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
introgression (Noor et al. 2001, Rieseberg 2001), accumulating genetic incompatibilities (Navarro and Barton 2003), and simplifying reinforcement (Trickett and Butlin 1994).

Despite their prevalence, the fixation of underdominant chromosomal rearrangements is still not well understood. The simplest model is that in sufficiently small populations, underdominant rearrangements can be stochastically fixed by genetic drift (Lewis 1966, Templeton 1981, Grant 1981). However, several other mechanisms have been proposed to explain chromosomal evolution (Rieseberg 2001). These include individually neutral chromosome fissions and fusions (Baker and Bickham 1986, Britton-Davidian 2000), meiotic drive (Novitski 1951, White 1978, Chmáňal et al. 2014) and association with locally adapted genes (Kirkpatrick and Barton 2006, Lowry and Willis 2010). The importance of each of these mechanisms will largely depend on the types of rearrangements fixed (e.g., chromosome fusions, pericentric inversions, or reciprocal translocations) and the cost/benefit of the genes associated with them. Consequently, mapping and characterizing chromosomal rearrangements is a critical step towards understanding their evolutionary dynamics.

The genus *Helianthus* is well known to have particularly labile genome structure. These sunflowers have several paleopolyploidy events in their evolutionary history (Barker et al. 2008), 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 extremely 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). Furthermore, while *Helianthus* species generally have large effective population sizes (but see Stebbins and Daly 1961, Carney et al. 2000), where drift should be relatively weak, chromosomal rearrangements appear to be strongly underdominant in this group (Chandler 1986, Lai et al. 2005). This means that chromosomal evolution is less likely to be driven by drift alone and more likely to be an important driver of reproductive isolation in sunflower.
Studying chromosomal evolution within any group requires high-density genetic maps and a method for systematic comparison of karyotypes. Recently, high-density genetic maps of *H. niveus* ssp. *tephrodes* and *H. argophyllus* were built and compared to *H. annuus* (Barb et al. 2014). This analysis clearly mapped previously inferred karyotypes (Heiser 1951, Chandler 1986, Quillet et al. 1995) and showed that chromosomal rearrangements affect introgression between sunflowers. That said, creating additional high-density maps for related species would yield a fuller picture of chromosomal evolution.

Previously constructed genetic maps with limited marker density suggest that several chromosomal rearrangements differentiate *H. annuus* and *H. petiolaris* (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). Creating high-density genetic maps for *H. petiolaris* and adding them to the above analysis will allow us to (1) precisely track additional rearrangements and (2) reconstruct ancestral karyotypes for the group, untangling overlapping rearrangements that can be obscured by directly comparing present-day karyotypes and that are likely to occur in lineages with extensive chromosomal rearrangement like sunflower. At the same time, these *H. petiolaris* maps will be useful to breeders looking to introgress beneficial alleles from crop wild relatives into domesticated sunflower lineages (Chetelat and Meglic 2000, Foulongne et al. 2003, Dirlewanger et al. 2004).

A key part of a multi-species comparative study of chromosome evolution is a systematic and repeatable method for identifying syntenic chromosomal regions (*sensu* Pevzner and Tesler 2003) using genetic map data. This is 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 counting all differences in marker order (Wu and Tanksley 2010), by visual inspection (Burke et al. 2004), or by manually applying simple rules like size thresholds (Heesacker et al. 2009, Barb et al. 2014) 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). 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.
The phylogenetic relationships between the sunflower taxa used in this study. A) Phylogenetic 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
Agriculture. When it was originally collected, accession PI435768 was classified *H. neglectus*. However, based on the location of the collection (Heiser 1961) and a more recent genetic analysis of the scale of differences between *H. petiolaris* ssp. *fallax* and *H. neglectus* (Raduski et al. 2010), we believe that this accession should be classified *H. petiolaris* ssp. *fallax*.

**Genotyping**

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

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

To call variants, we used a pipeline that combines the Burrows–Wheeler Aligner version 0.7.15 (BWA, 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 ‘HaplotypCaller’, 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.
Using the corrected genotypes, we made new linkage groups with only the most reliable markers. Namely, we used the function ‘mstmap.cross’ (with the parameter values: dist.fun = "kosambi", p.value = \(1 \times 10^{-6}\), noMap.size = 2, noMap.dist = 5) on markers with less than 10% missing data and without significant segregation distortion. We refined the resulting linkage groups by removing (1) markers with more than three double crossovers, (2) markers with aberrant segregation patterns (segregation distortion more than two standard deviations above or below the mean segregation distortion of the nearest 20 markers), and (3) linkage groups made up of fewer than four markers.

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

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

\textbf{SyntR development}

To aid in the identification of chromosomal rearrangements, we developed the R package ‘syntR’ (code and documentation available at \url{http://ksamuk.github.io/syntR}). This package implements a heuristic algorithm for systematically detecting synteny blocks from marker positions in two genetic maps. In short, syntR combines a biologically-informed noise reduction method and a cluster identification method better suited for detecting linear (as opposed to circular) clusters of data points.
The algorithm compares sets of homologous genetic markers from two genetic maps, clusters the markers into synteny blocks using a method that is similar to the “friends-of-friends” clustering algorithm (Huchra and Geller 1982), and identifies breakpoints. The algorithm correctly recovers breakpoints in simulated genetic maps under biologically reasonable scenarios and performs at least well as prior methods while being systematic and repeatable. We fully describe this method and its performance in the supplemental materials.

Finding synteny blocks

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

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

Table 1 – The syntR tuning parameters used to identify syntenic 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             |

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

Karyotype reconstruction and analysis

We used our inferred syntenic blocks and the software MGR v 2.01 (Bourque and Pevzner 2002) to infer ancestral karyotypes for our five *Helianthus* taxa and to determine the number of chromosomal rearrangements that occurred along each branch of the species tree. To run the MGR analysis, we needed the order and orientations of syntenic blocks in all five maps. However, individual syntenic blocks were often missing from one or more of our final maps. We took two approaches to addressing this problem. First, we inferred the likely position of missing syntenic blocks based on the location of markers that were too sparse to be grouped together by syntR and matched the location of syntenic blocks in other maps. In the second case, we simply dropped any syntenic blocks that were not universally represented. Because we already had two sets of syntenic blocks for each map (curated and individually optimized), we ran the MGR analyses using three different sets of syntenic blocks: (set 1) curated and inferred, (set 2) curated and present in all five maps, (set 3) individually optimized and 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* ssp. *fallax* - full     | 13335     | 2178        |
| *H. petiolaris* ssp. *petiolaris* - thinned | 2462     | 1576        |
| *H. petiolaris* ssp. *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
within the *H. petiolaris* ssp. *fallax* individual used for crosses (Fig S12). That is, this individual was homozygous for most of F2 and F4. Notably, we find the same pattern on the distal half of *H. annuus* chromosome 7 and find that this region is also not represented in the *H. petiolaris* ssp. *fallax* map.

Figure 2 – *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 \textit{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 \textit{H. petiolaris} ssp. \textit{petiolaris} and \textit{H. annuus} and 79 genetic regions that are syntenic between the \textit{H. petiolaris} ssp. \textit{fallax} and \textit{H. annuus} (Fig 2). We also recovered synteny blocks for the \textit{H. petiolaris} ssp. \textit{tephrodes} and \textit{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 \textit{H. annuus} and four other \textit{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 \textit{H. annuus} chromosome. The other markers are considered rogue.

| Genetic map            | N synteny blocks | Rogue markers | Map coverage | \textit{H. annuus} coverage | Collinear | Inverted | Translocated |
|------------------------|------------------|--------------|--------------|----------------------------|-----------|-----------|-------------|
| \textit{H. petiolaris} ssp. \textit{petiolaris} | 97               | 19%          | 80%          | 74%                        | 39%       | 36%       | 26%         |
| \textit{H. petiolaris} ssp. \textit{fallax}    | 79               | 17%          | 63%          | 65%                        | 32%       | 34%       | 34%         |
| \textit{H. niveus} ssp. \textit{tephrodes}     | 43               | 26%          | 78%          | 75%                        | 40%       | 21%       | 39%         |
| \textit{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 \textit{H. annuus} genetic map in direct comparisons (Table 2). However, because nested and shared
rearrangements can obscure patterns of chromosome evolution, we use the MGR analyses to predict the most likely sequence of rearrangements in a phylogenetic context before quantifying rearrangement rate.

Karotype reconstruction and chromosomal rearrangement

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

Table 4 – Overall patterns of chromosomal rearrangement inferred based on different sets of synteny blocks. Set 1 synteny blocks are curated based on multiple syntR runs and inferred when missing. Set 2 synteny blocks are curated but present in all five maps. Set 3 synteny blocks are the output from individually optimized syntR runs that are present in all five maps. Rate 1 is the number of rearrangements per million years based on a cumulative divergence time of 9.1 million years, while rate 2 is the based on 26 million years. P-value 1 is the probability of seeing the observed number of inversions and translocations if the rate of inversion was equal to the rate of translocation. P-value 2 is the probability of seeing the observed number of chromosomes involved in 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 x 10^{-11}                   | 8.0 x 10^{-8}                   |
| Set 2          | 67       | 45           | 15       | 6.6    | 2.3    | 7                           | 1.3 x 10^{-4}                    | 5.3 x 10^{-11}                   |
| Set 3          | 76       | 50           | 10       | 6.6    | 2.3    | 7                           | 1.6 x 10^{-7}                    | 3.0 x 10^{-6}                    |
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
reliability put an upper bound on this estimate because the lower bounds of divergence time estimates within this groups overlap 0. However, at 2.3 rearrangements per million years, even our most conservative estimate of rearrangement for this group, which uses the less complete sets of synteny blocks and the upper bounds of divergence time estimates, is among the highest measurements reported in Burke et al. (2004).

These 88 rearrangements are made up of 74 inversions and 14 translocations that are distributed across the phylogeny. Except for the branch between reconstructed karyotypes 1 and 2, where there are no translocations and many inversions, each branch has 2-12 inversions and 1-3 translocations. Although inversion and translocation rates are fairly even across the tree, it is unlikely that the rate of inversions is equal to the rate of translocation (binomial test, 5.1x10^{-11}). Furthermore, we find that only 8 of the 17 chromosomes are involved in the 14 translocations identified on the phylogeny. This extreme asymmetry is highly unlikely to have happened by chance if translocation is equally likely for all chromosomes (p = 8.0x10^{-8}, Fig S22). Therefore, this pattern of rearrangement suggests that some chromosomes are more likely than others to be involved in translocations. In line with this observation, we see that some sections of chromosomes are frequently found in new positions. For example, we see that A4 and A7 are involved in repeated translocations and that part of A6 has a different position in almost every map.

Discussion

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

Similarity of *H. petiolaris* maps to previous studies

Compared with previous work, we found more inversions and fewer translocations between *H. petiolaris* subspecies and *H. annuus* (Rieseberg et al. 1995, Burke et al. 2004). This is probably due to a combination of factors. First, there appears to be karyotypic variation within some *Helianthus* species (Heiser 1948, Heiser 1961, Chandler et al. 1986) and the individuals used to make maps in each study were not identical. Second, the maps presented here are made up of more markers and individuals, which allowed us to identify small inversions that were undetected in the previous comparisons, as well as to eliminate false linkages that can be problematic with small mapping populations. Lastly, we required more evidence to call rearrangements, especially translocations. Although we did recover some of the translocations that were supported multiple markers in Rieseberg et al. (1995) and Burke et al. (2004), for example a translocation between A6 and A15 in *H. petiolaris ssp. fallax* and the fragmentation of A16 across multiple chromosomes in both *H. petiolaris* maps, we did not recover any of the translocations supported by only a single sequence-based marker. Given the high proportion of “rogue markers” in our maps it is likely that some of the putative translocations recovered in those earlier comparisons are the result of the same phenomenon.

On the other hand, we found that rearrangements between our *H. petiolaris* maps match the translocations predicted from cytological studies quite well. Heiser (1961) predicted that *H. petiolaris* ssp. *petiolaris* and *H. petiolaris* ssp. *fallax* karyotypes would have three chromosomes involved in two translocations that form a ring during pairing at meiosis, as well as the possibility of a second independent rearrangement. This configuration of chromosomes would occur at meiosis in hybrids between the *H. petiolaris* subspecies maps we present here (Fig S23). Furthermore, one of the cytological configurations that included the most chromosomes in *H. petiolaris* and *H. annuus* hybrids reported by Whelan (1979) and one of the most common configurations in other cytological studies (Heiser 1947, Ferriera 1980, Chandler et al. 1986) was a hexavalent (a six-chromosome structure) plus...
a quadrivalent (a four-chromosome structure). Again, this is the configuration that we would expect in a hybrid between the *H. petiolaris* ssp. *petiolaris* and *H. annuus* individuals mapped here. Further, the complicated arrangement and relatively small size of A12, A16 and A17 synteny blocks in our *H. petiolaris* maps might explain why cytological chromosome configurations in hybrids between *H. petiolaris* and *H. annuus* are so variable. Interestingly, the rearrangements identified between *H. argophyllus* and *H. annuus* karyotypes here and in Barb et al. (2014) also match the cytological studies better than an earlier comparison of sparse genetic maps (Heesacker et al. 2009). It seems that in systems with the potential for high proportions of rogue markers, many markers are needed to reliably identify chromosomal rearrangements.

**Rearrangement rate**

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

**Type of rearrangements**

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

Reciprocal translocation heterozygotes affect fertility because of missegregation during meiosis and are expected to cause half of gametes to be unbalanced and therefore inviable (White 1973, King 1993). Inversion heterozygotes can also produce unbalanced gametes, but whether that happens is dependent on the size of the inversion and whether disrupted pairing during meiosis actually inhibits crossovers (Searle 1993). When inversions are small or have suppressed crossing over, they will not be strongly underdominant. Indeed, inversions typically are not strongly underdominant in sunflowers (L. Rieseberg, pers. comm.), which might explain why translocations are less common than inversions and
why pollen viability is accurately predicted by the number of translocations inferred from cytological studies (Chandler et al. 1986).

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

Non-random chromosomal rearrangement

We found that some sunflower chromosomes are involved in more translocations than others. A bias in the identity of chromosomes involved in translocation is common in species with sex chromosomes, where sex chromosomes are rearranged more often than autosomes (e.g., Neafsey et al. 2015), but this is not relevant for sunflowers. However, it is also known that centromeres and other repetitive regions can affect the rate of mutations that cause chromosomal rearrangements (Hardison et al. 2003, Vitte et al. 2014) and breakpoint reuse is a common phenomenon in comparative studies of karyotypes (Bailey et al. 2004, Larkin et al. 2009, Murphy et al. 2005, Pevzner and Tesler 2003).
Therefore, it is possible that the overactive chromosomes identified in this study are outliers for centromeric/repeat content. In fact, we see some hints that centromeric repeats might be associated with repeated translocation in this clade. Using the locations of the centromere specific retrotransposon sequence, HaCEN-LINE (Nagaki et al. 2015), to roughly identify the locations of centromeres in our reference, we find that some rearrangement breakpoints, for example the section of A16 with a different position in each map, are close to putative centromeres (Fig S24-S29).

An association between repeated translocations and centromere properties would be interesting because rearrangements that change the position of centromeres might cause meiotic drive to fix those rearrangements (Chmátal et al. 2014, Zanders et al. 2014). Although a more thorough analysis of centromeric repeat locations and their association with rearrangement breakpoints is required to draw strong conclusions about the importance of centromeres to chromosomal evolution in sunflower, the development of reference sequences for wild sunflower species is underway, which will allow those associations to be confirmed. Further, it is time to directly test for meiotic drive in this system by examining the transmission of rearrangements that affect centromeres in gametes produced by plants that have heterozygous karyotypes. However, regardless of how they arise, if translocations are primarily responsible for reducing pollen viability in sunflower and certain chromosomes are translocated more often, these chromosomes could be playing an outsized role in maintaining divergence between sunflower lineages.

Identifying rearrangements

Studying the evolution of chromosomal rearrangements requires dense genetic maps and systematic methods to analyze and compare these maps between species. Our new software, syntR, provides an end-to-end solution for systematic and repeatable identification of synteny blocks in pairs of genetic maps with any marker density. Our tests on real and simulated data find that syntR recovers chromosomal rearrangements identified by both previous manual comparisons and cytological study, suggesting that syntR is providing an accurate view of karyotypic differences between species.
Although the number of species with whole genome sequence and methods to detect synteny blocks from those sequences are rapidly accumulating, it is still rare to have multiple closely related whole genome sequences that are of sufficient quality to compare for karyotype differences. At the same time, the proliferation of reduced representation genome sequencing methods means that it is easy to generate many genetic markers and produce dense genetic maps. These maps coupled with this software will allow us to discover many karyotypes. Furthermore, syntR allows comparisons to include older genetic map data that would otherwise go unused. Overall, we believe syntR will be a valuable tool for the systematic study of chromosomal rearrangements in any species.

We also believe that syntR provides a baseline for the development of further computational and statistical methods for the study of chromosomal rearrangements. One fruitful direction would be to integrate syntR’s synteny block detection directly into the genetic map building process (much like GOOGA, Flagel et al. 2019). Another key extension would be to allow syntR to compare multiple genetic maps simultaneously to detect synteny blocks in a group of species (e.g. by leveraging information across species). Finally, formal statistical methods for evaluating the model fit and uncertainty involved with a particular inferred set of synteny blocks would be a major (albeit challenging) improvement to all existing methods, including syntR.

Conclusion

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

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

Author contributions

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Kawakami T., Dhakal P., Katterhenry A. N., Heatherington C. A., Ungerer M. C., 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., Barton N., 2006 Chromosome inversions, local adaptation and speciation. Genetics 173: 419–434.

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

Van der Auwera G. A., Carneiro M. O., Hartl C., Poplin R., del Angel G. et al., 2013 From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Current Protocols in Bioinformatics 43: 11.10.1–33.
Via S., 2012 Divergence hitchhiking and the spread of genomic isolation during ecological speciation with-gene-flow. Philosophical Transactions of the Royal Society B: Biological Sciences 367: 451–460.

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

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

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

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

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

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

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

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

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

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

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

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

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

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