of inversions. Most current methods of genome comparison focus
412 on pairwise comparisons (e.g., SyRI) but promising new methods using genomes graphs should allow for evolutionary
413 analyses of genome structure across multiple species (Garrison and Guarracino 2022).

414

CONCLUSION:

416 Our results reject clock-like fixation of inversions in plants, and support early theoretical work that emphasized the
417 central importance of selective and demographic effects on inversion fixation rates (Lande 1984). What we don't
418 know is what proportion of the inversions fixed were under positive selection versus weak negative selection. Future
419 projects should combine population level sampling with comparative genomics to determine how often newly fixed
420 inversions show signs of selection.

421

REFERENCES

423 Barth, J. M. I., P. R. Berg, P. R. Jonsson, S. Bonanomi, H. Corell, J. Hemmer-Hansen, K. S. Jakobsen, K. Johannesson, P.
424   E. Jorde, H. Knutsen, P.-O. Moksnes, B. Star, N. C. Stenseth, H. Svedäng, S. Jentoft, and C. André. 2017. Genome
425   architecture enables local adaptation of Atlantic cod despite high connectivity. Mol. Ecol. 26:4452–4466. John
426   Wiley & Sons, Ltd.
427 Bhutkar, A., S. W. Schaeffer, S. M. Russo, M. Xu, T. F. Smith, and W. M. Gelbart. 2008. Chromosomal rearrangement
428   inferred from comparisons of 12 Drosophila genomes. Genetics 179:1657–1680.
429 Bosse, M., H.-J. Megens, M. F. L. Derks, Á. M. R. de Cara, and M. A. M. Groenen. 2019. Deleterious alleles in the
430   context of domestication, inbreeding, and selection. Evol. Appl. 12:6–17.
431 Burssed, B., M. Zamariolli, F. T. Bellucco, and M. I. Melaragno. 2022. Mechanisms of structural chromosomal
432   rearrangement formation. Mol. Cytogenet. 15:1–15. BioMed Central.
433 Casals, F., and A. Navarro. 2007. Chromosomal evolution: Inversions: The chicken or the egg? Heredity (Edinb).
434   99:479–480.
435 Charlesworth, B., and D. Charlesworth. 1973. Selection of new inversions in multi-locus genetic systems. Genet. Res.
436   21:167–183.
437 Dobzhansky, T. 1933. On the Sterility of the Interracial Hybrids in Drosophila Pseudoobscura. Proc. Natl. Acad. Sci.
438   19:397–403.
439 Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed Phylogenetics and Dating with Confidence.
440   PLOS Biol. 4:e88. Public Library of Science.
441 Ellinghaus, D., S. Kurtz, and U. Willhoeft. 2020. LTRharvest, an efficient and flexible software for de novo detection
442   of LTR retrotransposons. BMC Bioinformatics 9.
443 Exposito-Alonso, M., C. Becker, V. J. Schuenemann, E. Reiter, C. Setzer, R. Slovak, B. Brachi, J. Hagmann, D. G. Grimm,
444   J. Chen, W. Busch, J. Bergelson, R. W. Ness, J. Krause, H. A. Burbano, and D. Weigel. 2018. The rate and potential
445   relevance of new mutations in a colonizing plant lineage. PLOS Genet. 14:e1007155. Public Library of Science.
446 Flynn, J. M., R. Hubley, C. Goubert, J. Rosen, A. G. Clark, C. Feschotte, and A. F. Smit. 2020. RepeatModeler2 for
447   automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. 117:9451–9457.
448   Proceedings of the National Academy of Sciences.
449 Garrison, E., and A. Guarracino. 2022. Unbiased pangenome graphs. bioRxiv 2022.02.14.480413.
450 Goel, M., H. Sun, W. B. Jiao, and K. Schneeberger. 2019. SyRI: Finding genomic rearrangements and local sequence
451   differences from whole-genome assemblies. Genome Biol. 20:1–13. Genome Biology.
452 Hager, E., O. Harringmeyer, T. Wooldridge, S. Theingi, J. Gable, S. McFadden, B. Neugeboren, K. Turner, J. Jensen,
453   and H. Hoekstra. 2022. A chromosomal inversion contributes to divergence in multiple traits between deer
454   mouse ecotypes. Science (80-.). 337:399–405.

455 Harringmeyer, O. S., and H. E. Hoekstra. 2022. Massive inversion polymorphisms shape the genomic landscape of
456   deer mice. bioRxiv.

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

458 Henderson, I. R. 2012. Control of meiotic recombination frequency in plant genomes. Curr. Opin. Plant Biol. 15:556–
459   561.

460 Ho, S., R. Lanfear, L. Bromham, M. Phillips, J. Soubrier, A. Rodrigo, and A. Cooper. 2011. Time-dependent rates of
461   molecular evolution. Mol. Ecol. 20:3087–3101. John Wiley & Sons, Ltd.

462 Hoffmann, A. A., and L. H. Rieseberg. 2008. Revisiting the impact of inversions in evolution: From population genetic
463   markers to drivers of adaptive shifts and speciation? Annu. Rev. Ecol. Evol. Syst. 39:21–42.

464 Huang, K., R. L. Andrew, G. L. Owens, K. L. Ostevik, and L. H. Rieseberg. 2020. Multiple chromosomal inversions
465   contribute to adaptive divergence of a dune sunflower ecotype. Mol. Ecol. 29:2535–2549.

466 Huang, K., K. L. Ostevik, C. Elphinstone, M. Todesco, N. Bercovich, G. L. Owens, and L. H. Rieseberg. 2022. Mutation
467   Load in Sunflower Inversions Is Negatively Correlated with Inversion Heterozygosity. Mol. Biol. Evol.
468   39:msac101.

469 Huang, K., and L. H. Rieseberg. 2020. Frequency, Origins, and Evolutionary Role of Chromosomal Inversions in Plants.
470   Front. Plant Sci. 11:1–13.

471 Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White,
472   E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K.
473   Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya, J. Baldwin, T. Bloom, D. B. Jaffe, R. Nicol, J.
474   Wilkinson, E. S. Lander, F. Di Palma, K. Lindblad-Toh, D. M. Kingsley, and B. I. G. S. P. & W. G. A. Team. 2012.
475   The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.

476 Kirkpatrick, M., and N. Barton. 2006. Chromosome inversions, local adaptation and speciation. Genetics 173:419–
477   434.

478 Kirubakaran, T. G., Ø. Andersen, M. Moser, M. Árnyasi, P. McGinnity, S. Lien, and M. Kent. 2020. A Nanopore Based
479   Chromosome-Level Assembly Representing Atlantic Cod from the Celtic Sea. G3 Genes|Genomes|Genetics
480   10:2903–2910.

481 Lande, R. 1984. The Expected Fixation Rate of Chromosomal Inversions. Evolution (N. Y). 38:743–752. [Society for
482   the Study of Evolution, Wiley].

483 Leitwein, M., J. C. Garza, and D. E. Pearse. 2017. Ancestry and adaptive evolution of anadromous, resident, and
484   adfluvial rainbow trout (Oncorhynchus mykiss) in the San Francisco bay area: application of adaptive genomic
485   variation to conservation in a highly impacted landscape. Evol. Appl. 10:56–67. John Wiley & Sons, Ltd.

486 Li, H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100.

487 Li, H. 2021. New strategies to improve minimap2 alignment accuracy. Bioinformatics 37:4572–4574.

488 Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome
489   Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics
490   25:2078–2079.

491 Lloyd, A. 2022. Crossover patterning in plants. Plant Reprod., doi: 10.1007/s00497-022-00445-4. Springer Berlin
492   Heidelberg.

493 Lowry, D. B., and J. H. Willis. 2010. A widespread chromosomal inversion polymorphism contributes to a major life-
494   history transition, local adaptation, and reproductive isolation. PLoS Biol. 8.

495 Mérot, C., R. A. Oomen, A. Tigano, and M. Wellenreuther. 2020. A Roadmap for Understanding the Evolutionary
496   Significance of Structural Genomic Variation. Trends Ecol. Evol. 35:561–572. Elsevier Ltd.

497 Naish, M., M. Alonge, P. Wlodzimierz, A. J. Tock, B. W. Abramson, A. Schmucker, T. Mandakova, B. Jamge, C. Lambing,
498   P. Kuo, N. Yelina, N. Hartwick, K. Colt, L. M. Smith, J. Ton, T. Kakutani, R. A. Martienssen, K. Schneeberger, M.
499   A. Lysak, F. Berger, A. Bousios, T. P. Michael, M. C. Schatz, and I. R. Henderson. 2021. The genetic and epigenetic
500   landscape of the Arabidopsis centromeres. Science (80-.). 374. Univ Cambridge, Dept Plant Sci, Downing St,
501   Cambridge CB2 3EA, England.

502 Ohta, T., and K. I. Kojima. 1968. Survival probabilities f new inversions in large populations. Biometrics 24:501–516.
503   United States.

504 Ou, S., and N. Jiang. 2019. LTR_FINDER_parallel: parallelization of LTR_FINDER enabling rapid identification of long
505   terminal repeat retrotransposons. Mob DNA 10:48.

506   Ou, S., and N. Jiang. 2018. LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal
507   Repeat Retrotransposons. Plant Physiol. 176:1410–22.
508   Ou, S., W. Su, Y. Liao, K. Chougule, J. R. A. Agda, A. J. Hellinga, C. S. B. Lugo, T. A. Elliott, D. Ware, T. Peterson, N. Jiang,
509   C. N. Hirsch, and M. B. Hufford. 2019. Benchmarking transposable element annotation methods for creation
510   of a streamlined, comprehensive pipeline. Genome Biol. 20:1–18. Genome Biology.
511   Pucker, B., I. Irisarri, J. de Vries, and B. Xu. 2022. Plant genome sequence assembly in the era of long reads: Progress,
512   challenges and future directions. Quant. Plant Biol. 3.
513   Quinlan, A. R., and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.
514   Bioinformatics 26:841–842.
515   Ranz, J. M., F. Casals, and A. Ruiz. 2001. How malleable is the eukaryotic genome? Extreme rate of chromosomal
516   rearrangement in the genus Drosophila. Genome Res. 11:230–239.
517   Ranz, J. M., D. Maurin, Y. S. Chan, M. von Grotthuss, L. W. Hillier, J. Roote, M. Ashburner, and C. M. Bergman. 2007.
518   Principles of genome evolution in the Drosophila melanogaster species group. PLoS Biol. 5:e152–e152. Public
519   Library of Science.
520   Rieseberg, L. H. 2001. Chromosomal rearrangements and speciation. TRENDS Ecol. Evol. 16:351–358.
521   Shi, J., and C. Liang. 2019. Generic Repeat Finder: A High-Sensitivity Tool for Genome-Wide De Novo Repeat
522   Detection. Plant Physiol. 180:1803–1815.
523   Sodeland, M., S. Jentoft, P. E. Jorde, M. Mattingsdal, J. Albretsen, A. R. Kleiven, A. E. W. Synnes, S. H. Espeland, E. M.
524   Olsen, C. Andrè, N. C. Stenseth, and H. Knutsen. 2022. Stabilizing selection on Atlantic cod supergenes through
525   a millennium of extensive exploitation. Proc. Natl. Acad. Sci. U. S. A. 119.
526   Sturtevant, A. H., and K. Mather. 1938. The Interrelations of Inversions, Heterosis and Recombination. Am. Nat.
527   72:447–452. [University of Chicago Press, American Society of Naturalists].
528   Todesco, M., G. L. Owens, N. Bercovich, J. S. Légaré, S. Soudi, D. O. Burge, K. Huang, K. L. Ostevik, E. B. M. Drummond,
529   I. Imerovski, K. Lande, M. A. Pascual-Robles, M. Nanavati, M. Jahani, W. Cheung, S. E. Staton, S. Muños, R.
530   Nielsen, L. A. Donovan, J. M. Burke, S. Yeaman, and L. H. Rieseberg. 2020. Massive haplotypes underlie ecotypic
531   differentiation in sunflowers. Nature 584:602–607. Springer US.
532   Trickett, A. J., and R. K. Butlin. 1994. Recombination suppressors and the evolution of new species. Heredity (Edinb).
533   73:339–345.
534   Wang, L., Y. Ji, Y. Hu, H. Hu, X. Jia, M. Jiang, X. Zhang, L. Zhao, Y. Zhang, Y. Jia, C. Qin, L. Yu, J. Huang, S. Yang, L. D.
535   Hurst, and D. Tian. 2019. The architecture of intra-organism mutation rate variation in plants. PLOS Biol.
536   17:e3000191. Public Library of Science.
537   White, M. J. D. 1973. Animal Cytology and Evolution. Cambridge University Press, Cambridge; London; New York.
538   White, M. J. D. 1978. Modes of Speciation. W.H. Freeman & Company, San Francisco.
539   Xiong, W., L. He, J. Lai, H. . Dooner, and C. Du. 2014. HelitronScanner uncovers a large overlooked cache of Helitron
540   transposons in many plant genomes. Proc. Natl. Acad. Sci. United States Am. 111:10263–8.
541   Xu, Z., and H. Wang. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons.
542   Nucleic Acids Res. 35.
543   Zhang, L., R. Reifová, Z. Halenková, and Z. Gompert. 2021. How important are structural variants for speciation?
544   Genes (Basel). 12.
545   Zhang, R. G., Z. . Wang, S. Ou, and G. Y. Li. 2019. TEsorter: lineage-level classification of transposable elements using
546   conserved protein domains. bioRxiv 800177.
547   Zhang, Z., S. Schwartz, L. Wagner, and W. Miller. 2000. A greedy algorithm for aligning DNA sequences. J. Comput.
548   Biol. 7:203–214. United States.
549
550
551

552   SUPPLEMENTARY FILES:

553

554   Supplementary Table 1: Species chosen for analysis and information collected, with full reference list for all genome
555   assemblies included in the study.

556

557   Supplementary Figure 1: Distribution of inverted (blue) and syntenic (red) regions identified by SyRI in 32 species
558   pair. The density of the sequence similarity score is plotted according to the corresponding nucleotide sequence
559   identity where 1.00 = 100% identical.

560

561   Supplementary Figure 2: a) syntenic region sequence identity score distribution with vertical line representing the
562   mean, b) inverted region sequence identity score distribution for the studied 32 species pair.

563

564   Supplementary Figure 3: Sequence divergence does not explain the proportion of the genome in inversions (p=0.506).

565

566   Supplementary Figure 4: The number of inversions, or proportion of the genome inverted for factors that may
567   influence the accumulation of inversions (n=32). Hybridization (red), reproductive strategy (yellow), assembly
568   methods parameters (blue).

569