Ancestral reconstruction of sunflower karyotypes reveals dramatic chromosomal evolution

Kate L. Ostevik[1,2], Kieran Samuk[1], and Loren H. Rieseberg[2]

1. Department of Biology, Duke University, Durham, NC, 27701
2. Department of Botany, University of British Columbia, Vancouver, BC, Canada, V6T 1Z4

Correspondence to: kate.ostevik@gmail.com

Abstract

Mapping the chromosomal rearrangements between species can inform our understanding of genome evolution, reproductive isolation, and speciation. Here we present a systematic survey of chromosomal rearrangements in the annual sunflowers, which is a group known for extreme karyotypic diversity. We build high-density genetic maps for two subspecies of the prairie sunflower, *Helianthus petiolaris* ssp. *petiolaris* and *H. petiolaris* ssp. *fallax.* Using a novel algorithm implemented in the accompanying R package *syntR*, we identify blocks of synteny between these two subspecies and previously published high-density genetic maps. We reconstruct ancestral karyotypes for annual sunflowers using those synteny blocks and conservatively estimate that there have been 9.7 chromosomal rearrangements per million years – a high rate of chromosomal evolution. Although the rate of inversion is even higher than the rate of translocation in this group, we further find that every extant karyotype is distinguished by between 1 and 3 translocations involving only 8 of the 17 chromosomes. This non-random sampling suggests that certain chromosomes are prone to translocation and may thus contribute disproportionately to widespread hybrid sterility in sunflowers. These data deepen our understanding of chromosome evolution and confirm that *Helianthus* has an exceptional rate of chromosomal rearrangement that likely facilitates similarly rapid diversification.

Introduction

Organisms vary widely in the number and arrangement of their chromosomes – i.e. their karyotype. Interestingly, karyotypic differences are often associated with species boundaries and, therefore, suggest a link between chromosomal evolution and speciation (White 1978, King 1993). Indeed, it is well established that chromosomal rearrangements can contribute to reproductive isolation. Although not a universal phenomenon, individuals heterozygous for divergent karyotypes are often sterile or inviable, illustrating a direct link to reproductive isolation (King 1987, Lai et al. 2005, Stathos and Fishman 2014). Apart from being underdominant *per se*, chromosomal rearrangements can also facilitate the evolution of reproductive barriers by extending genomic regions that are protected from

28 introgression (Noor et al. 2001, Rieseberg 2001), accumulating genetic incompatibilities (Navarro and
29 Barton 2003), and simplifying reinforcement (Trickett and Butlin 1994).
30
31 Despite their prevalence, the fixation of underdominant chromosomal rearrangements is still not well
32 understood. The simplest model is that in sufficiently small populations, underdominant
33 rearrangements can be stochastically fixed by genetic drift (Lewis 1966, Templeton 1981, Grant 1981).
34 However, several other mechanisms have been proposed to explain chromosomal evolution (Rieseberg
35 2001). These include individually neutral chromosome fissions and fusions (Baker and Bickham 1986,
36 Britton-Davidian 2000), meiotic drive (Novitski 1951, White 1978, Chmátal et al. 2014) and association
37 with locally adapted genes (Kirkpatrick and Barton 2006, Lowry and Willis 2010). The importance of
38 each of these mechanisms will largely depend on the types of rearrangements fixed (e.g., chromosome
39 fusions, pericentric inversions, or reciprocal translocations) and the cost/benefit of the genes
40 associated with them. Consequently, mapping and characterizing chromosomal rearrangements is a
41 critical step towards understanding their evolutionary dynamics.
42
43 The genus *Helianthus* is well known to have particularly labile genome structure. These sunflowers
44 have several paleopolyploidy events in their evolutionary history (Barker et al. 2008), have given rise to
45 three homoploid hybrid species (Rieseberg 1991), and are prone to transposable element activity
46 (Kawakami et al. 2011, Staton et al. 2012). Evidence in the form of hybrid pollen inviability, abnormal
47 chromosome pairings during meiosis, and genetic map comparisons suggests that *Helianthus*
48 karyotypes are extremely diverse (Heiser 1947, Heiser 1951, Heiser 1961, Whelan 1979, Chandler
49 1986, Rieseberg et al. 1995, Quillet et al. 1995, Burke et al. 2004, Heesacker et al. 2009, Barb et al.
50 2014). In fact, annual sunflowers have one of the highest described rates of chromosomal evolution
51 across all plants and animals (Burke et al. 2004). Furthermore, while *Helianthus* species generally have
52 large effective population sizes (but see Stebbins and Daly 1961, Carney et al. 2000), where drift should
53 be relatively weak, chromosomal rearrangements appear to be strongly underdominant in this group
54 (Chandler 1986, Lai et al. 2005). This means that chromosomal evolution is less likely to be driven by
55 drift alone and more likely to be an important driver of reproductive isolation in sunflower.
56

57 Studying chromosomal evolution within any group requires high-density genetic maps and a method
58 for systematic comparison of karyotypes. Recently, high-density genetic maps of *H. niveus* ssp.
59 *tephrodes* and *H. argophyllus* were built and compared to *H. annuus* (Barb et al. 2014). This analysis
60 clearly mapped previously inferred karyotypes (Heiser 1951, Chandler 1986, Quillet et al. 1995) and
61 showed that chromosomal rearrangements affect introgression between sunflowers. That said,
62 creating additional high-density maps for related species would yield a fuller picture of chromosomal
63 evolution.
64
65 Previously constructed genetic maps with limited marker density suggest that several chromosomal
66 rearrangements differentiate *H. annuus* and *H. petiolaris* (Rieseberg et al. 1995, Burke et al. 2004) and
67 evidence from cytological surveys suggests that subspecies within *H. petiolaris* subspecies carry
68 divergent karyotypes (Heiser 1961). Creating high-density genetic maps for *H. petiolaris* and adding
69 them to the above analysis will allow us to (1) precisely track additional rearrangements and (2)
70 reconstruct ancestral karyotypes for the group, untangling overlapping rearrangements that can be
71 obscured by directly comparing present-day karyotypes and that are likely to occur in lineages with
72 extensive chromosomal rearrangement like sunflower. At the same time, these *H. petiolaris* maps will
73 be useful to breeders looking to introgress beneficial alleles from crop wild relatives into domesticated
74 sunflower lineages (Chetelat and Meglic 2000, Foulongne et al. 2003, Dirlewanger et al. 2004).
75
76 A key part of a multi-species comparative study of chromosome evolution is a systematic and
77 repeatable method for identifying syntenic chromosomal regions (*sensu* Pevzner and Tesler 2003)
78 using genetic map data. This is especially important for cases with high marker density, because
79 breakpoints between synteny blocks can be blurred by mapping error, micro-rearrangements, and
80 paralogy (Hackett and Broadfoot 2003, Choi et al. 2007, Barb et al. 2014, Bilton et al. 2018). In previous
81 studies, synteny blocks have been found by counting all differences in marker order (Wu and Tanksley
82 2010), by visual inspection (Burke et al. 2004), or by manually applying simple rules like size thresholds
83 (Heesacker et al. 2009, Barb et al. 2014) and Spearman's rank comparisons (Berdan et al 2014,
84 Schlautman et al 2017). However, these methods become intractable and prone to error when applied
85 to very dense genetic maps. Furthermore, to our knowledge, there is no software available that

identifies synteny blocks based on relative marker positions alone (i.e., without requiring reference genomes, sequence data, or markers with known orientations). Therefore, there is a need for software that implements an algorithm to systematically and repeatably identify synteny blocks from any number of paired genetic map positions.

Here, with the goal of understanding chromosome evolution in *Helianthus* and more generally, we aimed to: (1) build high density genetic maps for two subspecies of *Helianthus petiolaris*, (2) develop a method and software to systematically identify chromosomal rearrangements from genetic maps, (3) reconstruct ancestral karyotypes for a subsection of annual sunflowers, and (4) detect general patterns of chromosomal rearrangement in *Helianthus.*

Methods

Study system

We focused on five closely related diploid (n = 17) taxa from the annual clade of the genus *Helianthus* (Fig 1). These sunflowers are native to North America and are naturally self-incompatible, although domesticated lineages of *H. annuus* are self-compatible.

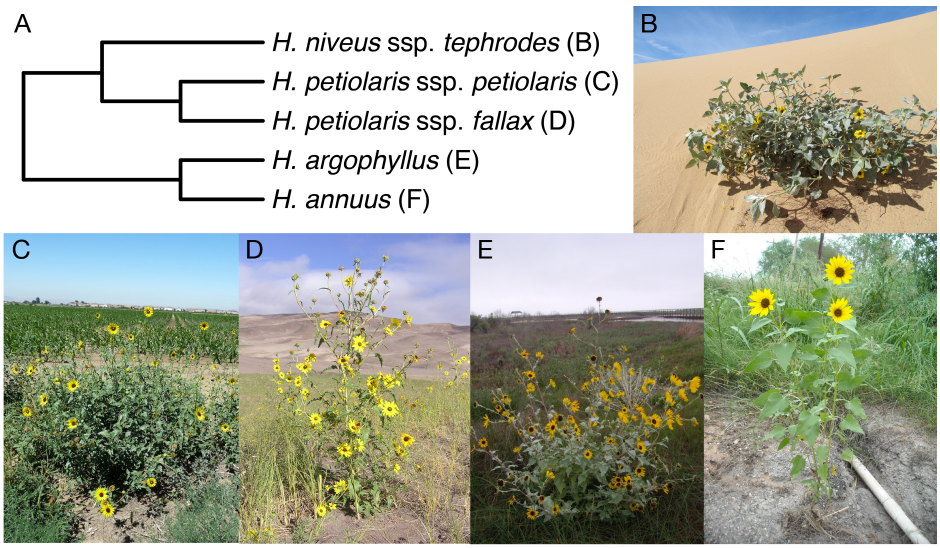


Fig 1 - The phylogenic relationships between the sunflower taxa used in this study. A) Phylogenic relationships are based on Stephens et al. (2015) and Baute et al. (2016). The divergence times between *H. petiolaris*, *H. niveus* ssp. *tephrodes* and *H. argophyllus* and *H. annuus* are estimated at 1.8 MYA (Sambatti et al. 2012), 1.89 MYA (0-5.25; Mason 2018) and 1.54 MYA (0-4.96, Mason 2018) respectively. Photo credits - B, C, E & F: Brook Moyers, D: Rose Andrew.

The five taxa are found in the United States (Fig S7, Rogers et al. 1982). *Helianthus annuus* is found throughout much of the central United States and commonly occurs in somewhat heavy soils and along roadsides (Heiser 1947). With a similarly widespread range, *H. petiolaris* occurs in sandier soils and is made up of two subspecies: *H. petiolaris* ssp. *petiolaris*, which is commonly found in the southern Great Plains, and *H. petiolaris* ssp. *fallax*, which is limited to more arid regions in Colorado, Utah, New Mexico, and Arizona (Heiser 1961). Where *H. petiolaris* is sympatric with *H. annuus,* gene flow occurs between the species (Strasburg and Rieseberg 2008). *Helianthus argophyllus* is primarily found along the east coast of Texas where it also overlaps and hybridizes with *H. annuus* (Baute et al. 2016). Finally, *H. niveus* ssp. *tephrodes* is a facultative perennial that grows in sandy dunes from the southwestern US into Mexico.

Controlled crosses

To make genetic maps, we crossed an outbred individual with presumably high heterozygosity from each *H. petiolaris* subspecies to a homozygous line of domesticated sunflower and genotyped the resulting F1 offspring. Specifically, we used pollen from a single *H. petiolaris ssp. petiolaris* plant (PI435836) and a single *H. petiolaris ssp. fallax* plant (PI435768) to fertilize individuals of a highly inbred and male sterile line of *H. annuus* (HA89cms). This design allows us to infer where recombination occurred in the heterozygous parents because we can reliably track the segregation of those parents' alleles against a predictable background.

The self incompatible *H. petiolaris* accessions were collected in central Colorado (PI435836, 39.741°, -105.342°, Boulder County) and the southeast corner of New Mexico (PI435768, 32.3°, -104.0°, Eddy County, Fig S7) and were maintained at large population sizes by the United States Department of

134 Agriculture. When it was originally collected, accession PI435768 was classified *H. neglectus*. However,
135 based on the location of the collection (Heiser 1961) and a more recent genetic analysis of the scale of
136 differences between *H. petiolaris* ssp. *fallax* and *H. neglectus* (Raduski et al. 2010), we believe that this
137 accession should be classified *H. petiolaris ssp. fallax*.

138 Genotyping

139

140 We collected leaf tissue from 116 *H. annuus* x *H. petiolaris* ssp. *petiolaris* F1 seedlings and 132 *H.*
141 *annuus* x *H. petiolaris* ssp. *fallax* F1 seedlings. We extracted DNA using a modified CTAB protocol
142 (Doyle and Doyle 1987) and prepared individually barcoded genotyping-by-sequencing (GBS) libraries
143 using a version the Poland et al. (2012) protocol. Our modified protocol includes steps to reduce the
144 frequency of high-copy fragments (e.g. chloroplast and repetitive sequence) based on Shagina et al.
145 (2010) and Matvienko et al. (2013) and steps to select specific fragment sizes for sequencing (see
146 Ostevik 2016 appendix B for the full protocol).

147

148 Briefly, we digested 100ng of DNA from each individual with restriction enzymes (either *Pst*I-HF or *Pst*I-
149 HF and *Msp*I) and ligated individual barcodes and common adapters to the digested DNA. We pooled
150 barcoded fragments from up to 192 individuals, cleaned and concentrated the libraries using SeraMag
151 Speed Beads made in-house (Rohland and Reich 2012), and amplified fragments using 12 cycles of PCR.
152 To deplete high-copy fragments, we used the following steps (M. Todesco unpublished): (1) denature
153 the libraries using high temperatures, (2) allow the fragments to re-hybridize, (3) digest the double
154 stranded fragments with duplex specific nuclease (Zhulidov et al 2004), and (4) amplify the undigested
155 fragments using another 12 cycles of PCR. We ran the libraries out on a 1.5% agarose gel and extracted
156 300-800 bp fragments using a Zymoclean Gel DNA Recovery kit (Zymo Research, Irvine, USA). Then,
157 following additional library cleanup and quality assessment, we sequenced paired-ends of our libraries
158 on an Illumina HiSeq 2000 (Illumina Inc., San Diego, CA, USA).

159

160 To call variants, we used a pipeline that combines the Burrows–Wheeler Aligner version 0.7.15 (BWA,
161 Li & Durbin 2010) and the Genome Analysis Toolkit version 3.7 (GATK, McKenna et al 2010). First, we

demultiplexed the data using sabre (https://github.com/najoshi/sabre, Accessed 27 Jan 2017). Next, we aligned reads to the *H. annuus* reference (HanXRQr1.0-20151230, Badouin et al. 2017) with 'bwa-mem' (Li 2013), called variants with GATK 'HaplotypeCaller', and jointly genotyped all samples within a cross type with GATK 'GentypeGVCFs'. We split variants into SNPs and indels and filtered each marker type using hard-filtration criteria suggested in the GATK best practices (DePristo et al. 2011, Van der Auwera et al. 2013). Specifically, we removed SNPs that had quality by depth scores (QD) less than 2, strand bias scores (FS) greater than 60, mean mapping quality (MQ) less than 40, or allele mapping bias scores (MQRankSum) less than -12.5 and indels that had QD < 2 or FS > 200. After further filtering variants for biallelic and triallelic markers with genotype calls in at least 50% of individuals, we used GATK 'VariantsToTable' to merge SNPs and indels into a single variant table for each cross type.

Finally, we converted our variant tables into AB format, such that the heterozygous parents contribute 'A' and 'B' alleles to offspring, while the *H. annuus* parent contributes exclusively 'A' alleles. At biallelic markers, sites with two reference alleles became 'AA' and sites with the reference allele and the alternate allele became 'AB'. At triallelic markers, sites with the reference allele and one alternate allele became 'AA' and sites with the reference allele and the other alternate allele became 'AB'. This method randomly assigns 'A' and 'B' alleles to the homologous chromosomes in each heterozygous parent, so our genetic maps initially consisted of pairs of mirror-imaged linkage groups that we later merged.

Genetic mapping

We used R/qtl (Broman et al. 2003) in conjunction with R/ASMap (Taylor and Butler 2017) to build genetic maps. After excluding markers with less than 20% or greater than 80% heterozygosity and individuals with less than 50% of markers scored, we used the function 'mstmap.cross' with a stringent significance threshold (p.value = $1x10^{-16}$) to form conservative linkage groups. We used the function 'plotRF' to identify pairs of linkage groups with unusually high recombination fractions and the function 'switchAlleles' to reverse the genotype scores of one linkage group in each mirrored pair. We did this until reversing genotype scores no longer reduced the number of linkage groups.

190

191 Using the corrected genotypes, we made new linkage groups with only the most reliable markers.
192 Namely, we used the function 'mstmap.cross' (with the parameter values: dist.fun = "kosambi", p.value
193 = $1x10^{-6}$, noMap.size = 2, noMap.dist = 5) on markers with less than 10% missing data and without
194 significant segregation distortion. We refined the resulting linkage groups by removing (1) markers
195 with more than three double crossovers, (2) markers with aberrant segregation patterns (segregation
196 distortion more than two standard deviations above or below the mean segregation distortion of the
197 nearest 20 markers), and (3) linkage groups made up of fewer than four markers.
198

199 We progressively pushed markers with increasing amounts of segregation distortion and missing data
200 into the maps using the function 'pushCross'. After adding each batch of markers, we reordered the
201 linkage groups and dropped markers and linkage groups as described above. Once all the markers had
202 been pushed back, we used the function 'calc.errorlod' to identify possible genotyping errors (error
203 scores greater than 2), and replaced those genotypes with missing data. We continued to drop linkage
204 groups, markers, and genotypes that did not meet our criteria until none remained.
205

206 Finally, we dropped five excess linkage groups, each made up of fewer than 30 markers, from each
207 map. The markers in these linkage groups mapped to regions of the *H. annuus* genome that were
208 otherwise represented in the final genetic maps but could not be explained by reversed genotypes.
209 Instead, it is likely that these markers were polymorphic in the HA89cms individual used for crosses
210 because of the 2-4% residual heterozygosity in sunflower inbred lines (Mandel et al. 2013).

211 SyntR development

212

213 To aid in the identification of chromosomal rearrangements, we developed the R package 'syntR'
214 (code and documentation available at http://ksamuk.github.io/syntR). This package implements a
215 heuristic algorithm for systematically detecting synteny blocks from marker positions in two genetic
216 maps. In short, syntR combines a biologically-informed noise reduction method and a cluster
217 identification method better suited for detecting linear (as opposed to circular) clusters of data points.

218 The algorithm compares sets of homologous genetic markers from two genetic maps, clusters the
219 markers into synteny blocks using a method that is similar to the "friends-of-friends" clustering
220 algorithm (Huchra and Geller 1982), and identifies breakpoints. The algorithm correctly recovers
221 breakpoints in simulated genetic maps under biologically reasonable scenarios and performs at least
222 well as prior methods while being systematic and repeatable. We fully describe this method and its
223 performance in the supplemental materials.

224 Finding synteny blocks

225

226 We used syntR to identify synteny blocks between our newly generated genetic maps and an ultra-
227 high-density map of *H. annuus* (Badouin et al. 2017). The *H. annuus* genetic map was used to build the
228 sunflower genome that we use as a reference (Badouin et al. 2017). This allowed us to easily convert
229 between physical position in the *H. annuus* reference and position in the *H. annuus* genetic map. Using
230 this property, we further compared two previously published genetic maps for the closely related
231 sunflower species, *H. niveus* ssp. *tephrodes* and *H. argophyllus* (Barb et al. 2014), to the same *H.*
232 *annuus* map. We aligned the published maps' marker sequences to the *H. annuus* reference using bwa
233 and converted the definitively aligned markers (MQ > 40) to their positions in the *H. annuus* genetic
234 map.

235

236 Initially, we ran syntR on each individual map comparison using the parameters that maximized the
237 genetic distance (i.e. coverage of the genome) represented across synteny blocks (Table 1) given that
238 synteny blocks should represent all positions in the genetic maps being compared (Chen et al. 2009).
239 However, varying the tuning parameters of syntR resulted in the identification of additional
240 rearrangements, which were clearly shared between the maps (Fig S8). Therefore, we ran syntR using
241 values of the parameter 'max_clust_dist' (see supplemental material for details) that covered the
242 range of best parameters for each species (max_clust_dist = 1, 1.5, 2, 2.5, 3, 3.5) and extracted a
243 curated set of synteny blocks from the output using the following hierarchical criteria. A synteny block
244 was retained if it: (1) was found in another species, (2) was identified in the majority of syntR runs for a
245 single species, (3) maximized the genetic distance represented by synteny blocks. We present this

246 curated set of synteny blocks below, but our results are unchanged if we used only the synteny blocks
247 recovered with the best fit for each map individually.
248
249 Table 1 – The syntR tuning parameters used to identify synteny blocks between our genetic maps and *H. annuus*.

Taxon	max_clust_range	max_nn_dist	min_block_size
H. petiolaris ssp. *petiolaris*	3.0	10	3
H. petiolaris ssp. *fallax*	2.5	10	3
H. niveus ssp. *tephrodes*	1.5	10	3
H. argophyllus	1.5	10	3

250
251 We named the chromosomes in our genetic maps based on their synteny with the standard order and
252 orientation of *H. annuus* chromosomes (Tang et al 2002, Bowers et al. 2012) following Barb et al.
253 (2014) but with shortened prefixes (A = *H. annuus*, R = *H. argophyllus*, N = *H. niveus* ssp. *tephrodes*, P =
254 *H. petiolaris* ssp. *petiolaris*, F = *H. petiolaris* ssp. *fallax*). For example, a *H. petiolaris* ssp. *fallax*
255 chromosome made up of regions that are syntenic with *H. annuus* chromosomes 4 and 7 is called F4-7.

256 Karyotype reconstruction and analysis

257
258 We used our inferred synteny blocks and the software MGR v 2.01 (Bourque and Pevzner 2002) to infer
259 ancestral karyotypes for our five *Helianthus* taxa and to determine the number of chromosomal
260 rearrangements that occurred along each branch of the species tree. To run the MGR analysis, we
261 needed the order and orientations of synteny blocks in all five maps. However, individual synteny
262 blocks were often missing from one or more of our final maps. We took two approaches to addressing
263 this problem. First, we inferred the likely position of missing synteny blocks based on the location of
264 markers that were too sparse to be grouped together by syntR and matched the location of synteny
265 blocks in other maps. In the second case, we simply dropped any synteny blocks that were not
266 universally represented. Because we already had two sets of synteny blocks for each map (curated and
267 individually optimized), we ran the MGR analyses using three different sets of synteny blocks: (set 1)
268 curated and inferred, (set 2) curated and present in all five maps, (set 3) individually optimized and
269 present in all five maps.

Data availability

All data and scripts are available upon request and will be sent to dryad before publication. The R program, syntR, is available on github: https://github.com/ksamuk/syntR

Results

Genetic maps

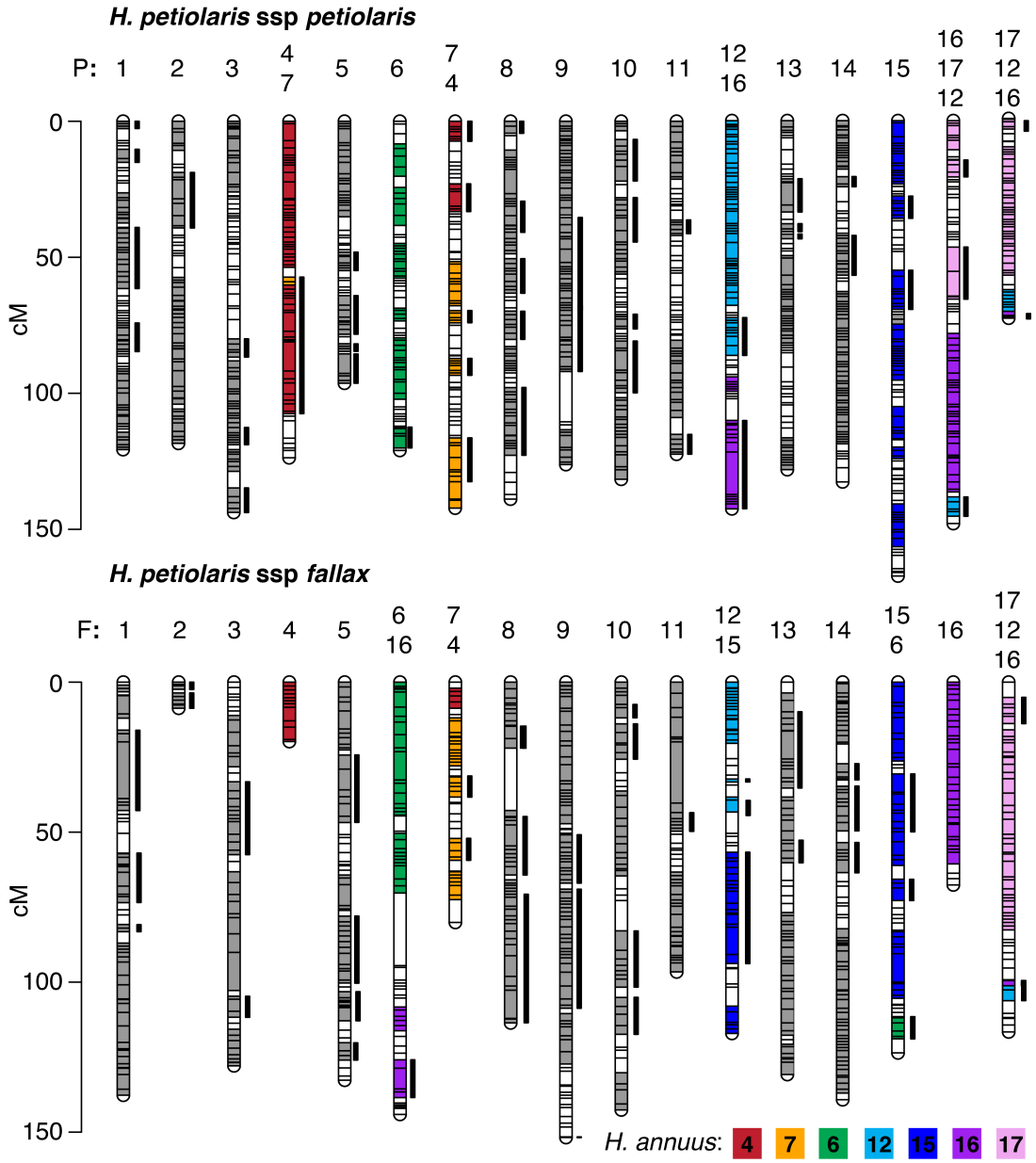
Both of the final *H. petiolaris* genetic maps are made up of the expected 17 chromosomes and have very high marker density (Table 1, Fig 2, Fig S9, Fig S10). Only 6% of the *H. petiolaris* ssp. *petiolaris* map and 10% of the *H. petiolaris* ssp. *fallax* map fails to have a maker within 2 cM. Overall, both maps are somewhat longer than the *H. petiolaris* map reported by Burke et al. (2004), but the higher marker density combined with some genotyping error likely explains this discrepancy. Indeed, building maps with variants that were thinned to 1 per 150 bp using vcftools version 0.1.13 (Danecek et al. 2011) yields collinear maps that are closer to the expected lengths (Table 1, Fig S11). We present subsequent results based on the full maps to improve our resolution for detecting smaller rearrangements.

Table 2 – Properties of new and previously published *H. petiolaris* maps.

Genetic map	N markers	Length (cM)
H. petiolaris ssp. *petiolaris* - full	8179	1850
H. petiolaris spp. *fallax* - full	13335	2178
H. petiolaris ssp. *petiolaris* - thinned	2462	1576
H. petiolaris spp. *fallax* - thinned	3368	1791
H. petiolaris - Burke et al. 2004	795	1592

Despite the general expansion of our maps, we find that chromosomes 2 and 4 in the *H. petiolaris* ssp. *fallax* map (F2 and F4) are unexpectedly short (Fig 2). When we look at the distribution of markers for this map relative to the *H. annuus* reference, we find very few variable sites in the distal half of these chromosomes. This suggests that the shortness of F2 and F4 is explained by a lack of variable sites

291 within the *H. petiolaris* ssp. *fallax* individual used for crosses (Fig S12). That is, this individual was
292 homozygous for most of F2 and F4. Notably, we find the same pattern on the distal half of *H. annuus*
293 chromosome 7 and find that this region is also not represented in the *H. petiolaris* spp. *fallax* map.



294

295 Figure 2 – *Helianthus petiolaris* genetic maps showing blocks of synteny with *H. annuus*. Each horizontal bar
296 represents a genetic marker. The thick vertical bars next to chromosomes represent synteny blocks that are
297 inverted relative to the *H. annuus* genetic map. Where there are no translocations between *H. petiolaris* and *H.*
298 *annuus* chromosomes (e.g. all synteny blocks in P1 and F1 are syntenic with A1), the synteny blocks are shown in

grey. Where there are translocations, the synteny blocks are color-coded based on their synteny with *H. annuus* chromosomes. The synteny blocks plotted are those curated based on multiple runs of syntR using different parameters. Please see Fig S13 for a labeled version. This figure was made with LinkageMapView (Ouellette et al. 2017).

Synteny blocks

Using syntR, we recovered 97 genetic regions that are syntenic between the *H. petiolaris* ssp. *petiolaris* and *H. annuus* and 79 genetic regions that are syntenic between the *H. petiolaris* ssp. *fallax* and *H. annuus* (Fig 2). We also recovered synteny blocks for the *H. petiolaris* ssp. *tephrodes* and *H. argophyllus* comparisons that are similar, although likely more conservative (see supplementary materials), to those presented in Barb et al. (2014). In all four comparisons, syntR successfully identified synteny blocks that cover large proportions (63%-90%) of their respective genetic maps even in the face of a very high proportion of markers that map to a different chromosome than their neighbours (Table 3). These "rogue markers" could be the result of very small translocations, poorly mapped markers, or extensive paralogy.

Table 3 – Properties of the synteny blocks found using a syntR analysis between genetic maps of *H. annuus* and four other *Helianthus* taxa (Fig 2, S13-S20). The proportion of rogue markers is based only on the chromosomes without translocations in any map (i.e., chromosomes 1-3, 5, 8-10, 11, and 14). For those chromosomes, the majority of marker mapped to a single *H. annuus* chromosome. The other markers are considered rogue.

Genetic map	N synteny blocks	Rogue markers	Map coverage	*H. annuus* coverage	Collinear	Inverted	Translocated
H. petiolaris ssp. *petiolaris*	97	19%	80%	74%	39%	36%	26%
H. petiolaris spp. *fallax*	79	17%	63%	65%	32%	34%	34%
H. niveus ssp. *tephrodes*	43	26%	78%	75%	40%	21%	39%
H. argophyllus	31	20%	90%	82%	45%	16%	39%

Over and above the prevalence of rogue markers, the karyotypes we recovered are substantially rearranged. Only between 32% and 45% of synteny blocks for each map are collinear with the *H. annuus* genetic map in direct comparisons (Table 2). However, because nested and shared

323 rearrangements can obscure patterns of chromosome evolution, we use the MGR analyses to predict
324 the most likely sequence of rearrangements in a phylogenetic context before quantifying
325 rearrangement rate.

326 Karotype reconstruction and chromosomal rearrangement

327

328 Using all sets of synteny block orders and orientations, our MGR analyses identified similar overall
329 patterns of chromosome evolution (Table 4). Many rearrangements are shared by multiple taxa and
330 the similarity of karyotypes matches known phylogenetic relationships. As expected, MGR analyses run
331 without a guide tree inferred the known species tree and MGR analyses run with all other topologies
332 identified an inflated number of chromosomal rearrangements.

333

334 Table 4 – Overall patterns of chromosomal rearrangement inferred based on different sets of synteny blocks. Set
335 1 synteny blocks are curated based on multiple syntR runs and inferred when missing. Set 2 synteny blocks are
336 curated but present in all five maps. Set 3 synteny blocks are the output from individually optimized syntR runs
337 that are present in all five maps. Rate 1 is the number of rearrangements per million years based on a
338 cumulative divergence time of 9.1 million years, while rate 2 is the based on 26 million years. P-value 1 is the
339 probability of seeing the observed number of inversions and translocations if the rate of inversion was equal to
340 the rate of translocation. P-value 2 is the probability of seeing the observed number of chromosomes involved in
341 translocation if all chromosomes were equally likely to be involved in a reciprocal translocation.

Synteny blocks	N blocks	N inversions	N trans.	Rate 1	Rate 2	N translocated chromosomes	P-value 1: Inversion rate = trans. rate	P-value 2: Trans. rate = across chromosomes
Set 1	97	74	14	9.7	3.4	8	5.1×10^{-11}	8.0×10^{-8}
Set 2	67	45	15	6.6	2.3	7	1.3×10^{-4}	5.3×10^{-11}
Set 3	76	50	10	6.6	2.3	7	1.6×10^{-7}	3.0×10^{-6}

342

343

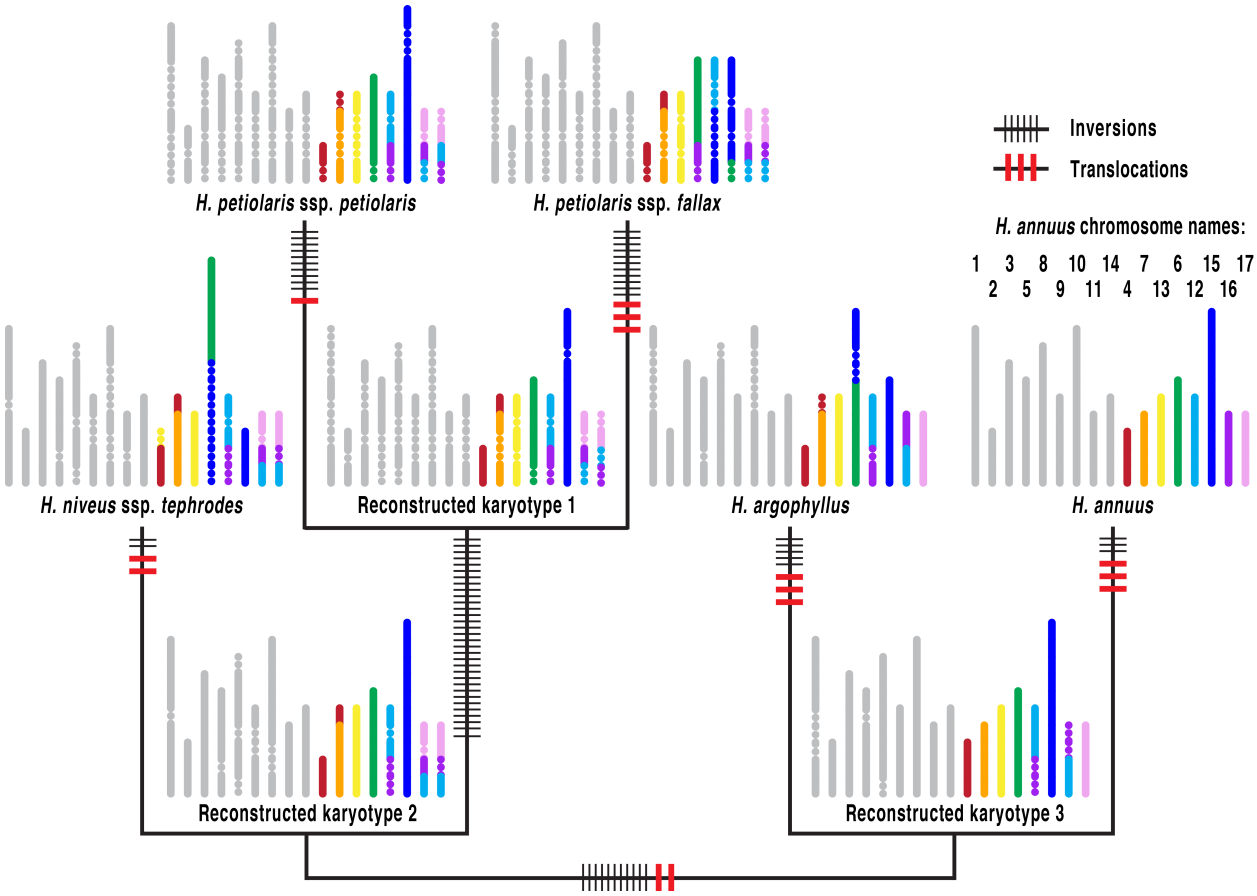


Figure 2 – Diagram of extant and ancestral karyotypes for 5 *Helianthus* taxa. The karyotypes are built using synteny block set 1, which are curated based on multiple syntR runs and inferred when missing. Each synteny block is represented using a line segment that is color-coded based on its position in the *H. annuus* genome (see Fig S21 for a labeled version). Chromosomes without translocations in any map are plotted in grey, and synteny blocks that are inverted relative to *H. annuus* are plotted using dotted lines. Also note that along some branches the same pair of chromosomes is involved in multiple translocations.

Using our most complete set of synteny blocks (set 1), we find 88 chromosomal rearrangements across the phylogeny (Fig 2). If we then use the most current divergence time estimates for this group (Mason 2018; *H. niveus* ssp. *tephrodes* and *H. annuus* = 1.89 MYA, range = 0-5.25; *H. argophyllus* and *H. annuus* = 1.54 MYA, range = 0-4.96) and conservatively assume that *H. niveus* ssp. *tephrodes* and both *H. petiolaris* subspecies arose at approximately the same time (such that they all have 1.89 Ma long branches), we find that 9.7 rearrangements occurred per million years in this clade. We cannot

358 reliability put an upper bound on this estimate because the lower bounds of divergence time estimates
359 within this groups overlap 0. However, at 2.3 rearrangements per million years, even our most
360 conservative estimate of rearrangement for this group, which uses the less complete sets of synteny
361 blocks and the upper bounds of divergence time estimates, is among the highest measurements
362 reported in Burke et al. (2004).
363
364 These 88 rearrangements are made up of 74 inversions and 14 translocations that are distributed
365 across the phylogeny. Except for the branch between reconstructed karyotypes 1 and 2, where there
366 are no translocations and many inversions, each branch has 2-12 inversions and 1-3 translocations.
367 Although inversion and translocation rates are fairly even across the tree, it is unlikely that the rate of
368 inversions is equal to the rate of translocation (binomial test, 5.1×10^{-11}). Furthermore, we find that only
369 8 of the 17 chromosomes are involved in the 14 translocations identified on the phylogeny. This
370 extreme asymmetry is highly unlikely to have happened by chance if translocation is equally likely for
371 all chromosomes ($p = 8.0 \times 10^{-8}$, Fig S22). Therefore, this pattern of rearrangement suggests that some
372 chromosomes are more likely than others to be involved in translocations. In line with this observation,
373 we see that some sections of chromosomes are frequently found in new positions. For example, we
374 see that A4 and A7 are involved in repeated translocations and that part of A6 has a different position
375 in almost every map.

376 Discussion

377
378 Large-scale chromosomal changes may be key contributors to the process of adaptation and
379 speciation, yet we still have a poor understanding of the basic evolutionary patterns and rates of
380 chromosomal rearrangement. Here, we performed comprehensive analysis of the evolution of
381 chromosomal rearrangements in a clade of sunflowers. We created multiple new genetic maps for taxa
382 in this clade, and devised a novel, systematic method for comparing these maps. Using our new
383 method, we identified a wide range of karyotypic variation in our new maps, as well as previously
384 published maps. Consistent with cytological data, we discovered an extremely high rate of
385 chromosomal evolution in this clade of plants. Here, we discuss the evolutionary and methodological

386 implications of this finding, and suggest next steps in understanding the dynamic process of
387 chromosomal rearrangement.

388 Similarity of *H. petiolaris* maps to previous studies

389

390 Compared with previous work, we found more inversions and fewer translocations between *H.*
391 *petiolaris* subspecies and *H. annuus* (Rieseberg et al. 1995, Burke et al. 2004). This is probably due to a
392 combination of factors. First, there appears to be karyotypic variation within some *Helianthus* species
393 (Heiser 1948, Heiser 1961, Chandler et al. 1986) and the individuals used to make maps in each study
394 were not identical. Second, the maps presented here are made up of more markers and individuals,
395 which allowed us to identify small inversions that were undetected in the previous comparisons, as
396 well as to eliminate false linkages that can be problematic with small mapping populations. Lastly, we
397 required more evidence to call rearrangements, especially translocations. Although we did recover
398 some of the translocations that were supported multiple markers in Rieseberg et al. (1995) and Burke
399 et al. (2004), for example a translocation between A6 and A15 in *H. petiolaris* ssp. *fallax* and the
400 fragmentation of A16 across multiple chromosomes in both *H. petiolaris* maps, we did not recover any
401 of the translocations supported by only a single sequence-based marker. Given the high proportion of
402 "rogue markers" in our maps it is likely that some of the putative translocations recovered in those
403 earlier comparisons are the result of the same phenomenon.
404
405 On the other hand, we found that rearrangements between our *H. petiolaris* maps match the
406 translocations predicted from cytological studies quite well. Heiser (1961) predicted that *H. petiolaris*
407 ssp. *petiolaris* and *H. petiolaris* ssp. *fallax* karyotypes would have three chromosomes involved in two
408 translocations that form a ring during pairing at meiosis, as well as the possibility of a second
409 independent rearrangement. This configuration of chromosomes would occur at meiosis in hybrids
410 between the *H. petiolaris* subspecies maps we present here (Fig S23). Furthermore, one of the
411 cytological configurations that included the most chromosomes in *H. petiolaris* and *H. annuus* hybrids
412 reported by Whelan (1979) and one of the most common configurations in other cytological studies
413 (Heiser 1947, Ferriera 1980, Chandler et al. 1986) was a hexavalent (a six-chromosome structure) plus

414 a quadrivalent (a four-chromosome structure). Again, this is the configuration that we would expect in
415 a hybrid between the *H. petiolaris* ssp. *petiolaris* and *H. annuus* individuals mapped here. Further, the
416 complicated arrangement and relatively small size of A12, A16 and A17 synteny blocks in our *H.*
417 *petiolaris* maps might explain why cytological chromosome configurations in hybrids between *H.*
418 *petiolaris* and *H. annuus* are so variable. Interestingly, the rearrangements identified between *H.*
419 *argophyllus* and *H. annuus* karyotypes here and in Barb et al. (2014) also match the cytological studies
420 better than an earlier comparison of sparse genetic maps (Heesacker et al. 2009). It seems that in
421 systems with the potential for high proportions of rogue markers, many markers are needed to reliably
422 identify chromosomal rearrangements.

423 Rearrangement rate

424

425 Our data suggest that annual sunflowers experience approximately 9.7 chromosomal rearrangements
426 per million years (the upper bound for this estimate is undefined because divergence time estimates
427 for the group include 0, while the lower bound is 2.3). This rate is higher than initially reported for
428 sunflower (5.5-7.3, Burke et al. 2004) and overlaps with more recent estimates for this group (7.4-10.3,
429 Barb et al. 2014). It is also much higher than recent estimates in other taxa (0.03-0.12, Wu and
430 Tanksley 2010), but is not the highest ever reported (e.g. a comparisons of genome sequence revealed
431 up to 35.7 rearrangements per million years in some grass lineages, Dvorak 2018). That said, 9.7 is very
432 likely an underestimate of the actual chromosomal rearrangement rate in sunflower for several
433 reasons. First, we were conservative in our estimates of divergence time and thresholds for calling
434 rearrangements. For example, some proportion of the rogue markers that we identified could, in fact,
435 be the result of very small but real chromosomal rearrangements. Second, our ability to resolve very
436 small synteny blocks and breakpoints between synteny blocks is dependent on marker density. Until
437 we have full genome sequences to compare (like for the grass lineages), we are likely failing to detect
438 very small rearrangements and potentially infer that some independent rearrangements are shared.
439 However, regardless of just how much we are underestimating the rate, sunflower chromosomes are
440 evolving quickly. This high rate of chromosomal evolution could be a consequence of a higher rate of
441 chromosomal mutation, a decreased chance that chromosomal polymorphisms are lost, or both

442 processes. Although there are factors that could increase the rate of all types of chromosomal
443 rearrangements (e.g. repetitive sequence or small inbred populations), other factors are dependent on
444 the specific type of rearrangement.

445 Type of rearrangements

446

447 All annual sunflowers have 17 chromosomes (Chandler et al. 1986). Therefore, it seemed unlikely that
448 chromosomal evolution was caused by fissions and fusions in sunflower, and our results are consistent
449 with that expectation. Instead, we see that inversions and interchromosomal translocations dominate
450 chromosomal evolution in *Helianthus,* which is a common pattern in angiosperm lineages (Weiss-
451 Schneeweiss and Schneeweis 2012). Furthermore, we found that inversions are happening at a higher
452 rate than translocations. This is consistent with examples of inversions occurring more often than
453 translocations in other plant lineages (Wu and Tanksley 2010, Amores et al. 2014) and reports that
454 intrachromosomal rearrangements are more common than interchromosomal rearrangements
455 generally (Pevzner and Tesler 2003). The difference between the rate of inversion and translocation is
456 notable because many studies do not clearly differentiate the evolutionary dynamics expected for
457 different rearrangement types. Both inversions and translocations are associated with reduced gene
458 flow between types in sunflower (Rieseberg et al. 1999, Barb et al. 2014) and in other systems (Basset
459 et al. 2006, Yannic et al. 2009). However, translocations have a more predictable effect on hybrid
460 fertility while the main effect of inversions is to reduce recombination.
461

462 Reciprocal translocation heterozygotes affect fertility because of missegregation during meiosis and
463 are expected to cause half of gametes to be unbalanced and therefore inviable (White 1973, King
464 1993). Inversion heterozygotes can also produce unbalanced gametes, but whether that happens is
465 dependent on the size of the inversion and whether disrupted pairing during meiosis actually inhibits
466 crossovers (Searle 1993). When inversions are small or have suppressed crossing over, they will not be
467 strongly underdominant. Indeed, inversions typically are not strongly underdominant in sunflowers (L.
468 Rieseberg, pers. comm.), which might explain why translocations are less common than inversions and

469 why pollen viability is accurately predicted by the number of translocations inferred from cytological
470 studies (Chandler et al. 1986).
471
472 In the case of inversions, the consequences of reduced recombination for adaptation and speciation
473 has been extensively examined (Hoffman and Rieseberg 2008). Recombination in inversions can be
474 suppressed because of disrupted pairing (Searle 1993) or ineffective because the products are inviable
475 (Rieseberg 2001). In either case, reduced recombination can facilitate adaptation in the face of gene
476 flow because regions of low recombination (e.g. an inversion) increase the chance that new mutations
477 with positive effects on fitness will arise either linked to a region that is already diverged (Via and West
478 2008, Via 2012) or linked to other similar loci (Yeaman and Whitlock 2011, Yeaman 2013) and will
479 therefore be more likely to persist under migration-selection balance. On the other hand, in a case of
480 multiple locally adapted alleles, a mutation that reduces the production of maladaptive recombinants
481 (e.g. an inversion) can be advantageous and thereby drive low recombining regions to fixation
482 (Kirkpatrick and Barton 2006). Both processes will lead to clusters of adaptive traits inside inversions
483 and examples of this pattern are accumulating quickly (Jones et al. 2012, Lee et al. 2017, Hooper and
484 Price 2017, Fuller et al. 2017, Coughlan and Willis 2019). It is not clear whether the same pattern will
485 be common for translocations (but see Fishman et al. 2013, Stathos and Fishman 2014 for one
486 example). Mispairing at translocation breakpoints could suppress crossing over, but recombination
487 inside reciprocal translocations will not necessarily produce inviable gametes.

488 Non-random chromosomal rearrangement

489
490 We found that some sunflower chromosomes are involved in more translocations than others. A bias
491 in the identity of chromosomes involved in translocation is common in species with sex chromosomes,
492 where sex chromosomes are rearranged more often than autosomes (e.g., Neafsey et al. 2015), but
493 this is not relevant for sunflowers. However, it is also known that centromeres and other repetitive
494 regions can affect the rate of mutations that cause chromosomal rearrangements (Hardison et al.
495 2003, Vitte et al. 2014) and breakpoint reuse is a common phenomenon in comparative studies of
496 karyotypes (Bailey et al. 2004, Larkin et al. 2009, Murphy et al. 2005, Pevzner and Tesler 2003).

497 Therefore, it is possible that the overactive chromosomes identified in this study are outliers for
498 centromeric/repeat content. In fact, we see some hints that centromeric repeats might be associated
499 with repeated translocation in this clade. Using the locations of the centromere specific
500 retrotransposon sequence, HaCEN-LINE (Nagaki et al. 2015), to roughly identify the locations of
501 centromeres in our reference, we find that some rearrangement breakpoints, for example the section
502 of A16 with a different position in each map, are close to putative centromeres (Fig S24-S29).
503
504 An association between repeated translocations and centromere properties would be interesting
505 because rearrangements that change the position of centromeres might cause meiotic drive to fix
506 those rearrangements (Chmátal et al. 2014, Zanders et al. 2014). Although a more thorough analysis of
507 centromeric repeat locations and their association with rearrangement breakpoints is required to draw
508 strong conclusions about the importance of centromeres to chromosomal evolution in sunflower, the
509 development of reference sequences for wild sunflower species is underway, which will allow those
510 associations to be confirmed. Further, it is time to directly test for meiotic drive in this system by
511 examining the transmission of rearrangements that affect centromeres in gametes produced by plants
512 that have heterozygous karyotypes. However, regardless of how they arise, if translocations are
513 primarily responsible for reducing pollen viability in sunflower and certain chromosomes are
514 translocated more often, these chromosomes could be playing an outsized role in maintaining
515 divergence between sunflower lineages.

516 Identifying rearrangements

517
518 Studying the evolution of chromosomal rearrangements requires dense genetic maps and systematic
519 methods to analyze and compare these maps between species. Our new software, syntR, provides an
520 end-to-end solution for systematic and repeatable identification of synteny blocks in pairs of genetic
521 maps with any marker density. Our tests on real and simulated data find that syntR recovers
522 chromosomal rearrangements identified by both previous manual comparisons and cytological study,
523 suggesting that syntR is providing an accurate view of karyotypic differences between species.
524

525 Although the number of species with whole genome sequence and methods to detect synteny blocks
526 from those sequences are rapidly accumulating, it is still rare to have multiple closely related whole
527 genome sequences that are of sufficient quality to compare for karyotype differences. At the same
528 time, the proliferation of reduced representation genome sequencing methods means that it is easy to
529 generate many genetic markers and produce dense genetic maps. These maps coupled with this
530 software will allow us to discover many karyotypes. Furthermore, syntR allows comparisons to include
531 older genetic map data that would otherwise go unused. Overall, we believe syntR will be a valuable
532 tool for the systematic study of chromosomal rearrangements in any species.
533
534 We also believe that syntR provides a baseline for the development of further computational and
535 statistical methods for the study of chromosomal rearrangements. One fruitful direction would be to
536 integrate syntR's synteny block detection directly into the genetic map building process (much like
537 GOOGA, Flagel et al. 2019). Another key extension would be to allow syntR to compare multiple
538 genetic maps simultaneously to detect synteny blocks in a group of species (e.g. by leveraging
539 information across species). Finally, formal statistical methods for evaluating the model fit and
540 uncertainty involved with a particular inferred set of synteny blocks would be a major (albeit
541 challenging) improvement to all existing methods, including syntR.

542 Conclusion

543
544 Understanding the evolution of chromosomal rearrangements remains a key challenge in evolutionary
545 genetics. By developing new software to systematically detect synteny blocks and building new genetic
546 maps, we show that sunflowers exhibit rapid and non-random patterns of chromosomal evolution.
547 These data expand the relatively short list of taxa in which karyotype evolution as been systematically
548 studied (e.g, Wu and Tanksley 2010, Huang et al. 2009, Vogel et al. 2010, Yogeeswaran et al. 2005,
549 Ferguson-Smith and Trifonov 2007) and generate specific and testable hypotheses about chromosomal
550 evolution in sunflower. We believe that our work will spur additional studies of karyotypic evolution
551 and diversity, and ultimately lead to a more comprehensive understanding of the interplay between
552 chromosomal evolution and speciation.

553 # Acknowledgements

554

555 We thank Jessica Barb for providing marker sequence data, Marcy Uyenoyama for help with random
556 walk, Greg Baute for sharing hybrid seed, Chris Grassa for growing seedlings and sharing helpful scripts,
557 Marco Todesco and Nadia Chaidir for help in the lab and Jenn Coughlan, Brook Moyers, Mariano
558 Alvarez, Dolph Schluter, Darren Irwin, and Sally Otto for thoughtful discussions and help with earlier
559 drafts of this manuscript. This work was supported by an NSERC Postgraduate Scholarship awarded to
560 KLO and an NSERC Discovery Grant awarded to LHR (327475).

561 # Author contributions

562

563 KLO and LHR planned the study. KLO and KS designed and built the R package syntR. KLO made genetic
564 maps, carried out data analysis, and drafted the manuscript. All authors read, edited, and approved the
565 final manuscript.

566 # References

567

568 Amores A., Catchen J., Nanda I., Warren W., Walter R. *et al.*, 2014 A RAD-tag genetic map for the
569 platyfish (*Xiphophorus maculatus*) reveals mechanisms of karyotype evolution among teleost fish.
570 Genetics **197**: 625–641.

571 Badouin H., Gouzy J., Grassa C. J., Murat F., Staton S. E. *et al.*, 2017 The sunflower genome provides
572 insights into oil metabolism, flowering and Asterid evolution. Nature **175**: 1823.

573 Bailey J. A., Baertsch R., Kent W., Haussler D., Eichler E. E., 2004 Hotspots of mammalian chromosomal
574 evolution. Genome Biology **5**: R23–7.

575 Baker R. J., Bickham J. W., 1986 Speciation by monobrachial centric fusions. Proceedings of the
576 National Academy of Sciences **83**: 8245–8248.

577 Barb J. G., Bowers J. E., Renaut S., Rey J. I., Knapp S. J., et al., 2014 Chromosomal evolution and
578 patterns of introgression in *Helianthus*. Genetics **197**: 969–979.

579 Barker M. S., Kane N. C., Matvienko M., Kozik A., Michelmore R. W., et al., 2008 Multiple
580 paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate

581 gene retention after millions of years. Molecular Biology and Evolution **25**: 2445–2455.

582 Baute G. J., 2015 Genomics of sunflower improvement: From wild relatives to a global oil seed.

583 Baute G. J., Owens G. L., Bock D. G., Rieseberg L. H., 2016 Genome-wide genotyping-by-sequencing
584 data provide a high-resolution view of wild Helianthus diversity, genetic structure, and interspecies
585 gene flow. American Journal of Botany.

586 Berdan E. L., Kozak G. M., Ming R., Rayburn A. L., Kiehart R., Fuller R. C., 2014 Insight into genomic
587 changes accompanying divergence: genetic linkage maps and synteny of *Lucania goodei* and *L.*
588 *parva* reveal a Robertsonian fusion. G3: Genes| Genomes| Genetics **4**: 1363–1372.

589 Bilton T. P., Schofield M. R., Black M. A., Chagné D., Wilcox P. L., Dodds K. G., 2018 Accounting for
590 errors in low coverage high-throughput sequencing data when constructing genetic maps using
591 biparental outcrossed populations. Genetics **209**: 65–76.

592 Bourque G., Pevzner P. A., 2002 Genome-scale evolution: reconstructing gene orders in the ancestral
593 species. Genome Research **12**: 26–36.

594 Bowers J. E., Bachlava E., Brunick R. L., Rieseberg L. H., Knapp S. J., Burke J. M., 2012 Development of a
595 10,000 locus genetic map of the sunflower genome based on multiple crosses. G3 **2**: 721–729.

596 Britton-Davidian J., Catalan J., da Graça Ramalhinho M., Ganem G., Auffray J. C. *et al.*, 2000 Rapid
597 chromosomal evolution in island mice. Nature **403**: 158–158.

598 Broman K. W., Wu H., Sen S., Churchill G. A., 2003 R/qtl: QTL mapping in experimental crosses.
599 Bioinformatics **19**: 889–890.

600 Burke J. M., Lai Z., Salmaso M., Nakazato T., Tang S., Heesacker A., Knapp S. J., Rieseberg L. H., 2004
601 Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*. Genetics **167**: 449–
602 457.

603 Carney S. E., Gardner K. A., Rieseberg L. H., 2000 Evolutionary changes over the fifty-year history of a
604 hybrid population of sunflowers (Helianthus). Evolution **54**: 462–474.

605 Chandler J. M., Jan C. C., Beard B. H., 1986 Chromosomal differentiation among the annual *Helianthus*
606 species. Systematic Botany **11**: 354–371.

607 Chen Z., Fu B., Jiang M., Zhu B., 2009 On recovering syntenic blocks from comparative maps. J Comb
608 Optim **18**: 307–318.

609 Chetelat R. T., Meglic V., 2000 Molecular mapping of chromosome segments introgressed from
610 *Solanum lycopersicoides* into cultivated tomato (*Lycopersicon esculentum*). Theor Appl Genet **100**:
611 232–241.

612 Chmátal L., Gabriel S. I., Mitsainas G. P., Martínez-Vargas J., Ventura J *et al.*, 2014 Centromere strength

provides the cell biological basis for meiotic drive and karyotype evolution in mice. Current Biology **24**: 2295–2300.

Choi V., Zheng C., Zhu Q., Sankoff D., 2007 Algorithms for the extraction of synteny blocks from comparative maps. In: International Workshop on Algorithms in Bioinformatics, pp. 277–288. Springer, Berlin, Heidelberg.

Coughlan J. M., Willis J. H., 2019 Dissecting the role of a large chromosomal inversion in life history divergence throughout the *Mimulus guttatus* species complex. Mol Ecol **28**: 1343–1357.

Danecek P., Auton A., Abecasis G., Albers C. A., Banks E. *et al.*, 1000 Genomes Project Analysis Group, 2011 The variant call format and VCFtools. Bioinformatics **27**: 2156–2158.

DePristo M. A., Banks E., Poplin R., Garimella K. V., Maguire J. R. *et al.*, 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet **43**: 491–501.

Dirlewanger E., Graziano E., Joobeur T., Garriga-Calderé F., Cosson P. *et al.*, 2004 Comparative mapping and marker-assisted selection in Rosaceae fruit crops. Proceedings of the National Academy of Sciences **101**: 9891–9896.

Doyle J., Doyle J., 1987 A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull **19**: 11–15.

Dvorak J., Wang L., Zhu T., Jorgensen C. M., Deal K. R. *et al.*, 2018 Structural variation and rates of genome evolution in the grass family seen through comparison of sequences of genomes greatly differing in size. Plant J **95**: 487–503.

Ferguson-Smith M. A., Trifonov V., 2007 Mammalian karyotype evolution. Nat Rev Genet **8**: 950–962.

Ferriera J. V., 1980 Introgressive hybrisization between *Helanthus annuus* L. and *Helianthus petiolaris* Nutt. Mendeliana: 81–93.

Flagel L. E., Blackman B. K., Fishman L., Monnahan P. J., Sweigart A., Kelly J. K., 2019 GOOGA: A platform to synthesize mapping experiments and identify genomic structural diversity (FA Feltus, Ed.). PLoS Comput Biol **15**: e1006949–25.

Foulongne M., Pascal T., Ar s P., Kervella J., 2003 The potential of *Prunus davidiana* for introgression into peach [*Prunus persica* (L.) Batsch] assessed by comparative mapping. Theor Appl Genet **107**: 227–238.

Fuller Z. L., Haynes G. D., Richards S., Schaeffer S. W., 2017 Genomics of natural populations: Evolutionary forces that establish and maintain gene arrangements in *Drosophila pseudoobscura*. Mol Ecol **26**: 6539–6562.

Grant V., 1981 *Plant Speciation*. Columbia University Press.

645 Hackett C. A., Broadfoot L. B., 2003 Effects of genotyping errors, missing values and segregation
646 distortion in molecular marker data on the construction of linkage maps. Heredity **90**: 33–38.

647 Hardison R. C., Roskin K. M., Yang S., Diekhans M., Kent W. J. *et al.*, 2003 Covariation in frequencies of
648 substitution, deletion, transposition, and recombination during eutherian evolution. Genome
649 Research **13**: 13–26.

650 Heesacker A. F., Bachlava E., Brunick R. L., Burke J. M., Rieseberg L. H., Knapp S. J., 2009 Karyotypic
651 Evolution of the Common and Silverleaf Sunflower Genomes. The Plant Genome **2**: 233–14.

652 Heiser C., 1948 Taxonomic and Cytological Notes on the Annual Species of *Helianthus*. Bulletin of the
653 Torrey Botanical Club **75**: 512–515.

654 Heiser C., 1951 Hybridization in the annual sunflowers: *Helianthus annuus* x *H. argophyllus*. The
655 American Naturalist **85**: 65–72.

656 Heiser C. B. Jr, 1947 Hybridization between the sunflower species *Helianthus annuus* and *H. petiolaris*.
657 Evolution **1**: 249–262.

658 Heiser C. B. Jr, 1961 Morphological and cytological variation in *Helianthus petiolaris* with notes on
659 related species. Evolution **15**: 247–258.

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

662 Hooper D. M., Price T. D., 2017 Chromosomal inversion differences correlate with range overlap in
663 passerine birds. Nature Ecology & Evolution **1**: 1526–1534.

664 Huang S., Li R., Zhang Z., Li L., Gu X., et al., 2009 The genome of the cucumber, *Cucumis sativus* L. Nat
665 Genet **41**: 1275–1281.

666 Huchra J. P., Geller M. J., 1982 Groups of galaxies. I-Nearby groups. The Astrophysical Journal **257**:
667 423–437.

668 Jones F. C., Grabherr M. G., Chan Y. F., Russell P., Mauceli E., et al., 2012 The genomic basis of adaptive
669 evolution in threespine sticklebacks. Nature **484**: 55–61.

670 Kawakami T., Dhakal P., Katterhenry A. N., Heatherington C. A., Ungerer M. C., 2011 Transposable
671 element proliferation and genome expansion are rare in contemporary sunflower hybrid
672 populations despite widespread transcriptional activity of LTR retrotransposons. Genome Biol Evol
673 **3**: 156–167.

674 King M., 1987 Chromosomal rearrangements, speciation and the theoretical approach. Heredity **59**: 1–
675 6.

676 King M., 1993 *Species Evolution*. Cambridge University Press.

677 Kirkpatrick M., Barton N., 2006 Chromosome inversions, local adaptation and speciation. Genetics **173**:
678 419–434.

679 Lai Z., Nakazato T., Salmaso M., Burke J. M., Tang S. *et al.*, 2005 Extensive chromosomal repatterning
680 and the evolution of sterility barriers in hybrid sunflower species. Genetics **171**: 291–303.

681 Larkin D. M., Pape G., Donthu R., Auvil L., Welge M., Lewin H. A., 2009 Breakpoint regions and
682 homologous synteny blocks in chromosomes have different evolutionary histories. Genome
683 Research **19**: 770–777.

684 Lee C.-R., Wang B., Mojica J. P., Mandáková T., Prasad K. V. S. K. *et al.*, 2017 Young inversion with
685 multiple linked QTLs under selection in a hybrid zone. Nature Ecology & Evolution **1**: 119.

686 Lewis H., 1966 Speciation in Flowering Plants. Science **152**: 167–172.

687 Li H., Durbin R., 2010 Fast and accurate long-read alignment with Burrows-Wheeler transform.
688 Bioinformatics **26**: 589–595.

689 Lowry D. B., Willis J. H., 2010 A widespread chromosomal inversion polymorphism contributes to a
690 major life-history transition, local adaptation, and reproductive isolation. PLoS Biol **8**: 2227–2241.

691 Mandel J. R., Nambeesan S., Bowers J. E., Marek L. F., Ebert D. *et al.*, 2013 Association mapping and the
692 genomic consequences of selection in sunflower (NM Springer, Ed.). PLoS Genetics **9**: e1003378.

693 Mason C. M., 2018 How old are sunflowers? A molecular clock analysis of key divergences in the origin
694 and diversification of *Helianthus* (Asteraceae). Int. J Plant Sci. **179**: 182–191.

695 Matvienko M., Kozik A., Froenicke L., Lavelle D., Martineau B. *et al.*, 2013 Consequences of normalizing
696 transcriptomic and genomic libraries of plant genomes using a duplex-specific nuclease and
697 tetramethylammonium chloride (L Herrera-Estrella, Ed.). PLoS ONE **8**: e55913–17.

698 McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K. *et al.*, 2010 The Genome Analysis Toolkit:
699 A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research
700 **20**: 1297–1303.

701 Murphy W. J., Larkin D. M., Everts-van der Wind A., Bourque G., Tesler G. et al., 2005 Dynamics of
702 mammalian chromosome evolution inferred from multispecies comparative maps. Science **309**:
703 613–617.

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

706 Navarro A., Barton N. H., 2003 Chromosomal speciation and molecular divergence--accelerated
707 evolution in rearranged chromosomes. Science **300**: 321–324.

708 Neafsey D. E., Waterhouse R. M., Abai M. R., Aganezov S. S., Alekseyev M. A. *et al.*, 2015 Highly

evolvable malaria vectors: The genomes of 16 *Anopheles* mosquitoes. Science **347**: 1258522–11.

Noor M. A., Grams K. L., Bertucci L. A., Reiland J., 2001 Chromosomal inversions and the reproductive isolation of species. Proceedings of the National Academy of Sciences **98**: 12084–12088.

Novitski E., 1951 Non-random disjunction in Drosophila. Genetics **36**: 267–280.

Ostevik K. L., 2016 The ecology and genetics of adaptation and speciation in dune sunflowers.

Ouellette L. A., Reid R. W., Blanchard S. G., Brouwer C. R., 2017 LinkageMapView - Rendering High Resolution Linkage and QTL Maps. Bioinformatics **34**: 306-307.

Pevzner P., Tesler G., 2003 Genome rearrangements in mammalian evolution: lessons from human and mouse genomes. Genome Research **13**: 37–45.

Poland J. A., Brown P. J., Sorrells M. E., Jannink J.-L., 2012 Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach (T Yin, Ed.). PLoS ONE **7**: e32253.

Quillet M. C., Madjidian N., Griveau Y., Serieys H., Tersac M., Lorieux M., Berville A., 1995 Mapping genetic factors controlling pollen viability in an interspecific cross in *Helianthus* sect. *Helianthus*. Theor Appl Genet **91**: 1195–1202.

Raduski A. R., Rieseberg L., Strasburg J., 2010 Effective population size, gene flow, and species status in a narrow endemic sunflower, *Helianthus neglectus*, compared to its widespread sister species, *H. petiolaris*. IJMS **11**: 492–506.

Rieseberg L., 1991 Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes American Journal of Botany **78:** 1218-1237.

Rieseberg L. H., 2001 Chromosomal rearrangements and speciation. Trends in Ecology & Evolution **16**: 351–358.

Rieseberg L. H., Linder C. R., Seiler G. J., 1995 Chromosomal and genic barriers to introgression in *Helianthus*. Genetics **141**: 1163–1171.

Rieseberg L. H., Whitton J., Gardner K., 1999 Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics **152**: 713–727.

Rogers C. E., Thompson T. E., Seiler G. J., 1982 *Sunflowers species of the United States*. National Sunflower Association.

Rohland N., Reich D., 2012 Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Research **22**: 939–946.

Sambatti J. B. M., Strasburg J. L., Ortiz-Barrientos D., Baack E. J., Rieseberg L. H., 2012 Reconciling

740 extremely strong barriers with high levels of gene exchange in annual sunflowers. Evolution **66**:
741 1459–1473.

742 Schlautman B., Diaz-Garcia L., Covarrubias-Pazaran G., Schlautman N., Vorsa N. et al., 2017
743 Comparative genetic mapping reveals synteny and collinearity between the American cranberry
744 and diploid blueberry genomes. Molecular Breeding **38**: 969.

745 Searle J. B., 1993 Chromosomal hybrid zones in eutherian mammals. In: *Hybrid zones and the*
746 *evolutionary process*, pp. 309–353.

747 Shagina I., Bogdanova E., Mamedov I., Lebedev Y., Lukyanov S., Shagin D., 2010 Normalization of
748 genomic DNA using duplex-specific nuclease. Biotechniques **48**: 455–459.

749 Stathos A., Fishman L., 2014 Chromosomal rearrangements directly cause underdominant F1 pollen
750 sterility in *Mimulus lewisii-Mimulus cardinalis* hybrids. Evolution **68**: 3109–3119.

751 Staton S. E., Bakken B. H., Blackman B. K., Chapman M. A., Kane N. C. *et al.*, 2012 The sunflower
752 (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable
753 elements. The Plant Journal **72**: 142–153.

754 Stebbins G. L., Daly K., 1961 Changes in the Variation Pattern of a Hybrid Population of *Helianthus* Over
755 an Eight-Year Period. Evolution **15**: 60–71.

756 Stephens J. D., Rogers W. L., Mason C. M., Donovan L. A., Malmberg R. L., 2015 Species tree estimation
757 of diploid *Helianthus* (Asteraceae) using target enrichment. American Journal of Botany **102**: 910–
758 920.

759 Strasburg J., Rieseberg L., 2008 Molecular demographic history of the annual sunflowers *Helianthus*
760 *annuus* and *H. petiolaris*—Large effective population sizes and rates of long-term gene flow.
761 Evolution **62**: 1936–1950.

762 Tang S., Yu J. K., Slabaugh M. B., Shintani D. K., Knapp S. J., 2002 Simple sequence repeat map of the
763 sunflower genome. TAG Theoretical and Applied Genetics **105**: 1124–1136.

764 Taylor J., Butler D., 2017 RPackage ASMap: Efficient Genetic Linkage Map Construction and Diagnosis.
765 J. Stat. Soft. **79**: 1–29.

766 Templeton A. R., 1981 Mechanisms of Speciation - a Population Genetic Approach. Annual Review of
767 Ecology and Systematics **12**: 23–48.

768 Trickett A. J., Butlin R. K., 1994 Recombination suppressors and the evolution of new species. Heredity
769 **73**: 339–345.

770 Van der Auwera G. A., Carneiro M. O., Hartl C., Poplin R., del Angel G. et al., 2013 From fastQ data to
771 high-confidence variant calls: The genome analysis toolkit best practices pipeline. Current
772 Protocols in Bioinformatics **43**: 11.10.1–33.

773 Via S., 2012 Divergence hitchhiking and the spread of genomic isolation during ecological speciation-
774 with-gene-flow. Philosophical Transactions of the Royal Society B: Biological Sciences **367**: 451–
775 460.

776 Via S., West J., 2008 The genetic mosaic suggests a new role for hitchhiking in ecological speciation.
777 Mol Ecol **17**: 4334–4345.

778 Vitte C., Fustier M. A., Alix K., Tenaillon M. I., 2014 The bright side of transposons in crop evolution.
779 Briefings in Functional Genomics **13**: 276–295.

780 Vogel J. P., Garvin D. F., Mockler T. C., Schmutz J., Rokhsar D. et al., 2010 Genome sequencing and
781 analysis of the model grass *Brachypodium distachyon*. Nature **463**: 763–768.

782 Weiss-Schneeweiss H., Schneeweiss G. M., 2012 Karyotype Diversity and Evolutionary Trends in
783 Angiosperms. In: *Plant Genome Diversity Volume 2*,

784 Whelan E. D., 1979 Interspecific hybrids between *Helianthus petiolaris* Nutt. and *H. annuus* L.: Effect of
785 backcrossing on meiosis. Springer

786 White M. J. D., 1973 *Animal Cytology and Evolution*. Cambridge University Press, London.

787 White M. J. D., 1978 *Modes of Speciation*. W. H. Freeman & Co., San Francisco.

788 Wu F., Tanksley S. D., 2010 Chromosomal evolution in the plant family Solanaceae. BMC Genomics **11**:
789 182.

790 Yannic G., Basset P., Hausser J., 2009 Chromosomal rearrangements and gene flow over time in an
791 inter-specific hybrid zone of the *Sorex araneus* group. **102**: 616–625.

792 Yeaman S., 2013 Genomic rearrangements and the evolution of clusters of locally adaptive loci.
793 Proceedings of the National Academy of Sciences **110**: E1743–E1751.

794 Yeaman S., Whitlock M., 2011 The genetic architecture of adaptation under migration-selection
795 balance. Evolution **65**: 1897–1911.

796 Yogeeswaran K., Frary A., York T. L., Amenta A., Lesser A. H., Nasrallah J. B., Tanksley S. D., Nasrallah M.
797 E., 2005 Comparative genome analyses of *Arabidopsis* spp.: inferring chromosomal rearrangement
798 events in the evolutionary history of A. thaliana. Genome Research **15**: 505–515.

799 Zanders S. E., Eickbush M. T., Yu J. S., Kang J.-W., Fowler K. R., Smith G. R., Malik H. S., 2014 Genome
800 rearrangements and pervasive meiotic drive cause hybrid infertility in fission yeast. eLife **3**: 419–
801 23.

802 Zhulidov P. A., 2004 Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic
803 Acids Research **32**: 37e–37.