Genetics: Early Online, published on February 7, 2020 as 10.1534/genetics.120.303026

# Ancestral reconstruction of karyotypes reveals an exceptional rate of non-random chromosomal evolution in sunflower

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

Running Title: Chromosomal evolution in sunflower

Keywords: chromosomal rearrangement, synteny block, *Helianthus*, syntR, dot plot

Corresponding author:
Kate Ostevik
Box 90338, 137 Biological Sciences, 130 Science Drive, Durham, NC, 27708
984-227-0832
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 novel algorithm for identifying regions of synteny in pairs of genetic maps, which is implemented in the accompanying R package, syntR. The syntR algorithm performs as well as previous ad-hoc methods while being systematic, repeatable, and is applicable to mapping chromosomal rearrangements in any group of species. In addition, 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 syntR, and 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 7.9 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 exchange suggests that specific 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 may facilitate 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. Individuals heterozygous for divergent karyotypes are often sterile or inviable (King 1987, Lai *et al.*

2005, Stathos and Fishman 2014). Apart from directly causing hybrid sterility and inviability, chromosomal rearrangements can also facilitate the evolution of other reproductive barriers by extending genomic regions that are protected from introgression (Noor *et al.* 2001, Rieseberg 2001), accumulating genetic incompatibilities (Navarro and Barton 2003), and simplifying reinforcement (Trickett and Butlin 1994). Despite its prevalence and potentially important role in speciation, the general patterns of karyotypic divergence are still not well understood. Mapping and characterizing chromosomal rearrangements in many taxa is a critical step towards understanding their evolutionary dynamics.

The genus *Helianthus* (sunflowers) is well known to have particularly labile genome structure and is thus a viable system in which to map and characterize a variety of rearrangements. These sunflowers have several paleopolyploidy events in their evolutionary history (Barker *et al.* 2008, Barker *et al.* 2016, Badouin *et al.* 2017), have given rise to three homoploid hybrid species (Rieseberg 1991), and are prone to transposable element activity (Kawakami *et al.* 2011, Staton *et al.* 2012). Evidence in the form of hybrid pollen inviability, abnormal chromosome pairings during meiosis, and genetic map comparisons suggests that *Helianthus* karyotypes are unusually diverse (Heiser 1947, Heiser 1951, Heiser 1961, Whelan 1979, Chandler 1986, Rieseberg *et al.* 1995, Quillet *et al.* 1995, Burke *et al.* 2004, Heesacker *et al.* 2009, Barb *et al.* 2014). In fact, annual sunflowers have one of the highest described rates of chromosomal evolution across all plants and animals (Burke *et al.* 2004).

Studying chromosomal evolution within any group requires high-density genetic maps. Recently, Barb *et al.* (2014) built high-density genetic maps for the sunflower species *H. niveus* ssp. *tephrodes* and *H. argophyllus* and compared them to *H. annuus*. This analysis precisely mapped previously inferred karyotypes (Heiser 1951, Chandler 1986, Quillet *et al.* 1995), but only captured a small amount of the chromosomal variation in the annual sunflowers. For example, comparisons of genetic maps with limited marker density suggest that several chromosomal rearrangements differentiate *H. petiolaris* from *H. annuus* (Rieseberg *et al.* 1995, Burke *et al.* 2004) and evidence from cytological surveys suggests that subspecies within *H. petiolaris* subspecies carry divergent karyotypes (Heiser 1961). Adding high-density genetic maps of *H. petiolaris* subspecies to the Barb *et al.* (2014) analysis will allow

us to: (1) precisely track additional rearrangements, (2) reconstruct ancestral karyotypes for the group, and (3) untangle overlapping rearrangements that can be obscured by directly comparing present-day karyotypes.

Another critical part of a multi-species comparative study of chromosome evolution using genetic map data is a systematic and repeatable method for identifying syntenic chromosomal regions (*sensu* Pevzner and Tesler 2003). These methods are especially important for cases with high marker density because breakpoints between synteny blocks can be blurred by mapping error, micro-rearrangements, and paralogy (Hackett and Broadfoot 2003, Choi *et al.* 2007, Barb *et al.* 2014, Bilton *et al.* 2018). In previous studies, synteny blocks have been found by a variety of ad-hoc methods, including counting all differences in marker order (Wu and Tanksley 2010), by visual inspection (Burke *et al.* 2004, Marone *et al.* 2012, Latta *et al.* 2019), or by manually applying simple rules like size thresholds (Heesacker *et al.* 2009, Barb *et al.* 2014, Rueppell *et al.* 2016) and Spearman's rank comparisons (Berdan *et al.* 2014, Schlautman *et al.* 2017). However, these methods become intractable and prone to error when applied 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).

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 and repeatably identify synteny blocks from any number of paired genetic map positions, (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 (2*n* = 34) taxa from the annual clade of the genus *Helianthus* (Fig 1). These sunflowers are native to North America (Fig S1, Rogers *et al.* 1982) and are naturally self-incompatible (domesticated lineages of *H. annuus* are self-compatible). *Helianthus annuus* occurs throughout much of the central United States, often in somewhat heavy soils and along roadsides (Heiser 1947). *Helianthus 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* and *H. annuus* are sympatric, 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 dunes from the southwestern US into Mexico.

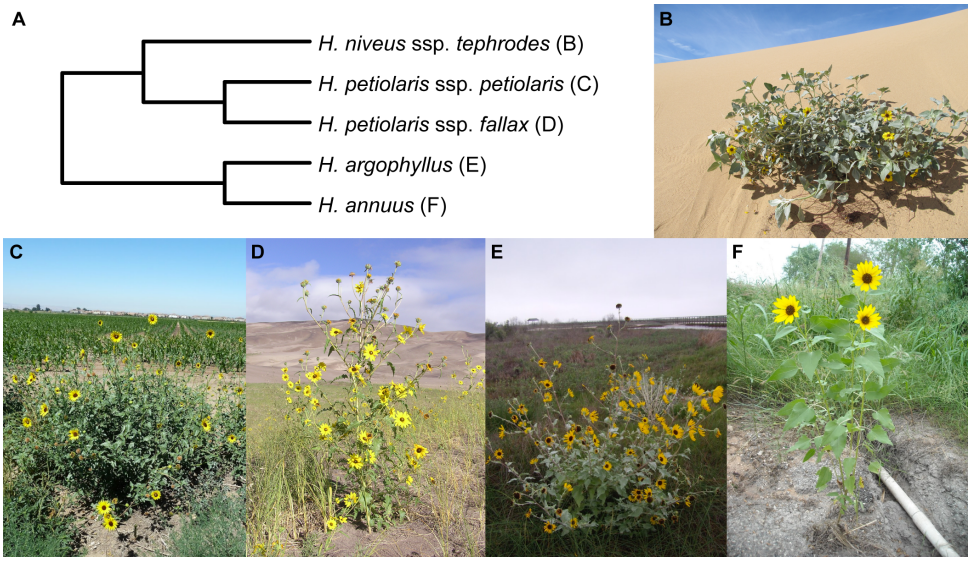
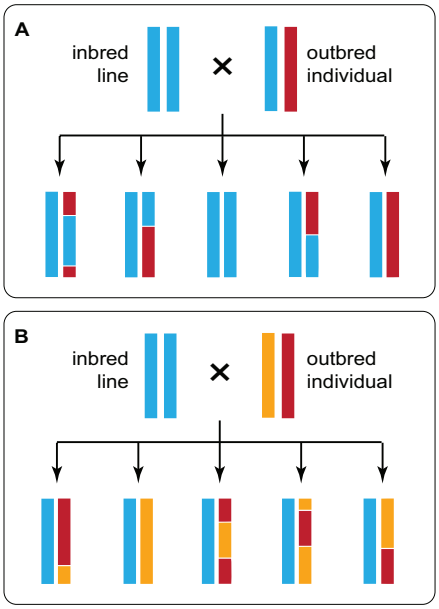



Figure 1 - The sunflower taxa used in this study. A) Phylogenetic relationships based on Stephens *et al.* (2015) and Baute *et al.* (2016). B) *H. niveus* ssp. *tephrodes.* C) *H. petiolaris* ssp. *petiolaris.* D) *H. petiolaris* ssp. *fallax.* E) *H. argophyllus.* F) *H. annuus.* Photo credits: Brook Moyers (B, C, E & F) and Rose Andrew (D).

## Controlled crosses

To make genetic maps, we crossed an outbred individual with presumably high heterozygosity from each *H. petiolaris* subspecies to a homozygous inbred line of domesticated sunflower and genotyped

102 the resulting F1 offspring. This test-cross design allows us to infer where recombination occurred in the
103 heterozygous parents because we can reliably track the segregation of those parents' alleles against a
104 predictable background (Fig 2).




105

106 Figure 2 – Diagram showing how a test-cross can be used to map the recombination events in an outbred
107 individual that may (A) or may not (B) share alleles with the inbred line. Each line represents a chromosome, and
108 the colors represent ancestry.

109

110 Specifically, we used pollen from a single *H. petiolaris ssp. petiolaris* plant (PI435836) and a single *H.*
111 *petiolaris ssp. fallax* plant (PI435768) to fertilize individuals of a highly inbred and male sterile line of *H.*
112 *annuus* (HA89cms). The self-incompatible *H. petiolaris* accessions were collected in central Colorado
113 (PI435836, 39.741°, -105.342°, Boulder County) and the southeast corner of New Mexico (PI435768,
114 32.3°, -104.0°, Eddy County, Fig S1) and were maintained at large population sizes by the United States
115 Department of Agriculture. When it was originally collected, accession PI435768 was classified *H.*
116 *neglectus*. However, based on the location of the collection (Heiser 1961) and a more recent genetic
117 analysis of the scale of differences between *H. petiolaris* ssp. *fallax* and *H. neglectus* (Raduski *et al.*
118 2010), we believe that this accession should be classified *H. petiolaris ssp. fallax*.

119 ## Genotyping

120

121 We collected leaf tissue from 116 *H. annuus* x *H. petiolaris* ssp. *petiolaris* F1 seedlings and 132 *H.*
122 *annuus* x *H. petiolaris* ssp. *fallax* F1 seedlings. We extracted DNA using a modified CTAB protocol
123 (Doyle and Doyle 1987) and prepared individually barcoded genotyping-by-sequencing (GBS) libraries
124 using a version of the Poland *et al.* (2012) protocol. Our modified protocol includes steps to reduce the
125 frequency of high-copy fragments (e.g., chloroplast and repetitive sequence) based on Shagina *et al.*
126 (2010) and Matvienko *et al.* (2013) and steps to select specific fragment sizes for sequencing (see
127 Ostevik 2016 appendix B for the full protocol).

128

129 Briefly, we digested 100ng of DNA from each individual with restriction enzymes (either *Pst*I-HF or *Pst*I-
130 HF and *Msp*I) and ligated individual barcodes and common adapters to the digested DNA. We pooled
131 barcoded fragments from up to 192 individuals, cleaned and concentrated the libraries using SeraMag
132 Speed Beads made in-house (Rohland and Reich 2012), and amplified fragments using 12 cycles of PCR.
133 We depleted high-copy fragments based on Todesco *et al.* (2019) using the following steps: (1)
134 denature the libraries using high temperatures, (2) allow the fragments to re-hybridize, (3) digest the
135 double-stranded fragments with duplex specific nuclease (Zhulidov *et al.* 2004), and (4) amplify the
136 undigested fragments using another 12 cycles of PCR. We ran the libraries out on a 1.5% agarose gel
137 and extracted 300-800 bp fragments using a Zymoclean Gel DNA Recovery kit (Zymo Research, Irvine,
138 USA). Then, following additional library cleanup and quality assessment, we sequenced paired-ends of
139 our libraries on an Illumina HiSeq 2000 (Illumina Inc., San Diego, CA, USA).

140

141 To call variants, we used a pipeline that combines the Burrows-Wheeler Aligner version 0.7.15 (BWA, Li
142 & Durbin 2010) and the Genome Analysis Toolkit version 3.7 (GATK, McKenna *et al.* 2010). First, we
143 demultiplexed the data using sabre (https://github.com/najoshi/sabre, Accessed 27 Jan 2017). Next,
144 we aligned reads to the *H. annuus* reference (HanXRQr1.0-20151230, Badouin *et al.* 2017) with 'bwa-
145 mem' (Li 2013), called variants with GATK 'HaplotypeCaller', and jointly genotyped all samples within a
146 cross type with GATK 'GentypeGVCFs'. We split variants into SNPs and indels and filtered each marker
147 type using hard-filtration criteria suggested in the GATK best practices (DePristo *et al.* 2011, Van der

148 Auwera *et al.* 2013). Specifically, we removed SNPs that had quality by depth scores (QD) less than 2,
149 strand bias scores (FS) greater than 60, mean mapping quality (MQ) less than 40, or allele mapping bias
150 scores (MQRankSum) less than -12.5 and indels that had QD < 2 or FS > 200. After further filtering
151 variants for biallelic and triallelic markers with genotype calls in at least 50% of individuals, we used
152 GATK 'VariantsToTable' to merge SNPs and indels into a single variant table for each cross type.
153
154 Finally, we converted our variant tables into AB format, such that the heterozygous parents contribute
155 'A' and 'B' alleles to offspring, while the *H. annuus* parent contributes exclusively 'A' alleles. At biallelic
156 markers (Fig 2A), sites with two reference alleles became 'AA' and sites with the reference allele, and
157 the alternate allele became 'AB'. At triallelic markers (Fig 2B), sites with the reference allele and one
158 alternate allele became 'AA' and sites with the reference allele, and the other alternate allele became
159 'AB'. This method randomly assigns 'A' and 'B' alleles to the homologous chromosomes in each
160 heterozygous parent, so our genetic maps initially consisted of pairs of mirror-imaged linkage groups
161 that we later merged.

## 162 Genetic mapping

163
164 We used R/qtl (Broman *et al.* 2003) in conjunction with R/ASMap (Taylor and Butler 2017) to build
165 genetic maps. After excluding markers with less than 20% or greater than 80% heterozygosity and
166 individuals with less than 50% of markers scored, we used the function 'mstmap.cross' with a stringent
167 significance threshold (p.value = $1x10^{-16}$) to form conservative linkage groups. We used the function
168 'plotRF' to identify pairs of linkage groups with unusually high recombination fractions and the function
169 'switchAlleles' to reverse the genotype scores of one linkage group in each mirrored pair. We did this
170 until reversing genotype scores no longer reduced the number of linkage groups.
171
172 Using the corrected genotypes, we made new linkage groups with only the most reliable markers.
173 Namely, we used the function 'mstmap.cross' (with the parameter values: dist.fun = "kosambi", p.value
174 = $1x10^{-6}$, noMap.size = 2, noMap.dist = 5) on markers with less than 10% missing data and without
175 significant segregation distortion. We refined the resulting linkage groups by removing (1) markers

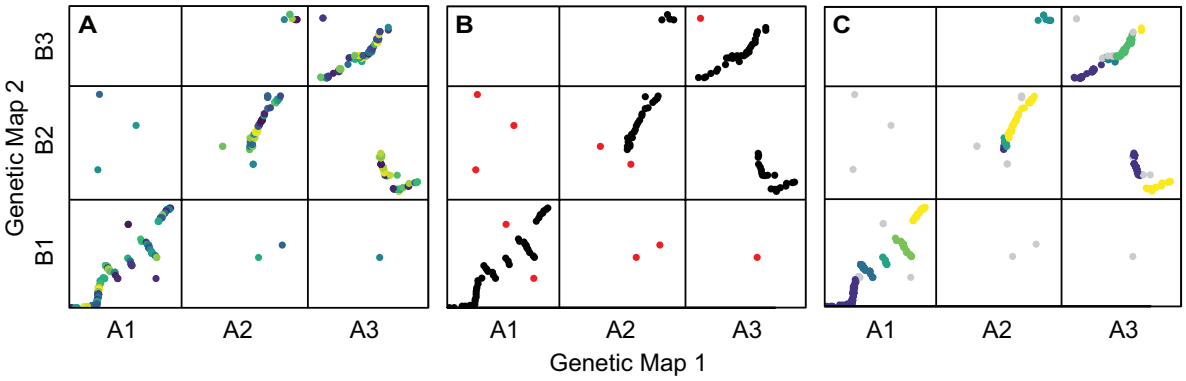
176 with more than three double crossovers, (2) markers with aberrant segregation patterns (segregation
177 distortion more than two standard deviations above or below the mean segregation distortion of the
178 nearest 20 markers), and (3) linkage groups made up of fewer than four markers.
179
180 We progressively pushed markers with increasing amounts of segregation distortion and missing data
181 into the maps using the function 'pushCross'. After adding each batch of markers, we reordered the
182 linkage groups and dropped markers and linkage groups as described above. Once all the markers had
183 been pushed back, we used the function 'calc.errorlod' to identify possible genotyping errors (error
184 scores greater than 2) and replaced those genotypes with missing data. We continued to drop linkage
185 groups, markers, and genotypes that did not meet our criteria until none remained.
186
187 Finally, we dropped five excess linkage groups, each made up of fewer than 30 markers, from each
188 map. The markers in these linkage groups mapped to regions of the *H. annuus* genome that were
189 otherwise represented in the final genetic maps but could not be explained by reversed genotypes.
190 Instead, these markers were likely polymorphic in the HA89cms individual used for crosses because of
191 the 2-4% residual heterozygosity in sunflower inbred lines (Mandel *et al.* 2013).

## 192 SyntR development

193
194 To aid in the identification of chromosomal rearrangements, we developed the R package 'syntR'
195 (code and documentation available at http://ksamuk.github.io/syntR). This package implements a
196 heuristic algorithm for systematically detecting synteny blocks from marker positions in two genetic
197 maps. The key innovation of the syntR algorithm is coupling a biologically-informed noise reduction
198 method with a cluster identification method better suited for detecting linear (as opposed to circular)
199 clusters of data points.
200
201 We based the syntR algorithm on the following statistical and biological properties of genetic maps and
202 chromosomal rearrangements:

203 (1) Synteny blocks appear as contiguous sets of orthologous markers in the same or reversed order
204 in pairs of genetic maps (Pevzner and Tesler 2003, Choi *et al.* 2007).
205 (2) The inferred order of markers in individual genetic maps is subject to error due to genotyping
206 errors and missing data (Hackett and Broadfoot 2003). This error manifests as slight differences
207 in the order of nearby markers within a linkage group between maps. This mapping error
208 (which we denote 'error rate one') results in uncertainty in the sequence of markers in synteny
209 blocks.
210 (3) In genomes with a history of duplication, seemingly orthologous markers can truly represent
211 paralogs. These errors ('error rate two') look like tiny translocations and also disrupt marker
212 orders within synteny blocks.
213 (4) When comparing genetic maps derived from genomes without duplications or deletions, every
214 region of each genome will be uniquely represented in the other. Because syntR is made for
215 comparing homoploid genomes with this property, we expect each point in each genetic map
216 to be contained within a single unique synteny block. Therefore, overlaps between synteny
217 blocks are likely errors. Note that this assumption precludes the identification of duplications.
218 (5) Chromosomal rearrangements can be of any size, but smaller rearrangements are difficult to
219 distinguish from error (Pevzner and Tesler 2003). A key decision in synteny block detection is
220 thus the choice of a detection threshold for small rearrangements, which results in a trade-off
221 between error reduction and the minimum size of detectable synteny blocks.
222
223 The first step of the syntR algorithm is to smooth over mapping error (error rate one) by identifying
224 highly localized clusters of markers based on a genetic distance threshold (cM) in both maps using
225 hierarchical clustering (Fig 3a). The number of clusters formed is determined by the parameter
226 maximum cluster range ($CR_{max}$) that defines the maximum genetic distance (cM) that any cluster can
227 span in either genetic map. After determining these initial clusters, we smooth the maps by collapsing
228 each multi-marker cluster down into a single representative point (the centroid of the cluster) for
229 processing in subsequent steps. Next, we address errors introduced by poorly mapped or paralogous
230 markers (error rate two) by flagging and removing outlier clusters that do not have a neighboring

231 cluster within a specified maximum genetic distance (cM), a parameter we denote nearest neighbor
232 distance ($NN_{dist}$, Fig 3b).
233




234
235 Figure 3 – The stages of the syntR algorithm. Each plot shows the relationship between markers or clusters of
236 markers from three chromosomes in two genetic maps. A) Highly localized markers are clustered. Each shade
237 represents an individual cluster of markers that will be collapsed into a single representative point. B) Clusters
238 without another cluster nearby are dropped. Red points represent clusters without a neighbor within 10 cM. C)
239 Clusters are grouped into synteny blocks based on their rank positions. Grey points represent markers that were
240 dropped in previous steps, and each other color represents a different synteny block.
241
242 After the noise reduction steps, we define preliminary synteny blocks using a method similar to the
243 "friends-of-friends" clustering algorithm (Huchra and Geller 1982). First, we transform the genetic
244 position of each cluster into rank order to minimize the impact of gaps between markers. We then
245 group clusters that are (1) adjacent in rank position in one of the maps and (2) within two rank
246 positions in the other map (Fig S2). This grouping method further reduces the effect of mapping error
247 by aggregating over pairs (but not triplets) of clusters that have reversed orientations. If a minimum
248 number of clusters per synteny block has been (optionally) defined, we sequentially eliminate blocks
249 that fall below the minimum number of clusters, starting with blocks made up of one cluster and
250 ending with blocks made up of clusters equal to one less than the minimum. After each elimination, we
251 regroup the clusters into new synteny blocks. Finally, we adjust the extents of each synteny block by
252 removing overlapping sections from both synteny blocks so that every position in each genetic map is
253 uniquely represented (Fig 3c).

## Assessing the performance of the syntR algorithm

To evaluate the performance of this method and explore the effect of parameter choice on outcomes, we simulated genetic map comparisons with known inversion breakpoints and error rates in R. The genetic map comparisons were made by randomly placing 200 of markers at 100 positions along a 100 cM chromosome in two maps, reversing marker positions within a defined inversion region in one map, and then repositioning markers based on simulated mapping noise using the following two error parameters: (1) $ER_1$ is the standard deviation of a normal distribution used to pick the distances markers are pushed out of their correct positions (e.g., when $ER_1$ is 1 cM 95% of markers will be within 2 cM of their true position); (2) $ER_2$ is the proportion of markers that are repositioned according to a uniform distribution (i.e., these markers can be moved to any position on the simulated chromosome).

We initially ran syntR using fixed syntR parameters ($CR_{max}$ = 2 and $NN_{dist}$ = 10) on multiple simulated maps, which were made using variable parameters (inversion size: 2.5-50 cM, $ER_1$: 0-2.0 cM, and $ER_2$: 0-20%), and counted the number of times the known breakpoints were identified within 1 cM (Fig S3). As expected, we find that rearrangement size affects the false negative rate (i.e., failing to detect known breakpoints), such that smaller inversions are more likely to be missed (Fig S3c), but does not affect the false positive rate (i.e., detecting breakpoints where there are none). We also find that increasing both types of error in the genetic maps tends to increase both the false positive and false negative rates, although $ER_1$ has a much stronger effect on the false positive rate than any other combination (Fig S3a,b).

Using the same simulation methods as above but now varying the syntR parameter $CR_{max}$, we find that small values of $CR_{max}$ yield high false positive rates while large values yield high false negative rates (Fig S4a). In addition, the $ER_1$ parameter has a strong effect on the relationship between $CR_{max}$ and the false positive rate. Higher values of $CR_{max}$ are needed to reduce the false positive rate when $ER_1$ is also high (Fig S4b). This means that picking an appropriate $CR_{max}$ value is key to the accuracy of this method. Although $NN_{dist}$ has a much weaker effect on outcomes than $CR_{max}$, it is useful to consider both parameter values carefully.

When the syntR heuristic algorithm is performing well, the final synteny blocks should represent all positions in the two genetic maps being compared (Chen *et al.* 2009). Based on this characteristic, we developed a method to choose optimal syntR tuning parameters ($CR_{max}$ and $NN_{dist}$) that maximize the representation of the genetic maps and markers in synteny blocks. In this method a user: (1) runs syntR with a range of parameter combinations; (2) saves summary statistics about the genetic distance of each map represented in the synteny blocks and the number of markers retained for each run; and (3) finds the parameter combination that maximizes a composite statistic that equally weights these three measures. In cases where there are multiple local maxima, we suggest choosing the local maximum with the smallest value of $CR_{max}$ to reduce the number of potential false positives.

The "maximize representation" method for choosing syntR parameters has several benefits. First, it does not rely on any additional information (e.g., error rate estimates from the genetic maps compared). Second, when we use this method to choose the best parameters for simulated genetic maps, we find that these parameter values also minimize false positive and false negative rates (Fig S5). Third, when we simulate biologically realistic genetic map comparisons, the absolute value of false positives and false negatives are small. For example, when comparing two genetic maps in which ~95% of markers are within 1 cM of their true position ($ER_1 = 0.5$) and 5% of markers are randomly permuted ($ER_2 = 0.05$), nonexistent breakpoints will be identified 0.1 times and a breakpoint of a 20 cM inversion will be missed 0.04 times. These low error rates also highlight the overall robustness and accuracy of the syntR algorithm.

In addition to performing simulations, we compared the synteny blocks identified by syntR to those identified by other means in a previously published comparison of *H. niveus* ssp. *tephrodes* and *H. argophyllus* maps to *H. annuus* (Barb *et al.* 2014). To do this, we formatted the original datasets for input into syntR and used the "maximize representation" method to determine the optimal parameter values for the two comparisons (*H. niveus vs. H. annuus*: $CR_{max} = 1.5$, $NN_{dist} = 30$; *H. argophyllus vs. H. annuus*: $CR_{max} = 2$, $NN_{dist} = 20$). We found that syntR was in strong agreement with previous work (Fig S6), recovering all the same translocations and most of the same inversions as the Barb *et al.* (2014)

312 maps. Most of the cases of mismatches were very small or weakly supported inversions in the Barb *et*
313 *al.* (2014) maps that syntR did not identify.
314

315 ## Finding synteny blocks

316

317 We used syntR to identify synteny blocks between our newly generated genetic maps and an ultra-
318 high-density map of *H. annuus* that was used to build the sunflower genome that we use as a reference
319 (Badouin *et al.* 2017). This allowed us to easily convert between physical position in the *H. annuus*
320 reference and position in the *H. annuus* genetic map. Using this property, we further compared two
321 previously published genetic maps for the closely related sunflower species, *H. niveus* ssp. *tephrodes*
322 and *H. argophyllus* (Barb *et al.* 2014), to the same *H. annuus* map. We aligned marker sequences from
323 the published maps to the *H. annuus* reference using bwa and converted well-aligned markers (MQ >
324 40) to their positions in the *H. annuus* genetic map.
325

326 Initially, we ran syntR using parameters identified through the "maximize representation" method for
327 each map comparison separately (Table S1). However, varying $CR_{max}$ revealed rearrangements that
328 were shared between the maps (Fig S7). Therefore, we ran syntR again using a range of $CR_{max}$ values
329 that included the best fit for each comparison (1.0 - 3.5 in 0.5 increments) and extracted a curated set
330 of synteny blocks from the output. A synteny block was retained if it fulfilled any of the following
331 criteria (in decreasing order of importance): (1) it was found in another species, (2) it was identified in
332 the majority of syntR runs for a single species, (3) it maximized the genetic distance represented by
333 synteny blocks. We present this curated set of synteny blocks below, but our results are unchanged if
334 we use the individually-fit synteny blocks.
335

336 We named the chromosomes in our genetic maps based on their synteny with the standard order and
337 orientation of *H. annuus* chromosomes (Tang *et al.* 2002, Bowers *et al.* 2012) following Barb *et al.*
338 (2014) but with shortened prefixes (A = *H. annuus*, R = *H. argophyllus*, N = *H. niveus* ssp. *tephrodes*, P =

339 *H. petiolaris* ssp. *petiolaris*, F = *H. petiolaris* ssp. *fallax*). For example, an *H. petiolaris* ssp. *fallax*
340 chromosome made up of regions that are syntenic with *H. annuus* chromosomes 4 and 7 is called F4-7.
341

## 342 Karyotype reconstruction and analysis

343

344 We used our inferred synteny blocks and the software MGR v 2.01 (Bourque and Pevzner 2002) to infer
345 ancestral karyotypes for our five *Helianthus* taxa and to determine the number of chromosomal
346 rearrangements that occurred along each branch of the species tree. To run the MGR analysis, we
347 needed the order and orientations of synteny blocks in all five maps. However, individual synteny
348 blocks were often missing from one or more of our final maps. We approached this problem in two
349 ways. First, we inferred the likely position of missing synteny blocks based on the location of markers
350 that were too sparse to be grouped by syntR and matched the location of synteny blocks in other
351 maps. In the second case, we dropped any synteny blocks that were not universally represented.
352 Because we already had two sets of synteny blocks for each map (curated and individually optimized),
353 we ran the MGR analyses using three different sets of synteny blocks: (set 1) curated and inferred, (set
354 2) curated and present in all five maps, (set 3) individually optimized and present in all five maps.
355

## 356 Data availability

357

358 The R program, syntR, is available on GitHub: https://github.com/ksamuk/syntR. The sequences used
359 to generate genetic maps are available on the SRA: http://www.ncbi.nlm.nih.gov/bioproject/598366.
360 All other data and scripts are available on dryad: https://doi.org/10.5061/dryad.7sqv9s4pc.

# 361 Results

## 362 Genetic maps

363

364 Both *H. petiolaris* genetic maps are made up of the expected 17 chromosomes and have very high
365 marker density (Fig 4, Fig S8). Only 6% of the *H. petiolaris* ssp. *petiolaris* map and 10% of the *H.*
366 *petiolaris* ssp. *fallax* map fails to have a marker within 2 cM (Fig S9). Overall, both maps are somewhat
367 longer than the *H. petiolaris* map reported by Burke *et al.* (2004). Although this could represent real
368 variation between genotypes, it could also be the result of spurious crossovers that are inferred based
369 on genotyping errors. Because genotyping errors are proportional to the number of markers, maps
370 with high marker densities are more likely to be inflated. Indeed, building maps with variants that were
371 thinned to 1 per 150 bp using vcftools version 0.1.13 (Danecek *et al.* 2011) yields collinear maps that
372 are closer to the expected lengths (Table S2, Fig S10). We present subsequent results based on the full
373 maps to improve our resolution for detecting small rearrangements.
374
375 Despite the general expansion of our maps, we find that chromosomes 2 and 4 in the *H. petiolaris* ssp.
376 *fallax* map (F2 and F4) are unexpectedly short (Fig 4). When we look at the distribution of markers for
377 this map relative to the *H. annuus* reference, we find very few variable sites in the distal half of these
378 chromosomes (Fig S11). That is, this individual was homozygous along vast stretches of F2 and F4.
379 These runs of homozygosity could be explained by recent common ancestry (i.e., inbreeding) or a lack
380 of variation in the population (e.g, because of background selection or a recent selective sweep).
381 Regardless, the lack of variable sites within the *H. petiolaris* ssp. *fallax* individual used for crosses
382 explains the shortness of F2 and F4. Notably, we find the same pattern on the distal half of *H. annuus*
383 chromosome 7 and find that this region is also not represented in the *H. petiolaris* spp. *fallax* map.
384

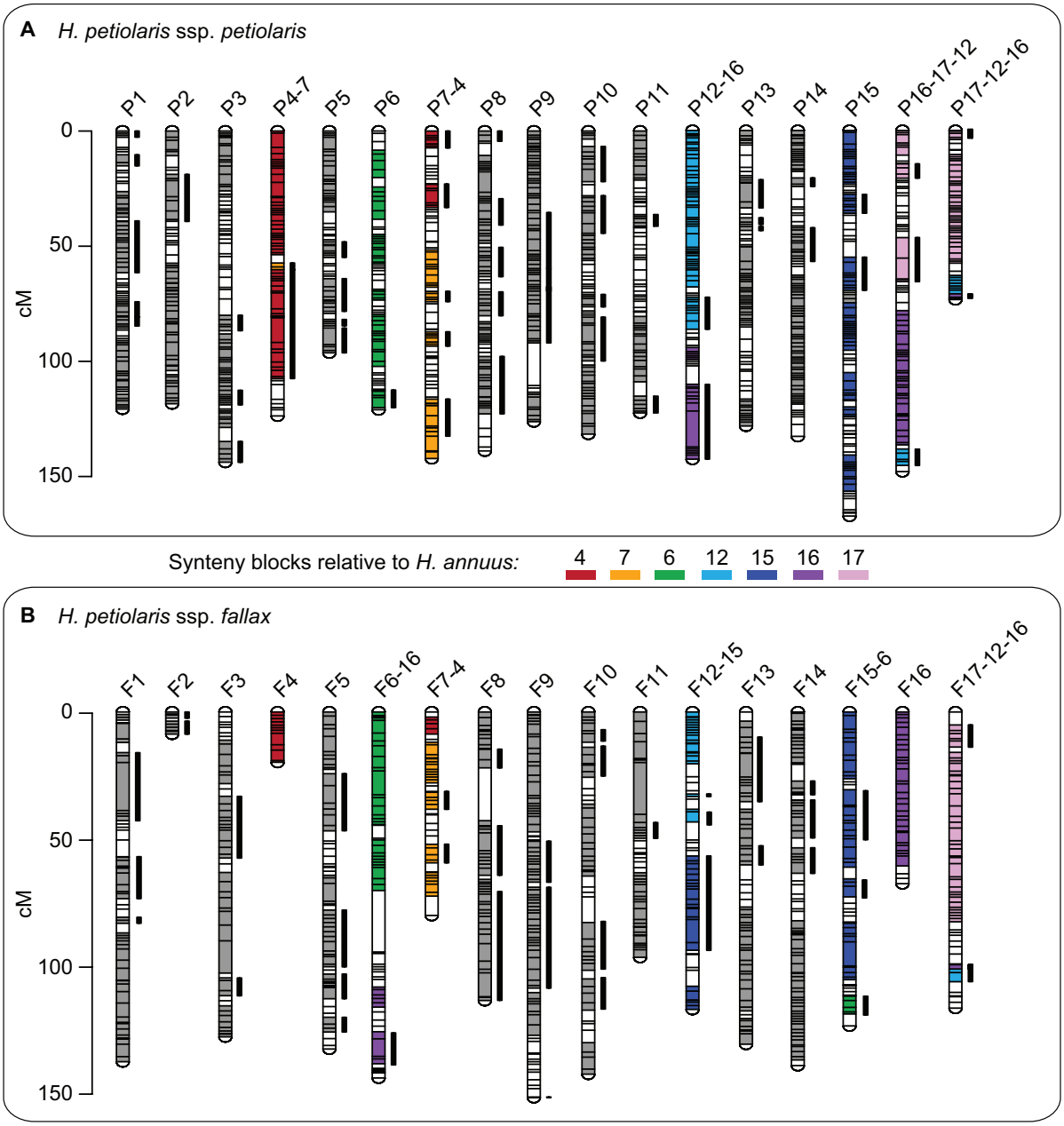



Figure 4 – *Helianthus petiolaris* genetic maps showing blocks of synteny with *H. annuus*. Each horizontal bar represents a genetic marker. The thick vertical bars next to chromosomes represent synteny blocks that are inverted relative to the *H. annuus* genetic map. Where there are no translocations between *H. petiolaris* and *H. 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. Regions that are not assigned to a synteny block remain white. The synteny blocks plotted are those curated based on multiple runs of syntR using different parameters. Please see Fig S12 for a labeled version. This figure was made with LinkageMapView (Ouellette *et al.* 2017).

394 ## Synteny blocks

395

396 Using syntR, we recovered 97 genetic regions that are syntenic between the *H. petiolaris* ssp. *petiolaris*
397 and *H. annuus* and 79 genetic regions that are syntenic between the *H. petiolaris* ssp. *fallax* and *H.*
398 *annuus* (Fig 4). We also recovered synteny blocks for the *H. niveus* ssp. *tephrodes* and *H. argophyllus*
399 comparisons that are similar to those found previously (Fig S13). In all four comparisons, syntR
400 successfully identified synteny blocks that cover large proportions (63%-90%) of each genetic map even
401 in the face of a very high proportion of markers that map to a different chromosome than their
402 neighbors (Table 1). These "rogue markers" could be the result of very small translocations, poorly
403 mapped markers, or extensive paralogy. Over and above the prevalence of rogue markers, the
404 karyotypes we recovered are substantially rearranged. Only between 32% and 45% of synteny blocks
405 for each map are collinear with the *H. annuus* genetic map in direct comparisons (Table 1).

406

407 Table 1 – Properties of the synteny blocks found using a syntR analysis between genetic maps of *H. annuus* and
408 four other *Helianthus* taxa. The proportion of rogue markers is based only on the chromosomes without
409 translocations in any map (i.e., chromosomes 1-3, 5, 8-10, 11, and 14). For those chromosomes, the majority of
410 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% |

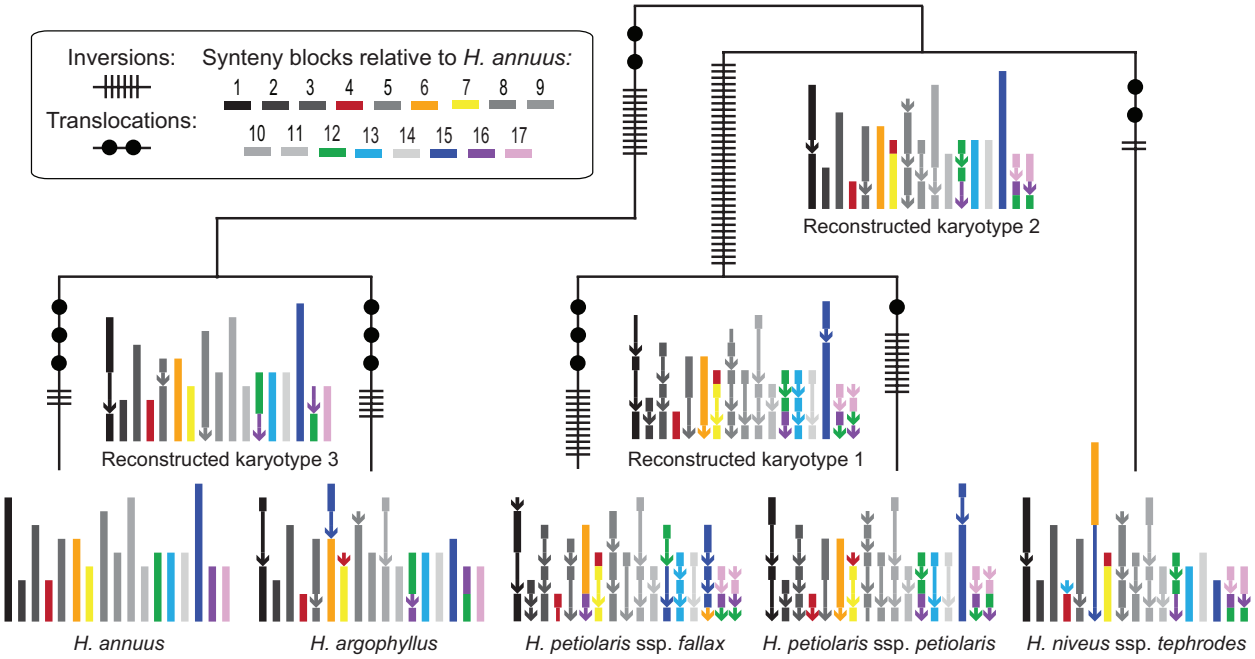
411

412 ## Karyotype reconstruction and chromosomal rearrangement

413

414 Because nested and shared rearrangements can obscure patterns of chromosome evolution, we use
415 the MGR analyses to predict the most likely sequence of rearrangements in a phylogenetic context
416 before quantifying the rearrangement rate. These MGR analyses identified similar patterns of
417 chromosome evolution regardless of the exact set of synteny blocks that we used (Table S5). Multiple

418 taxa share many rearrangements, and the similarity of karyotypes matches known phylogenetic
419 relationships. Moreover, MGR analyses run without a guide tree inferred the known species tree, and
420 MGR analyses run with all other topologies identified an inflated number of chromosomal
421 rearrangements.
422




423
424 Figure 5 – Diagram showing the karyotypes of 5 *Helianthus* taxa as well as reconstructed ancestral karyotypes
425 and the locations of chromosomal rearrangements. The karyotypes were built using synteny block set 1, which
426 were curated based on multiple syntR runs and inferred when missing. Each synteny block is represented using a
427 line segment that is color-coded based on its position in the *H. annuus* genome (see Fig S14 for a labeled
428 version). Chromosomes without translocations in any map are plotted in grey, and synteny blocks that are
429 inverted relative to *H. annuus* are plotted using arrows. Also, note that along some branches the same pair of
430 chromosomes is involved in multiple translocations.
431
432 Using the most complete set of synteny blocks (set 1), we find that 88 chromosomal rearrangements
433 occurred across the phylogeny (Fig 5). Then, using the most current divergence time estimates for this
434 group (Todesco *et al.* 2019) and conservatively assuming that *H. niveus* ssp. *tephrodes* diverged at the
435 earliest possible point, we estimate that 7.9 (7.8-8) rearrangements occurred per million years in this
436 clade (Tables S3-S5). To further explore the potential range of rearrangement rates, we considered

437 other estimates of divergence times in sunflower (Sambatti *et al.* 2012, Mason 2018) and the other
438 sets of synteny blocks. Overall, the lowest rate we identified was 2.6 rearrangements per million years,
439 while the highest rate was indeterminable because some minimum divergence time estimates for the
440 group include 0 (Tables S3-S5).
441
442 The 88 rearrangements include 74 inversions and 14 translocations that are quite evenly distributed
443 across the phylogeny. However, the excess inversions indicate that it is unlikely that the rate of
444 inversions is equal to the rate of translocation (binomial test, $5.1 \times 10^{-11}$). Furthermore, we find that only
445 8 of the 17 chromosomes are involved in the 14 translocations we identified. If translocations were
446 equally likely for all chromosomes, this asymmetry is very unlikely to have happened by chance (the
447 probability of sampling $\leq 8$ chromosomes in 14 translocations is $8.0 \times 10^{-8}$, Fig S15), suggesting that
448 some chromosomes are more likely to be involved in translocations than other. In line with this
449 observation, we see that some chromosome segments are repeatedly translocated. For example, A4
450 and A7 are involved in several exchanges, and part of A6 has a different position in almost every map
451 (Fig 5).

452 # Discussion

453 Large-scale chromosomal changes may be key contributors to the process of adaptation and
454 speciation, yet we still have a poor understanding of rates of chromosomal rearrangement and the
455 evolutionary forces underlying those rates. Here, we devised a novel, systematic method for
456 comparing any pair of genetic maps, and performed a comprehensive analysis of the evolution of
457 chromosomal rearrangements in a clade of sunflowers. We created two new genetic maps for
458 *Helianthus* species and used our new method to identify a wide range of karyotypic variation in our
459 new maps, as well as previously published maps. Consistent with previous studies, we discovered a
460 high rate of chromosomal evolution in the annual sunflowers. Further, we found that inversions are
461 more common than translocations and that certain chromosomes are more likely to be translocated.
462 Below, we discuss the evolutionary and methodological implications of this work and suggest some
463 next steps in understanding the dynamic process of chromosomal rearrangement.

464 ## Identifying rearrangements

465

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

472

473 Overall, we believe syntR will be a valuable tool for the systematic study of chromosomal
474 rearrangements in any species. The only data syntR needs to identify synteny blocks is relative marker
475 positions in two genetic maps. This fact is significant because, although the number of species with
476 whole genome sequence and methods to detect synteny blocks from those sequences are rapidly
477 accumulating, such as Mauve (Darling *et al.* 2004), Cinteny (Sinha and Meller 2007), syMAP (Soderlund
478 *et al.* 2011), SynChro (Drillion *et al.* 2014) and SyRI (Goel *et al.* 2019), it is still uncommon to have

479 multiple closely related whole genome sequences that are of sufficient quality to compare for
480 karyotype differences. At the same time, the proliferation of reduced representation genome
481 sequencing methods means that it is easy to generate many genetic markers for non-model species
482 and produce very dense genetic maps. Furthermore, syntR allows comparisons to include older genetic
483 map data that would otherwise go unused. The simplicity of the syntR algorithm will facilitate rapid
484 karyotype mapping in a wide range of taxa.
485
486 We also believe that syntR provides a baseline for the development of further computational and
487 statistical methods for the study of chromosomal rearrangements. One fruitful direction would be to
488 integrate the syntR algorithm for synteny block detection directly into the genetic map building
489 process (much like GOOGA, Flagel *et al.* 2019). Another key extension would be to allow syntR to
490 compare multiple genetic maps simultaneously to detect synteny blocks in a group of species (e.g., by
491 leveraging information across species). Finally, formal statistical methods for evaluating the model fit
492 and the uncertainty involved with any set of synteny blocks would be a major (albeit challenging)
493 improvement to all existing methods, including syntR.

## 494 The similarity of *H. petiolaris* maps to previous studies

495
496 Compared with previous work, we found more inversions and fewer translocations between *H.*
497 *petiolaris* subspecies and *H. annuus* (Rieseberg *et al.* 1995, Burke *et al.* 2004). This is probably due to a
498 combination of factors. First, there appears to be karyotypic variation within some *Helianthus* species
499 (Heiser 1948, Heiser 1961, Chandler *et al.* 1986). Second, the maps presented here are made up of
500 more markers and individuals, which allowed us to identify small inversions that were previously
501 undetected as well as to eliminate false linkages that can be problematic in small mapping populations.
502 Lastly, we required more evidence to call rearrangements. Although we recovered some of the
503 translocations supported by multiple markers in Rieseberg *et al.* (1995) and Burke *et al.* (2004), we did
504 not recover any of the translocations supported by only a single sequence-based marker. Given the
505 high proportion of “rogue markers” in our maps, it is likely that some of the putative translocations
506 recovered in those earlier comparisons are the result of the same phenomenon.

507

508 On the other hand, we found that rearrangements between our *H. petiolaris* maps match the
509 translocations predicted from cytological studies quite well. Heiser (1961) predicted that *H. petiolaris*
510 ssp. *petiolaris* and *H. petiolaris* ssp. *fallax* karyotypes would have three chromosomes involved in two
511 translocations that form a ring during pairing at meiosis, as well as the possibility of a second
512 independent rearrangement. This exact configuration is likely to occur at meiosis in hybrids between
513 the *H. petiolaris* subspecies maps we present here (Fig S16). Also, the most noteworthy chromosome
514 configuration in cytological studies of *H. annuus-H. petiolaris* hybrids (Heiser 1947, Whelan 1979,
515 Ferriera 1980, Chandler *et al.* 1986) was a hexavalent (a six-chromosome structure) plus a quadrivalent
516 (a four-chromosome structure). Again, this is the configuration that we would expect in a hybrid
517 between *H. annuus* and the *H. petiolaris* ssp. *petiolaris* individual mapped here. Furthermore, the
518 complicated arrangement and relatively small size of A12, A16 and A17 synteny blocks in *H. petiolaris*
519 might explain why cytological configurations in *H. annuus-H. petiolaris* hybrids are so variable.
520 Interestingly, the rearrangements identified between *H. argophyllus* and *H. annuus* karyotypes here
521 and in Barb *et al.* (2014) also match the cytological studies better than an earlier comparison of sparse
522 genetic maps (Heesacker *et al.* 2009). It seems that, in systems with the potential for high proportions
523 of rogue markers, many markers are needed to identify chromosomal rearrangements reliably.
524

## 525 Total rearrangement rates

526

527 Our data suggest that annual sunflowers experience approximately 7.9 chromosomal rearrangements
528 per million years. This rate overlaps with recent estimates for this group (7.4-10.3, Barb *et al.* 2014)
529 and is even higher than the estimate that highlighted sunflower as a group with exceptionally fast
530 chromosomal evolution (5.5-7.3, Burke *et al.* 2004). However, since Burke *et al.* (2004), chromosomal
531 rearrangements have been tracked in many additional groups, including mammals (Ferguson-Smith
532 and Trifonov 2007, Martinez *et al.* 2016, da Silva *et al.* 2019), fish (Molina *et al.* 2014, Ayres-Alves *et al.*
533 2017), insects (Rueppell *et al.* 2016, Corbett-Detig *et al.* 2019), fungi (Sun *et al.* 2017) and plants
534 (Yogeeswaran *et al.* 2005, Schranz *et al.* 2006, Huang *et al.* 2009, Vogel *et al.* 2010, Latta *et al.* 2019).

535 Of these analyses, relatively few have systematically studied karyotypes evolution across multiple
536 species and estimated total rearrangement rates. Of those that do, most studies report less than 7.9
537 chromosomal rearrangements per million years, for example, in *Solanum* (0.36-1.44, Wu and Tanksley
538 2010), *Drosophila* (0.44-2.74, Bhutkar *et al.* 2008) and mammals (0.05-2.76, Murphy *et al.* 2005). But
539 there are exceptions, such as a comparison of genome sequences that revealed up to 35.7
540 rearrangements per million years in some grass lineages (Dvorak *et al.* 2018).
541
542 At the same time, we are likely underestimating rearrangement rates here for two reasons. First, we
543 used conservative thresholds for calling rearrangements. For example, some proportion of the rogue
544 markers that we identified could be the result of very small but real chromosomal rearrangements.
545 Second, our ability to resolve very small synteny blocks and breakpoints between synteny blocks
546 depends on marker density. Until we have full genome sequences to compare (like for the grass
547 lineages), we could be failing to detect very small rearrangements and falsely inferring that
548 independent rearrangements are shared. However, regardless of just how much we are
549 underestimating the rate, sunflower chromosomes are evolving quickly. This high rate of chromosomal
550 evolution could be a consequence of a higher rate of chromosomal mutation, a decreased chance that
551 chromosomal polymorphisms are lost, or both processes.
552

553 ## Type of rearrangements

554
555 We found that inversions and interchromosomal translocations dominate chromosomal evolution in
556 *Helianthus.* This pattern is common in angiosperm lineages (Weiss-Schneeweiss and Schneeweis 2012)
557 and fits with the consistent chromosome counts across annual sunflowers (2n = 34, Chandler *et al.*
558 1986). In addition, we found more inversions than translocations, which has previously been seen in
559 both plant (Wu and Tanksley 2010, Amores *et al.* 2014) and animal systems (Rueppell *et al.* 2016) and
560 echoes general reports that intrachromosomal rearrangements are more common than
561 interchromosomal rearrangements (Pevzner and Tesler 2003). These consistent rate differences are
562 notable because, although both rearrangement types depend on double strand breaks, two of the

563 major consequences of chromosomal rearrangements, underdominance (i.e., rearrangement
564 heterozygotes are less fit than either homozygote) and recombination modification, might be more
565 common for some types of rearrangements.
566
567 Translocations have a more predictable effect on hybrid fertility, while inversions consistently reduce
568 recombination. Reciprocal translocation heterozygotes can affect fertility because missegregation
569 during meiosis can cause half of the gametes to be unbalanced and thus inviable (White 1973, King
570 1993). Although inversion heterozygotes can also produce unbalanced gametes, whether that happens
571 is dependent on the size of the inversion and whether disrupted pairing during meiosis inhibits
572 crossovers (Searle 1993). When inversions are small or have suppressed crossing over, they will not be
573 strongly underdominant. On the other hand, inversions often exhibit reduced recombination either
574 because recombination is suppressed through disrupted pairing (Searle 1993) or ineffective through
575 the production of inviable gametes (Rieseberg 2001). While interactions between reduced
576 recombination and adaptation with gene flow have been extensively examined in the case of
577 inversions (Kirkpatrick and Barton 2006, Hoffman and Rieseberg 2008, Yeaman and Whitlock 2011,
578 Yeaman 2013), it is not clear whether the same pattern will be common for translocations (but see
579 Fishman *et al.* 2013, Stathos and Fishman 2014 for one example). Translocations bring together
580 previously unlinked alleles and mispairing at translocation breakpoints could suppress crossing over,
581 but recombination inside reciprocal translocations will not necessarily produce inviable gametes and
582 thus reduce effective recombination.
583
584 Although any selective force could be responsible for the evolution of any chromosomal
585 rearrangement, potential differences in the relative magnitude of underdominance versus
586 recombination suppression may contribute to the evolution of sunflower chromosomes. While many
587 chromosomal rearrangements in sunflowers appear to be strongly underdominant (Chandler 1986, Lai
588 *et al.* 2005), inversions typically are not (L. Rieseberg, unpublished). If translocations tend to be more
589 underdominant than inversions, they would be less likely to evolve through drift and more likely to
590 cause reproductive isolation directly. This could explain why translocations are less common than
591 inversions and why pollen viability is accurately predicted by the number of translocations inferred

592 from cytological studies (Chandler *et al.* 1986). At the same time, recent genomic analyses have
593 identified several extensive regions of very low recombination caused by large inversions segregating
594 in natural sunflower populations (Todesco *et al.* 2019, Huang *et al.* 2019). Mutations that segregate for
595 extended periods are unlikely to be strongly underdominant, and these inversions are associated with
596 multiple adaptive alleles (Todesco *et al.* 2019), which is consistent with a role for selection in their
597 origin or maintenance.
598

## 599 Non-random chromosomal rearrangement

600

601 We also found that some sunflower chromosomes are involved in more translocations than others.
602 This pattern has been observed in wheat (Badaeva *et al.* 2007) and breakpoint reuse is a common
603 phenomenon in comparative studies of karyotypes (Pevzner and Tesler 2003, Bailey *et al.* 2004,
604 Murphy *et al.* 2005, Larkin *et al.* 2009). Many studies support the idea that chromosomal regions with
605 greater sequence similarity are more likely to recombine and thus potentially generate novel
606 chromosomal arrangements. Some of the clearest examples of this come from the polyploidy
607 literature, where chromosomes with ancestral homology are more likely to recombine (Nicolas *et al.*
608 2007, Marone *et al.* 2012, Mason *et al.* 2014, Tennessen *et al.* 2014, Nguepjop *et al.* 2016). However,
609 centromeres and other repetitive regions can also affect the rate of mutations that cause
610 chromosomal rearrangements (Hardison *et al.* 2003, Murphy *et al.* 2005, Raskina *et al.* 2008, Molnár *et*
611 *al.* 2010, Vitte *et al.* 2014, Ayers-Alves *et al.* 2017, Li *et al.* 2017, Corbett-Detig *et al.* 2019). Given that
612 sunflowers have several genome duplications and a burst of transposable element activity in their
613 evolutionary history (Barker *et al.* 2008, Kawakami *et al.* 2011, Staton *et al.* 2012, Barker *et al.* 2016,
614 Badouin *et al.* 2017) it is plausible that ancestral homology or repeat content could be associated with
615 translocation propensity.
616

617 Of the above possibilities, an association between repeated translocations and centromeres would be
618 particularly compelling. Beyond the repeat content of centromeres explaining non-random mutation
619 (Kawabe *et al.* 2006, Sun *et al.* 2017, but see Lin *et al.* 2018, Okita *et al.* 2019), the position and size of

620 centromeres on chromosomes is known to affect meiotic drive and thus the repositioning of
621 centromeres through rearrangement could cause non-random fixation of translocations (Kaszás *et al.*
622 1998, Chmátal *et al.* 2014, Zanders *et al.* 2014). The relative placement of centromeres has been
623 associated with chromosome evolution in *Brassica* (Schranz *et al.* 2006) and wheat (Badaeva *et al.*
624 2007), and associations between meiotic drive and chromosome evolution have been found in several
625 animal taxa (Bidau and Martí 2004, Palestis *et al.* 2004, Molina *et al.* 2014, Blackmon *et al.* 2019). In
626 sunflower, we see some hints that centromeric repeats might be associated with repeated
627 translocation. Using the locations of the centromere-specific retrotransposon sequence, HaCEN-LINE
628 (Nagaki *et al.* 2015), to roughly identify the locations of centromeres in our reference, we find that
629 some rearrangement breakpoints, for example, the section of A16 with a different position in each
630 map, are close to putative centromeres (Fig S17-S20). Although a more thorough analysis of
631 centromeric repeat locations and their association with rearrangement breakpoints is required to draw
632 firm conclusions about the importance of centromeres to chromosomal evolution in sunflower, the
633 development of reference sequences for wild sunflower species is underway, which will allow those
634 and other associations to be confirmed. Further, it is time to directly test for meiotic drive in this
635 system by examining the transmission of rearrangements that affect centromeres in gametes produced
636 by plants that have heterozygous karyotypes.
637

638 ## Conclusion

639

640 Understanding the evolution of chromosomal rearrangements remains a key challenge in evolutionary
641 genetics. By developing new software to systematically detect synteny blocks and building new genetic
642 maps, we show that sunflowers exhibit rapid and non-random patterns of chromosomal evolution.
643 These data generate specific and testable hypotheses about chromosomal evolution in sunflower. We
644 believe that our work will spur additional studies of karyotypic evolution and diversity, and ultimately
645 lead to a more comprehensive understanding of the interplay between chromosomal evolution and
646 speciation.

647 # Acknowledgments

648

649 We thank Jessica Barb for providing marker sequence data, Marcy Uyenoyama for help with our
650 random walk analysis, Greg Baute for sharing hybrid seed, Chris Grassa for growing seedlings and
651 sharing scripts, and both Marco Todesco and Nadia Chaidir for help in the lab. We also thank Jenn
652 Coughlan, Andrew MacDonald, Brook Moyers, Mariano Alvarez, Dolph Schluter, Darren Irwin, Sally
653 Otto, and three anonymous reviewers for thoughtful discussions and help with earlier drafts of this
654 manuscript. This work was supported by an NSERC Postgraduate Scholarship awarded to KLO and an
655 NSERC Discovery Grant awarded to LHR (327475).

656

657 # Author contributions

658

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

662 # References

663

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

667 Ayres-Alves T., A. L. Cardoso, C. Y. Nagamachi, L. M. de Sousa, J. C. Pieczarka *et al.*, 2017 Karyotypic
668 evolution and chromosomal organization of repetitive DNA sequences in species of *Panaque*,
669 *Panaqolus*, and *Scobinancistrus* (Siluriformes and Loricariidae) from the Amazon Basin. Zebrafish
670 14: 251–260.

671 Badaeva E. D., O. S. Dedkova, G. Gay, V. A. Pukhalskyi, A. V. Zelenin *et al.*, 2007 Chromosomal
672 rearrangements in wheat: their types and distribution. Génome 50: 907–926.

673

674

675 Badouin H., J. Gouzy, C. J. Grassa, F. Murat, S. E. Staton *et al.*, 2017 The sunflower genome provides
676 insights into oil metabolism, flowering and Asterid evolution. Nature 546: 148-152.

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

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

681 Barker M. S., N. C. Kane, M. Matvienko, A. Kozik, R. W. Michelmore *et al.*, 2008 Multiple
682 paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate
683 gene retention after millions of years. Molecular Biology and Evolution 25: 2445–2455.

684 Barker M. S., Z. Li, T. I. Kidder, C. R. Reardon, Z. Lai *et al.*, 2016 Most Compositae (Asteraceae) are
685 descendants of a paleohexaploid and all share a paleotetraploid ancestor with the Calyceraceae.
686 American Journal of Botany 103: 1203–1211.

687 Baute G. J., G. L. Owens, D. G. Bock, and L. H. Rieseberg, 2016 Genome-wide genotyping-by-sequencing
688 data provide a high-resolution view of wild Helianthus diversity, genetic structure, and interspecies
689 gene flow. American Journal of Botany 103: 2170-2177.

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

693 Bhutkar A., S. W. Schaeffer, S. M. Russo, M. Xu, T. F. Smith *et al.*, 2008 Chromosomal rearrangement
694 inferred from comparisons of 12 *Drosophila* genomes. Genetics 179: 1657–1680.

695 Bidau C. J., and D. A. Martí, 2004 B chromosomes and Robertsonian fusions of *Dichroplus pratensis*
696 (Acrididae): Intraspecific support for the centromeric drive theory. Cytogenet Genome Res 106:
697 347–350.

698 Bilton T. P., M. R.Schofield, M. A. Black, D. Chagné, P. L. Wilcox *et al.*, 2018 Accounting for errors in low
699 coverage high-throughput sequencing data when constructing genetic maps using biparental
700 outcrossed populations. Genetics 209: 65–76.

701 Blackmon H., J. Justison, I. Mayrose, and E. E. Goldberg, 2019 Meiotic drive shapes rates of karyotype
702 evolution in mammals. Evolution 73: 511–523.

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

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

707 Broman K. W., H. Wu, S. Sen, and G. A. Churchill, 2003 R/qtl: QTL mapping in experimental crosses.

708 Bioinformatics 19: 889–890.

709 Burke J. M., Z. Lai, M. Salmaso, T. Nakazato, S. Tang *et al.*, 2004 Comparative mapping and rapid
710 karyotypic evolution in the genus *Helianthus*. Genetics 167: 449–457.

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

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

715 Chmátal L., S. I. Gabriel, G. P. Mitsainas, J. Martínez-Vargas, J. Ventura *et al.*, 2014 Centromere strength
716 provides the cell biological basis for meiotic drive and karyotype evolution in mice. Current Biology
717 24: 2295–2300.

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

721 Corbett-Detig R. B., I. Said, M. Calzetta, M. Genetti, J. McBroome *et al.*, 2019 Fine-mapping complex
722 inversion breakpoints and investigating somatic pairing in the *Anopheles gambiae* species complex
723 using proximity-ligation sequencing. Genetics 213: 1495–1511.

724 da Silva W. O., J. C. Pieczarka, M. J. R. da Costa, M. A. Ferguson-Smith, P. C. M. O'Brien *et al.*, 2019
725 Chromosomal phylogeny and comparative chromosome painting among *Neacomys* species
726 (Rodentia, Sigmodontinae) from eastern Amazonia. BMC Evolutionary Biology 19: 1–13.

727 Danecek P., A. Auton, G. Abecasis, C. A. Albers, E. Banks *et al.*, 1000 Genomes Project Analysis Group,
728 2011 The variant call format and VCFtools. Bioinformatics 27: 2156–2158.

729 Darling A. C. E., B. Mau, F. R. Blattner, and N. T. Perna, 2004 Mauve: multiple alignment of conserved
730 genomic sequence with rearrangements. Genome Research 14: 1394–1403.

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

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

735 Drillon G., A. Carbone, and G. Fischer, 2014 SynChro: A fast and easy tool to reconstruct and visualize
736 synteny blocks along eukaryotic chromosomes. PLoS ONE 9: e92621–8.

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

740 Ferguson-Smith M. A., and V. Trifonov, 2007 Mammalian karyotype evolution. Nat Rev Genet 8: 950–
741 962.

742 Ferriera J. V., 1980 Introgressive hybrisization between *Helanthus annuus* L. and *Helianthus petiolaris*
743 Nutt. Mendeliana 4: 81–93.

744 Fishman L., A. Stathos, P. M.Beardsley, C. F. Williams, and J. P. Hill, 2013 Chromosomal rearrangements
745 and the genetics of reproductive barriers in *Mimulus* (monkey flowers). Evolution 67: 2547–2560.

746 Flagel L. E., B. K. Blackman, L. Fishman, P. J. Monnahan, A. Sweigart *et al.*, 2019 GOOGA: A platform to
747 synthesize mapping experiments and identify genomic structural diversity. PLoS Comput Biol 15:
748 e1006949–25.

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

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

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

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

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

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

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

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

766 Hoffmann A. A., and L. H. Rieseberg, 2008 Revisiting the impact of inversions in evolution: From
767 population genetic markers to drivers of adaptive shifts and speciation? Annu. Rev. Ecol. Evol. Syst.
768 39: 21–42.

769 Huang K., R. L. Andrew, G. L. Owens, K. L. Ostevik, and L. H. Rieseberg, 2019 Multiple chromosomal
770 inversions contribute to adaptive divergence of a dune sunflower ecotype. bioRxiv. doi:
771 10.1101/829622 (Preprint posted November 4, 2019).

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

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

Kaszás E., and J. B. Genetics, 1998 Meiotic transmission rates correlate with physical features of rearranged centromeres in maize. Genetics 150: 1683-1692.

Kawabe A., B. Hansson, J. Hagenblad, A. Forrest, and D. Charlesworth, 2006 Centromere locations and associated chromosome rearrangements in *Arabidopsis lyrata* and *A. thaliana.* Genetics 173: 1613–1619.

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

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

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

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

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

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

Latta R. G., W. A. Bekele, C. P. Wight, and N. A. Tinker, 2019 Comparative linkage mapping of diploid, tetraploid, and hexaploid Avena species suggests extensive chromosome rearrangement in ancestral diploids. Scientific Reports 9: 1–12.

Li H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. doi: 1303.3997v2. (Preprint posted May 26, 2013).

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

804 Li S. F., T. Su, G. Q. Cheng, B. X. Wang, X. Li *et al.*, 2017 Chromosome evolution in connection with
805 repetitive sequences and epigenetics in plants. Genes 8: 290–19.

806 Lin C. Y., A. Shukla, J. Grady, J. Fink, E. Dray *et al.*, 2018 Translocation breakpoints preferentially occur
807 in euchromatin and acrocentric chromosomes. Cancers 10: 13–19.

808 Mandel J. R., S. Nambeesan, J. E. Bowers, L. F. Marek, D. Ebert *et al.*, 2013 Association mapping and the
809 genomic consequences of selection in sunflower. PLoS Genetics 9: e1003378.

810 Marone D., G. Laidò, A. Gadaleta, P.Colasuonno, D. B. M. Ficco *et al.* Mastrangelo A. M., 2012 A high-
811 density consensus map of A and B wheat genomes. Theor Appl Genet 125: 1619–1638.

812 Martinez P. A., U. P. Jacobina, R. V. Fernandes, C. Brito, C. Penone *et al.* 2016 A comparative study on
813 karyotypic diversification rate in mammals. Heredity 118: 366–373.

814 Mason A. S., M. N. Nelson, J. Takahira, W. A. Cowling, G. M. Alves *et al.* 2014 The fate of chromosomes
815 and alleles in an allohexaploid *Brassica* population. Genetics 197: 273–283.

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

818 Matvienko M., A. Kozik, L. Froenicke, D. Lavelle, B. Martineau *et al.*, 2013 Consequences of normalizing
819 transcriptomic and genomic libraries of plant genomes using a duplex-specific nuclease and
820 tetramethylammonium chloride. PLoS ONE 8: e55913–17.

821 McKenna A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The Genome Analysis Toolkit:
822 A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research
823 20: 1297–1303.

824 Molina W. F., P. A. Martinez, L. A. C. Bertollo, and C. J. Bidau, 2014 Evidence for meiotic drive as an
825 explanation for karyotype changes in fishes. Marine Genomics 15: 29–34.

826 Molnár I., M. Cifuentes, A. Schneider, E. Benavente, and M. Molnár-Láng, 2010 Association between
827 simple sequence repeat-rich chromosome regions and intergenomic translocation breakpoints in
828 natural populations of allopolyploid wild wheats. Annals of Botany 107: 65–76.

829 Murphy W. J., D. M. Larkin, A. Everts-van der Wind, G. Bourque, G. Tesler *et al.*, 2005 Dynamics of
830 mammalian chromosome evolution inferred from multispecies comparative maps. Science 309:
831 613–617.

832 Nagaki K., K. Tanaka, N. Yamaji, H. Kobayashi, and M. Murata, 2015 Sunflower centromeres consist of a
833 centromere-specific LINE and a chromosome-specific tandem repeat. Front. Plant Sci. 6: 1-12.

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

836 Nguepjop J. R., H. A. Tossim, J. M. Bell, J. F. Rami, S. Sharma *et al.* 2016 Evidence of genomic exchanges
837 between homeologous chromosomes in a cross of peanut with newly synthetized allotetraploid
838 hybrids. Front. Plant Sci. 7: 87–12.

839 Nicolas S. D., G. L. Mignon, F. Eber, O. Coriton, H. Monod *et al*., 2007 Homeologous recombination
840 plays a major role in chromosome rearrangements that occur during meiosis of *Brassica napus*
841 haploids. Genetics 175: 487–503.

842 Noor M. A., K. L. Grams, L. A. Bertucci, and J. Reiland, 2001 Chromosomal inversions and the
843 reproductive isolation of species. Proceedings of the National Academy of Sciences 98: 12084–
844 12088.

845 Okita A. K., F. Zafar, J. Su, D. Weerasekara, T. Kajitani *et al.* 2019 Heterochromatin suppresses gross
846 chromosomal rearrangements at centromeres by repressing Tfs1/TFIIS-dependent transcription.
847 Communications Biology 2: 1–13.

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

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

851 Palestis B. G., A. Burt, R. N. Jones, and R. Trivers, 2004 B chromosomes are more frequent in mammals
852 with acrocentric karyotypes: Support for the theory of centromeric. drive. Proc. Biol. Sci. 271: 1–3.

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

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

858 Quillet M. C., N. Madjidian, Y. Griveau, H. Serieys, M. Tersac *et al.* 1995 Mapping genetic factors
859 controlling pollen viability in an interspecific cross in *Helianthus* sect. *Helianthus*. Theor Appl Genet
860 91: 1195–1202.

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

864 Raskina O., J. C. Barber, E. Nevo, and A. Belyayev, 2008 Repetitive DNA and chromosomal
865 rearrangements: speciation-related events in plant genomes. Cytogenet Genome Res 120: 351–
866 357.

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

869 Rieseberg L. H., 2001 Chromosomal rearrangements and speciation. Trends in Ecology & Evolution 16:
870 351–358.

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

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

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

877 Rueppell O., R. Kuster, K. Miller, B. Fouks, S. Rubio Correa *et al.*, 2016 A new metazoan recombination
878 rate record and consistently high recombination rates in the honey bee genus *Apis* accompanied
879 by frequent inversions but not translocations. Genome Biol Evol 8: 3653-3660.

880 Sambatti J. B. M., J. L. Strasburg, D. Ortiz-Barrientos, E. J. Baack, and L. H. Rieseberg, 2012 Reconciling
881 extremely strong barriers with high levels of gene exchange in annual sunflowers. Evolution 66:
882 1459–1473.

883 Schlautman B., L. Diaz-Garcia, G. Covarrubias-Pazaran, N. Schlautman, N. Vorsa *et al.*, 2017
884 Comparative genetic mapping reveals synteny and collinearity between the American cranberry
885 and diploid blueberry genomes. Molecular Breeding 38: 1-19.

886 Schranz M. E., T. Mitchell-Olds, and M. A. Lysak, 2006 The ABC's of comparative genomics in the
887 Brassicaceae: Building blocks of crucifer genomes. Trends in Plant Science 11: 535–542.

888 Searle J. B., 1993 Chromosomal hybrid zones in eutherian mammals, pp. 309–353 in Hybrid zones and
889 the evolutionary process, edited by R. G. Harrison. Oxford University Press on Demand.

890 Shagina I., E. Bogdanova, I. Mamedov, Y. Lebedev, S. Lukyanov *et al.* 2010 Normalization of genomic
891 DNA using duplex-specific nuclease. Biotechniques 48: 455–459.

892 Sinha A. U., and J. Meller, 2007 Cinteny: Flexible analysis and visualization of synteny and genome
893 rearrangements in multiple organisms. BMC Bioinformatics 8: 82–9.

894 Soderlund C. M. Bomhoff, and W. M. Nelson, 2011 SyMAP v3.4: A turnkey synteny system with
895 application to plant genomes. Nucleic Acids Research 39: e68–e68.

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

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

901 Stephens J. D., W. L. Rogers, C. M. Mason, L. A. Donovan, and R. L. Malmberg, 2015 Species tree
902 estimation of diploid *Helianthus* (Asteraceae) using target enrichment. American Journal of Botany
903 102: 910–920.

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

907 Sun S., V. Yadav, R. B. Billmyre, C. A. Cuomo, M. Nowrousian *et al.* 2017 Fungal genome and mating
908 system transitions facilitated by chromosomal translocations involving intercentromeric
909 recombination. PLoS Biol 15: e2002527–31.

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

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

914 Tennessen J. A., R. Govindarajulu, T. L. Ashman, and A. Liston, 2014 Evolutionary origins and dynamics
915 of octoploid strawberry subgenomes revealed by dense targeted capture linkage maps. Genome
916 Biol Evol 6: 3295–3313.

917 Todesco M., G. L. Owens, N. Bercovich, J. S. Légaré, S. Soudi *et al.* 2019 Massive haplotypes underlie
918 ecotypic differentiation in sunflowers. bioRxiv. doi: 10.1101/790279 (Preprint posted October 2,
919 2019)

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

922 Van der Auwera G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel *et al.*, 2013 From fastQ data to
923 high-confidence variant calls: The genome analysis toolkit best practices pipeline. Curr. Protoc.
924 Bioinformatics 43: 11.10.1–33.

925 Vitte C., M. A. Fustier, K. Alix, and M. I. Tenaillon, 2014 The bright side of transposons in crop evolution.
926 Brief. Funct. Genomics 13: 276–295.

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

929 Weiss-Schneeweiss H., and G. M. Schneeweiss, 2012 Karyotype Diversity and Evolutionary Trends in
930 Angiosperms, pp. 209-230 in *Plant Genome Diversity Volume 2: Physical Structure, Behaviour and*
931 *Evolution of Plant Genomes.* Edited by J. Greilhuber, J. Dolezel and J. F. Wendel. Springer Science &
932 Business Media

933 Whelan E. D., 1979 Interspecific hybrids between *Helianthus petiolaris* Nutt. and *H. annuus* L.: Effect of
934 backcrossing on meiosis. Euphytica 28: 297-308.

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

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

937 Wu F., and S. D. Tanksley, 2010 Chromosomal evolution in the plant family Solanaceae. BMC Genomics
938 11: 182.

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

941 Yeaman S., and M. Whitlock, 2011 The genetic architecture of adaptation under migration-selection
942 balance. Evolution 65: 1897–1911.

943 Yogeeswaran K., A. Frary, T. L. York, A. Amenta, A. H. Lesser *et al.* 2005 Comparative genome analyses
944 of *Arabidopsis* spp.: inferring chromosomal rearrangement events in the evolutionary history of *A.*
945 *thaliana.* Genome Research 15: 505–515.

946 Zanders S. E., M. T. Eickbush, J. S. Yu, J. W. Kang, K. R. Fowler *et al.* 2014 Genome rearrangements and
947 pervasive meiotic drive cause hybrid infertility in fission yeast. eLife 3: 419–23.

948 Zhulidov P. A., 2004 Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic
949 Acids Research 32: e37.