

# Testing the evolutionary theory of inversion polymorphisms in the yellow monkeyflower (*Mimulus guttatus*)

Received: 16 May 2024

Accepted: 13 November 2024

Published online: 29 November 2024

 Check for updatesParis Veltsos <sup>1,2</sup>, Luis J. Madrigal-Roca <sup>1</sup> & John K. Kelly <sup>1</sup> 

Chromosomal inversions have been implicated in a remarkable range of natural phenomena, but it remains unclear how much they contribute to standing genetic variation. Here, we evaluate 64 inversions that segregate within a single natural population of the yellow monkeyflower (*Mimulus guttatus*). Nucleotide diversity patterns confirm low internal variation for the derived orientation (predicted by recent origin), elevated diversity between orientations (predicted by natural selection), and localized fluctuations (predicted by gene flux). Sequence divergence between orientations varies idiosyncratically by position, not following the suspension bridge pattern predicted if the breakpoints are the targets of selection. Genetic variation in gene expression is not inflated close to inversion breakpoints but is clearly partitioned between orientations. Like sequence variation, the pattern of expression variation suggests that the capture of coadapted alleles is more important than the breakpoints for the fitness effects of inversions. This work confirms several evolutionary predictions for inversion polymorphisms, but clarity emerges only by synthesizing estimates across many loci.

Chromosomal inversions have been an enduring focus of evolutionary biology. Dobzhansky's demonstration of natural selection on inversions of *Drosophila pseudoobscura*<sup>1</sup> was integral to the modern synthesis and inversions were routinely invoked in the extensive studies of 'supergenes' by the British ecological genetics school<sup>2–4</sup>. They can generate under-dominance for fertility if recombination within inverted regions generates genetically unbalanced gametes. Across both plants and animals, there is a remarkable association between inversions and meiotic drive<sup>5–9</sup>. Recent studies have implicated inversions in local adaptation<sup>10–16</sup>, accelerated speciation<sup>6,17–26</sup>, the evolution of migration and species ranges<sup>27,28</sup>, sexual selection and mate choice<sup>25,29,30</sup>, and sex-chromosome evolution<sup>29,31–34</sup>. Reviewing the literature, Faria et al.<sup>35</sup> conclude that the fixation of inversions by positive selection is essential to the evolution of many species.

Despite many examples, we have no quantitative sense for the contribution of inversions to genetic variation. The argument that inversions are fundamental to adaptation is enormously burdened by

ascertainment bias. Inversions have been studied when identified from conspicuous, often dramatic, phenotypic effects<sup>4,5,36–40</sup>. They are not scored like Single Nucleotide Polymorphisms (SNPs) where surveys provide an enumeration sufficient to test broad evolutionary hypotheses. For example, the relative frequency of synonymous versus non-synonymous SNPs can estimate the fraction of substitutions that are fixed by natural selection as opposed to genetic drift. In contrast, inversion studies tend towards depth—detailed investigation of individual variants—rather than breadth. In this investigation, we evaluate the quantitative importance of inversions by comprehensively enumerating variants among a random sample of ten genomes from one natural population.

A distinguishing feature of inversion polymorphisms is that the composition of alternative alleles changes through time. Upon emergence, the derived orientation is a single haplotype (nucleotide sequence), and in the short term, it is likely to remain internally homogeneous owing to suppressed recombination with the ancestral

<sup>1</sup>Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS, USA. <sup>2</sup>Present address: Ecology, Evolution and Genetics Research Group, Biology Department, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. ✉e-mail: [jkk@ku.edu](mailto:jkk@ku.edu)

orientation. The ancestral orientation should be internally variable, composed of many different haplotypes. The amount and pattern of variation will change through time contingent on selection and the position-dependent rate of gene flux. Flux refers to the combined effect of double crossovers and gene conversion exchanging variation between orientations<sup>41–43</sup>. A suspension bridge pattern<sup>44,45</sup>, where divergence between inversion orientations is greatest at the breakpoints and gradually declines to its lowest value in the middle, is predicted if gene flux occurs most frequently in the middle of inversions<sup>42,46</sup> or if there is epistatic selection on the breakpoints themselves<sup>47</sup>. To test these predictions, we score nucleotide sequence variation across inverted genome regions and compare the estimates within and between orientations to the genome wide pattern.

Nucleotide differences can reveal the historical signature of natural selection<sup>48,49</sup>, but they do not indicate how inversions generate phenotypic differences or how those phenotypic effects translate into fitness variation. Altered gene expression is frequently posited as the cause of phenotypic effects simply because the physical rearrangement of DNA can affect expression (a position effect). This is most likely for genes near the inversion breakpoints where rearrangement can move coding DNA away from regulatory elements such as promoters or enhancers. Alternatively, orientations may differ in gene expression because rearrangement captures pre-existing variation, e.g. SNPs or indels within the reversed sequence, and it is these variants that are responsible for expression differences. We test the breakpoint and capture hypotheses by reanalyzing data from a recently published eQTL mapping experiment. The eQTL experiment used the same collection of genomes that we score for inversions in the present paper enabling reanalysis of expression differences in direct relation to inversion genotype calls.

In this paper, we provide a comprehensive survey of inversion polymorphisms that segregate within a single natural population of *M. guttatus*. We interrogate 64 inversions, each of which contains at least two genes, for patterns of gene expression and nucleotide sequence variation. We test alternative hypotheses for how fitness variation emerges from inversions. The “breakpoint hypothesis” predicts elevated gene expression variation at genes neighboring inversion breakpoints, relative both to genes within the interior of inversions and the remainder of the colinear genome. We also expect higher sequence divergence close to breakpoints. The captured alleles hypothesis predicts no specificity to breakpoints (in either expression or sequence variation) but a non-random and asymmetric distribution of variation within and between ancestral and derived orientations. The selection regime almost certainly varies among loci in this population, but we find patterns that clearly favor the capture hypothesis over the breakpoint hypothesis.

## Results

### Inversion discovery

We identify 64 distinct inversions (summarized in Supplementary Data 1) from ten whole genome assemblies of *M. guttatus*. Each assembly is from an unrelated plant derived from one natural population (Iron Mountain, hereafter IM<sup>50,51</sup>). Each inversion is characterized by a stretch of sequence that aligns negatively to the reference genome with high homology (IM767, one of the ten assemblies, is used as the reference). Thirty five of the 64 inversions are present in multiple alternative lines, exhibiting the same breakpoints relative to the reference genome. We treat these as distinct copies of the same alternative allele. The size of inversions ranges from 5 kb to 4.1 mb (median  $\approx$  100 kb) and the number of genes within inversions ranges from 2 to 791 (median = 9). Our criteria for calling inversions are conservative and many small rearrangements, particularly those outside of genic regions or encompassing fewer than two genes, are not included in our survey (see “Genome assemblies and inversion calling” section of the Methods).

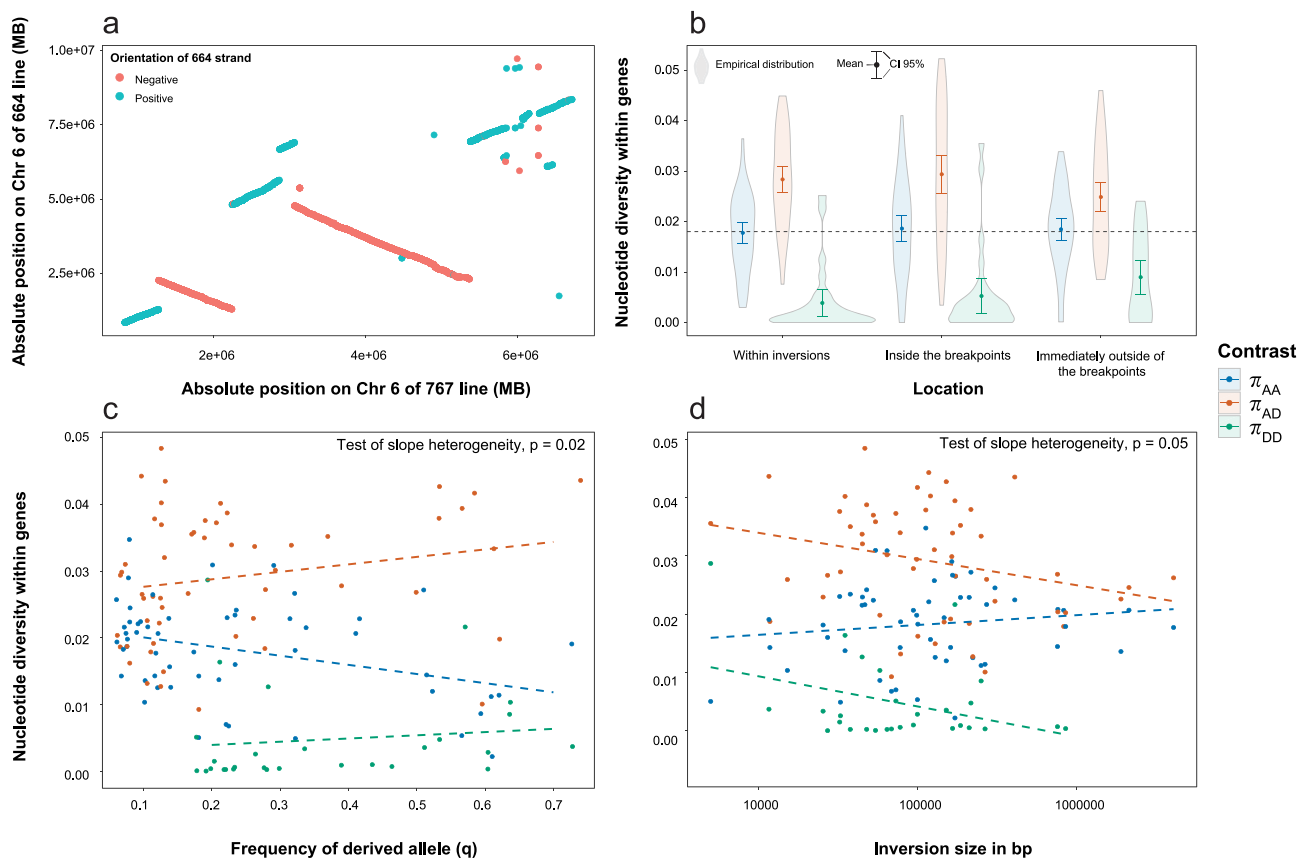
We find two large inversions, on chromosomes 6 and 11, previously discovered from genetic mapping studies<sup>40,52</sup>, but the current genome assemblies reveal aspects of their morphology. *inv\_24* (on chromosome 6, 1.3 mb to 5.4 mb coordinates on the IM767 reference) appears to be a multi-part rearrangement involving at least three mutational changes from the ancestral sequence (Fig. 1a). The middle portion of *inv\_24* is colinear with the ancestral orientation, but the entire region shows complete recombination suppression in inversion heterozygotes<sup>52,53</sup>. We also find two cases of nested inversion: *inv\_6a/inv\_6b* (chromosome 1) and *inv\_10a/inv\_10b* (chromosome 2). For each, two alternative lines exhibit inversions from the reference genome within the same general region, but with distinct breakpoints in each alternative line. Thus, three distinct alleles segregated among ten lines for *inv\_6a/6b* and *inv\_10a/10b*. These cases contribute to the growing catalog of complex, multi-part inversions segregating in *Mimulus*<sup>54,55</sup>.

### Nucleotide diversity

To evaluate nucleotide diversity patterns, we scored genes within inversions distinguishing those closest to the breakpoints from those in the interior. These were compared to genes just outside the inversion as well as the remainder of the genome. Using genome assemblies from four outgroup species, we were able to polarize 54 of the inversions, calling the reference line orientation Ancestral or Derived. As expected from theory<sup>44,45</sup>, nucleotide diversity patterns differ substantially between Ancestral and Derived orientations ( $F_{2,335} = 143.77$ ,  $p < 0.001$ ). The average nucleotide diversity within the ancestral orientation ( $\pi_{AA}$ ) is nearly 4.6 times higher than within the derived orientation ( $\pi_{DD}$ ) and sequence divergence between arrangements ( $\pi_{AD}$ ) is about 60% greater than  $\pi_{AA}$  (Fig. 1b). Importantly,  $\pi_{AA}$  within/near inversions is equivalent to the genomewide average (horizontal in Fig. 1b); the inversion effect is evident only in  $\pi_{DD}$  and  $\pi_{AD}$ . Among the genes within inversions, there is no difference in genetic diversity between those close to the breakpoints and those to the interior. There is a significant attenuation of the differences ( $\pi_{AA} > \pi_{DD}$  and  $\pi_{AD} > \pi_{AA}$ ) as we move to those genes nearest the breakpoints but outside the inversion ( $F_{4,335} = 2.86$ ,  $p < 0.05$ ).

Diversity patterns also depend on the relative frequencies of the two orientations (Fig. 1c) and the size of the inverted region (Fig. 1d). For both frequency ( $F_{2,130} = 4.03$ ,  $p = 0.02$ ) and size ( $F_{2,130} = 3.08$ ,  $p = 0.05$ ), there is significant slope heterogeneity, but the regressions for individual components ( $\pi_{AA}$ ,  $\pi_{DD}$  and  $\pi_{AD}$ ) are only marginally significant or non-significant. Intra-orientation variation appears positively correlated with population frequency (Fig. 1c), but only the  $\pi_{AA}$  regression is significant when considered in isolation. This may just reflect lower power for  $\pi_{DD}$ ; there are fewer estimates for this contrast because it can be calculated only when the derived orientation is carried by at least two lines (28 of 54 cases). Both  $\pi_{DD}$  and  $\pi_{AD}$  decline with inversion size and the  $\pi_{DD}$  reduction is probably stronger than evident in Fig. 1d. The largest inversion, *inv\_24* (Fig. 1a), does not yield a  $\pi_{DD}$  for the present analysis because it is carried by only one of the ten lines. However, a previous survey of this genomic region revealed complete internal homogeneity ( $\pi_{DD} \approx 0$ ) of *inv\_24* over its entire 4.1 mb length<sup>52</sup>.

The equivalence of diversity in the interior of inversions with genes neighboring the breakpoints suggests that divergence is not elevated at breakpoints. However, Fig. 1b–d are based on genes and the breakpoints generally reside in intergenic regions. To evaluate diversity in intergenic regions, we calculated statistics within 2 kb windows across each inversion including 20 kb beyond each breakpoint. Four examples are reported in Fig. 2, the full set as Supplementary Figs. 1–63. These trajectories are routinely incomplete, broken in places where reliable sequence alignments between lines are impossible owing to the very high sequence and insertion/deletion variation that segregates in *M. guttatus*. Despite these difficulties, the overall pattern is clear: Nucleotide diversity fluctuates idiosyncratically across



**Fig. 1 | The physical arrangement of DNA for the largest inversion and the distribution of nucleotide variation across all inversions. a** Inv\_24: Each exon of IM767 from the interval of 1 to 7 mb of Chromosome 6 was mapped to IM664 (y-axis): blue = forward alignments (+), red = reverse alignments (-). The inversion contains two distinct sections of negative alignment separated by a positive alignment. **b** The average (and distribution) of nucleotide diversity ( $\pi$ ) is reported ancestral-ancestral ( $\pi_{AA}$ ), ancestral-derived ( $\pi_{AD}$ ), and derived-derived ( $\pi_{DD}$ ) comparisons for genes within inversion interiors (left), inside the breakpoints (middle), and immediately outside breakpoints (right). The dotted line is the mean  $\pi$  for

genes across the colinear genome. Means were compared by ANOVA and the bands are 95% confidence intervals. Samples sizes for interior genes are  $n = 51, 51, 26$  for AA, AD, and DD, respectively. The corresponding  $n$  are 54, 53, 28 for inside breakpoints and  $n = 54, 53, 27$  for outside breakpoints. The diversity statistics are reported as a function of (c) the frequency of the derived orientation within IM and (d) the physical size of the inversion. In c, no estimate is reported for  $\pi_{DD}$  where  $q = 0.1$ , because interval variation cannot be calculated unless there are at least two lines carrying the orientation. Points have been ‘jittered’ slightly for  $q$  to reveal distinct observations. Source data are provided as a Source Data file.

inversions. There is no consistent tendency for  $\pi_{AD}$  to peak at or near breakpoints.

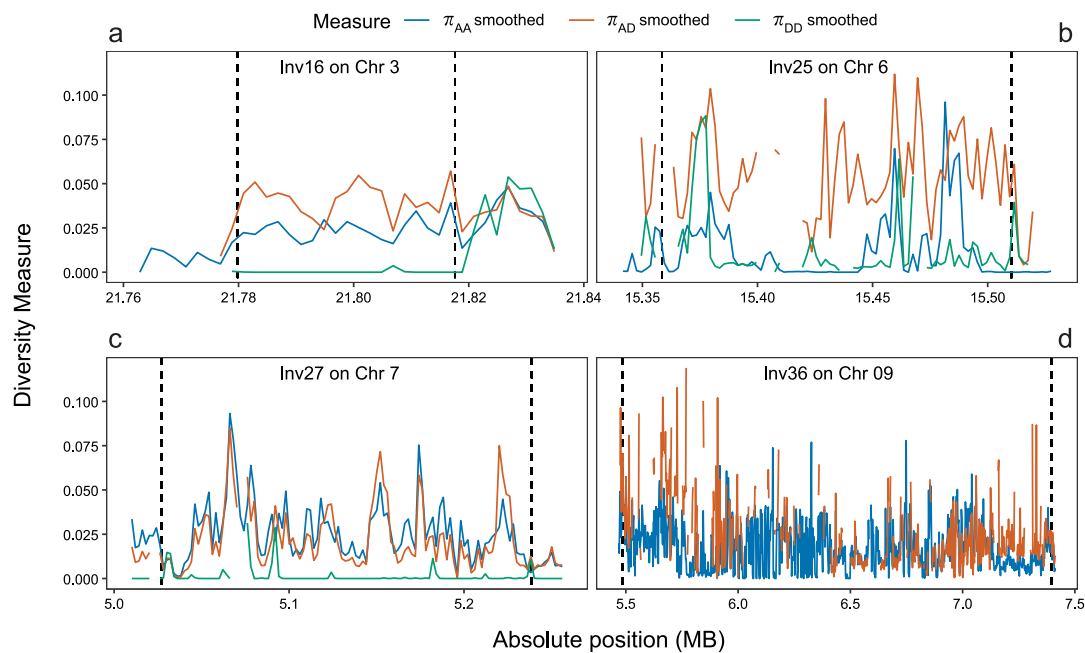
The window-based diversity calculations also indicate the respective effects of mutation and gene flux in generating variation within the derived orientation ( $\pi_{DD}$ ). Both processes will increase  $\pi_{DD}$  but they yield different spatial signals. Mutation will randomly distribute differences across an inversion while flux should generate localized introgression tracts. The latter are evident as highly localized peaks in  $\pi_{DD}$  trajectories, particularly in cases where the derived orientation harbors little variation on average (e.g. *inv\_16* and *inv\_27* in Fig. 2). Mutation will introduce variation in a statistically homogeneous way across the derived orientation. In the short term, this corresponds to a Poisson expectation for SNPs per window. The observed distribution is much ‘clumpier’ than Poisson suggesting that gene flux is the primary source of variants for  $\pi_{DD}$  (Supplementary Figs. 1–63).

### Gene regulation

We find no evidence that the overall genetic variance in gene expression is inflated by inversions. The ten genomes characterized here for inversions were recently used in a multi-parental eQTL mapping experiment<sup>53</sup>. Each of the nine ‘alternative’ genomes was crossed to IM767 and each of these nine  $F_1$  plants was self-fertilized to produce  $F_2$  plants. Each  $F_2$  was measured for gene expression in leaf tissue

( $n = 1588$ ). This experiment produced estimates for the genetic variance in expression for nearly 13,000 genes and for the proportion of that variation at each gene that is generated by the DNA surrounding the gene, i.e. the cis-eQTL. Locating genes inside/outside inversions, we find no difference between genes within inversions versus those in colinear portions of the genome, not for the overall genetic variance nor for the cis component of that variance. There is no difference between genes in the interior of inversions relative to those close to the breakpoints (Supplementary Fig. 64).

Despite the negative result on magnitude, there is clear evidence that inversions ‘partition’ expression variation. This can be shown by reanalyzing the eQTL data in direct relation to our inversion calls of each line for each inversion (see ‘Gene expression calculations’ section of the Methods). In each of the nine  $F_2$  populations (Fig. 3a), we can score each plant at the cis-locus of any expressed genes as 0, 1, or 2, depending on how many copies of the alternative line allele they carry. The regression of expression onto allele count is the ‘average effect’ of the alternative allele<sup>56</sup> in that particular cross. These regressions are reported for one gene across all nine  $F_2$  populations in Fig. 3b. The cis locus at each of the nine alternative genomes could differentially affect expression relative to IM767, and in principle, each may do so in differing ways. The eQTL mapping experiment demonstrated that cis eQTLs are usually allelic series with 3–5 distinct alleles<sup>53</sup>.



**Fig. 2 | Nucleotide diversity by position across four inversions.** The trajectories for  $\pi_{AA}$ ,  $\pi_{AD}$ , and  $\pi_{DD}$  are reported over the genomic regions containing inversions (a) 16, (b) 25, (c) 27, and (d) 36. The vertical broken lines indicate the estimated

breakpoints for each inversion. There is  $\pi_{DD}$  trajectory for *inv\_36* because the derived orientation is carried by only one line.

To assess inversion effects, we partition the full set of estimates (the regression slope from each cross) based on whether each alternative line shares the same inversion orientation as the reference genome (homogenous crosses) or carries the inverted orientation (segregating crosses). If the inversion is inconsequential for expression (our null hypothesis), average effect estimates should be no more similar within cross types than between cross types. Figure 3b illustrates a gene that exhibits an inversion effect. Those lines that share the reference orientation tend to reduce expression while the lines carrying the inversion (and thus differ in orientation from the reference) increase expression. Consequently, the average pairwise difference within cross types (0.236) is much lower than between cross types (0.960). Applying this calculation across all genes within inversions, the average between cross types is significantly greater than within ( $p < 0.0001$ , Fig. 3c left). Significance was established by permuting estimates across the inversion classification of lines within genes. The permutation distributions for each statistic (histograms in Fig. 3d) do not encompass the observed values (arrows in Fig. 3d): cis-eQTL effects are strongly predicted by inversion status ( $p < 0.0001$ ).

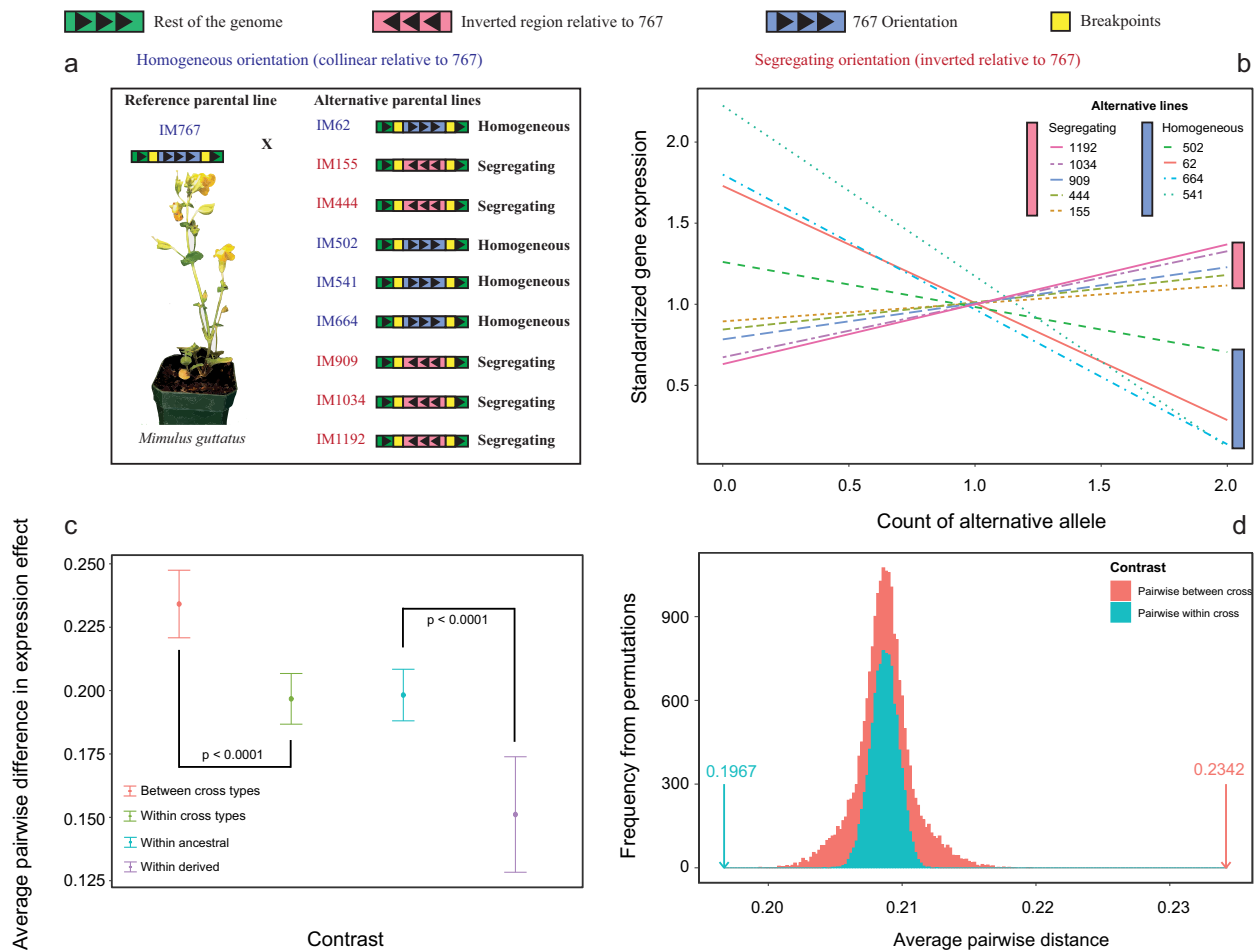
A second notable feature of the distribution of expression variation is revealed by the subset of inversions where both Ancestral and Derived orientations are carried by at least two lines. For these inversions, we can subdivide the intra-cross divergence into within-Ancestral and within-Derived components (Fig. 3c, right). We observe greater similarity of genetic effect estimates within the Derived orientation than within the Ancestral ( $p < 0.0001$  from permutation, Supplementary Fig. 65). This pattern is also evident in the specific case of *MgIM767.06G148700* (Fig. 3b) where the derived orientation alleles not only have a different average effect than ancestral orientation alleles, but the lines are more consistent in their effect. As a final analysis, we ignored the inversion classification and considered the average pairwise difference for genes within the inversion to those throughout the rest of the genome. We find no difference in the magnitude of differences (Supplementary Fig. 64) for genes within inversions versus the remainder of the colinear genome. This confirms the negative result—the genetic variance in expression is

not greater within inversions than across the colinear genome—that we obtained by classifying estimates taken direct from the eQTL mapping experiment.

## Discussion

The major fitness effects hypothesized to emerge from inversions are (a) meiotic aberrations produced by inversion heterozygotes, (b) position effects of sequence rearrangement, and (c) the collection of coadapted alleles into recombination-suppressed haplotypes<sup>44</sup>. We cannot test whether the production of genetically unbalanced gametes (hypothesis a) is important from the present experiment, but the accumulation of evidence in *Mimulus* is not supportive. Eight large inversions, previously mapped within the *M. guttatus* species complex, have been implicated in the full array of ‘inversion phenomena’ including meiotic drive<sup>7,57</sup>, local adaptation and speciation<sup>58–63</sup>, and the persistence of fitness variation within natural populations<sup>40,52</sup>. There is no evidence for underdominance owing to meiotic imbalances. Even for the traits most prone to this effect (e.g. pollen viability), inversion heterozygotes have intermediate phenotypes<sup>52</sup> or the heterozygotes resemble the more fit homozygote<sup>40,64</sup>. Here, we present results for a new collection of inversions, only 2 of 64 known previously. Both sequence evolution and gene expression variation at these loci favor hypothesis (c) over (b); the internal content of inversions appears to be more evolutionarily important than their breakpoints.

Considering gene sequence evolution, Berdan et al.<sup>44</sup> recently reviewed theoretical predictions for the evolution of  $\pi_{AA}$ ,  $\pi_{AD}$ , and  $\pi_{DD}$  over the genomic span of inversions. When a sequence is initially inverted by mutation, and shortly thereafter, we statistically expect  $\pi_{DD} = 0$  and  $\pi_{AA} = \pi_{AD}$ , with the latter diversity statistics equaling the genome average. However, the way that these statistics change as the inversion ages should be different depending on whether selection acts mainly on the breakpoints or on loci to the interior. If the selection maintaining the inversion is focused on the breakpoints, a suspension bridge pattern is predicted for  $\pi_{AD}$ , greatest at the breakpoints with a gradual decline to its lowest value in the middle. In contrast, if genes to the interior are the actual targets of selection,  $\pi_{AD}$  peaks should not be



**Fig. 3 | The effect of inversions on gene expression variation.** **a** The estimation of inversion effects on gene expression is illustrated using data from an exocyst complex protein gene (MgIM767.06G148700). This gene resides within *inv\_25* which is carried by five alternative lines (155, 444, 909, 1034, and 1192) which makes five of the F<sub>2</sub> populations segregating. **b** The linear regression of gene expression level onto alternative allele count is reported for each of the nine F<sub>2</sub> populations. Expression levels have been standardized to a mean value of 1 within each population. **c** The average pairwise distance is reported for four categories. The *p*-values

refer to comparisons between versus within cross type (left) and within-Ancestral versus within-Derived (right). 932 genes assayed for between/within cross types, 872 for within-Ancestral, and 168 for within-Derived. The 95% confidence bands are from averages across genes, while the test *p*-values were obtained by permutation. **d** The results of permutation for between cross-type (orange) and within cross-type (blue) reveal the observed pairwise distances (indicated by arrows) to be well outside the null distribution. Source data are provided as a Source Data file.

specific to the breakpoints. For the *Mimulus* inversions, the average  $\pi_{AD}$  for genes close to the breakpoints is indistinguishable from the average of the interior (Fig. 1b) and peaks of divergence occur idiosyncratically across the sequence space from one breakpoint to the other (Fig. 2). The haphazard location of  $\pi_{AD}$  peaks suggests that selection is acting more on the variants captured by inversions, which will naturally vary in location, than by breakpoint effects.

The variation internal to the ancestral orientation,  $\pi_{AA}$ , is equal (on average) to the overall nucleotide diversity ( $\pi$ ) in collinear parts of the genome (dotted line in Fig. 1b). In contrast, variation within the derived orientation ( $\pi_{DD}$ ) is much lower, which suggests that most inversions segregating in the IM population are evolutionarily young (Fig. 1b). Despite their apparent youth and the effect of gene flux, most inversions exhibit substantially elevated sequence divergence between orientations relative to ancestral ( $\pi_{AD} > \pi_{AA}$ ). This  $\pi_{AD} > \pi_{AA}$  tendency should emerge through time, because we expect  $\pi_{AD} \approx \pi_{AA}$  when an inversion is first produced by mutation. The spatial scale of sampling is a critical consideration here. The ten genomes of this study are from unrelated plants within one natural population (IM). Our estimates for  $\pi_{AA}$  and genome-wide  $\pi$  are within this population. *M. guttatus* exhibits substantial variation among populations with  $F_{st}$  ranging from 0.1 to

0.5 depending on geographical distance<sup>62</sup>. The mean  $\pi_{AD}$  (0.029) for inversions is a moderate value for nucleotide diversity between gene sequences from different populations. It is possible that inversion orientations are sampled from this larger pool of variation with derived orientation introduced to IM through gene flow. In this model, where geographical distance in selection is essential to inversion persistence<sup>10,54</sup>,  $\pi_{AD} > \pi_{AA}$  should emerge early in the lifetime of polymorphisms.

Nucleotide diversity within inversion orientations is correlated with their physical size and population frequency (Fig. 1c, d). The slight tendency towards lower variation within the less common orientation could reflect a population size effect. An inversion effectively splits the overall population of alleles into two sub-populations based on orientation. These subpopulations are genetically connected (gene flux is analogous to gene flow) but drift might remove variation more rapidly from the smaller subpopulation, i.e. the less frequent orientation. Alternatively, the decline of  $\pi_{AA}$  with derived orientation frequency (Fig. 1c) might result because the latter quantity is correlated some other critical variable such as inversion age.

The decline of both  $\pi_{AD}$  and  $\pi_{DD}$  with inversion size (Fig. 1d) is notable because it suggests a relationship between inversion size and

age. Previous studies<sup>7,52</sup> on two of the largest inversions (*inv\_48* which is part of the meiotic drive locus on chromosome 11 and *inv\_24* on chromosome 6) had concluded recent origin based on very low interval variation combined with limited geographic spread. Both inversions are maintained at intermediate frequencies in IM owing to conflicting selection pressures. A recent theoretical study<sup>65</sup> suggests that small rearrangements are likely to be fixed by selection, while intermediate-to-large inversions are more likely to be maintained as balanced polymorphisms through associative overdominance. Associative overdominance is entirely plausible for both *inv\_24* and *inv\_48*, but an incisive test of the model based on a survey across inversions will require information on the rate that inversions of different sizes are produced by mutation.

The breakpoint hypothesis for inversions makes a clear prediction for gene expression. If the physical reversal of the DNA sequence affects fitness by disrupting gene regulation, we expect that breakpoints should emerge as cis-eQTLs, i.e. loci that affect the expression of proximal genes<sup>66</sup>. Breakpoints clearly are cis-eQTLs in *Mimulus*, but this is not compelling evidence that they are causal to variation. In the IM population, nearly all genes exhibit cis-regulatory variation, often-times as an allelic series with three to five distinct alleles<sup>33</sup>. A random partitioning of ten genomes into two groups (simulated inversion orientations) will usually yield a difference in mean expression between partitions. When we ask the more informative question of whether the genetic variance is elevated in breakpoint-proximal genes, the result is negative (Supplementary Fig. 64). Breakpoint genes are not more variable in expression than those to the interior of inversions or throughout the rest of the genome.

If an inversion is favored because the derived orientation captures pre-existing variation in expression, there will be no immediate effect on the overall expression variance in the population. Incremental shifts in the genetic variance in expression are likely to occur subsequently as orientations change in frequency, but these shifts can have positive or negative effects on the overall variance. The capture model does predict a partitioning of expression variation between inversion types, which is strongly supported by the *Mimulus* data (Fig. 3c, d). Moreover, the partitioning is predicted to be asymmetric. The derived orientation should contain less variation than the ancestral orientation, particularly when most genes have multiple cis-regulatory alleles segregating in the population and only a single allele is captured by the inversion. The predicted asymmetry in internal expression variability (ancestral versus derived) is confirmed by these inversions (Fig. 3c, right).

In summary, both aspects of the data—sequence and expression variation—support the captured alleles hypothesis for inversions. Peaks of divergence between orientations are idiosyncratically distributed across inverted regions for both features. Of course, the selection regime is likely heterogeneous across inversions and breakpoint effects may be critical for some loci. More generally, the actual selection regime acting on alleles to the interior of inversions are not identified by these patterns. Polymorphism can be maintained locally, because each orientation is subject to conflicting costs and benefits within the IM population. The meiotic drive locus (associated with *inv\_48*) is an example where conflicting gametic and zygotic selection make the polymorphism locally stable<sup>7,40</sup>. Alternatively, varying selection across the range of a species could maintain polymorphism because the inversion captures alleles co-adapted for local environmental conditions<sup>10</sup>. A large inversion on chromosome 8 is associated with inter-population difference in life history, specifically the annual/perennial phenotypic polymorphism of *M. guttatus*<sup>58</sup>. The ‘perennial’ orientation at this locus does occur in annual populations<sup>67</sup>, but that may be due to recurrent immigration rather than local advantage. The selection pressures acting on inversion polymorphisms are most directly estimated by measuring survival and reproductive success of alternative genotypes under natural conditions. Field studies assessing inversion effects on survival and reproductive success have historically

focused on individual inversions. An important step for future studies will be to develop and implement methods to dissect fitness effects owing to multiple loci given that inversions are likely to segregate at many positions across the genome.

## Methods

### Genome assemblies and inversion calling

The 10 genomes are derived from a large panel of lines, each created by single seed descent from an independent maternal family. These families were sampled from the Iron Mountain population<sup>50</sup> of *M. guttatus* (syn *Erythranthe guttata*) in central Oregon, U.S.A. (44.402217 N, 122.153317 W). We assembled genomes for IM155, IM444, IM502, IM541, IM664, IM909, IM1034, and IM1192 by first extracting high molecular weight DNA from each line. These lines had been confirmed to be individually homozygous and unrelated to each other through previous Illumina sequencing<sup>68</sup>. The assemblies are haploid (given line homozygosity). The University of Georgia genomics facility made Sequel II CLR libraries and sequenced each using PacBio according to the manufacturer’s instructions. Each line was sequenced to ~250–300x coverage. We assembled sequence contigs using canu 2.1.1<sup>69</sup>, which yielded contigs 4–15 mb in size. Next, we used genetic mapping data<sup>53</sup> to order and orient the contigs into pseudo-chromosomes. Importantly, all the inversions identified below are within contigs and thus inference does not depend on the ordering of contigs from mapping data. The other two genome assemblies (IM767 and IM62) were produced by the Joint Genome Institute<sup>70</sup> from DNA extracted with the same protocol as the other lines. The IM767 and IM62 assemblies are used with permission.

After testing many bioinformatic tools, we identified inversions from the synthesis of results from three distinct analyses. Sequence variation is remarkably high in *Mimulus*, and the abundance of point mutations and insertion/deletions makes alignment very difficult outside of genic regions. We first applied AnchorWave<sup>71</sup> to detect inversions by aligning each alternative genome, chromosome by chromosome, to IM767. This program, which was designed for highly variable organisms like corn (or *Mimulus*), establishes ‘anchors’ as colinear blocks of sequence between aligned chromosomes. Coding sequences are the primary building blocks of these anchors. Anchors can be positive or negative alignments between assemblies; inversions are called from negative alignments. We implemented AnchorWave, chromosome by chromosome, using custom Python scripts. We used default parameter values except for “l” which we reduced from 2 to 1. This code and all other python programs described below, are publicly accessible on github (<https://zenodo.org/records/14013030>) with documentation explaining their use.

AnchorWave detected 178 inversions across the nine alternative lines relative to IM767. Oftentimes, the same inversion was detected in multiple lines, and after distilling these cases, we identify 64 non-redundant inversions. To refine the location of inversion breakpoints (which emerge as inter-anchor intervals from AnchorWave), we mapped exon sequence from all genes annotated in the IM767 assembly (our reference) to each alternative genome using Minimap2<sup>72</sup> with settings updated to account for the high variation present even in coding sequences. Treating exons as genetic markers, we confirmed the reversal of both order and mapping orientation (positive or negative) of markers within the Anchorwave predicted inversions. We could resolve breakpoint regions to smaller intervals for many inversions using the exon mappings. For the third analysis, we applied ‘pipeline 3’ from Zhou et al.<sup>73</sup> mapping each chromosome of each alternative genome to IM767 using Minimap2 followed by inversion calling using Syri<sup>74</sup>. This method, which is not explicitly focused on coding sequences, identified 715 total inversions. This reduces to 400 non-redundant inversions across lines (reported as Supplementary Data 2). We find that pipeline 3 usually ‘captures’ Anchorwave inversions (those in the set of 64 are contained within the larger collection).

Syri identified many additional inversions, but these are typically small and most do not contain even a single complete gene. While many of the pipeline 3 predicted variants may be real, we do not consider them further in this study. Our focus here is on sequence divergence within, and expression of, genes (Figs. 1–3). The pipeline 3 calls corresponding to AnchorWave inversions were used to refine breakpoint estimation. Since Syri yields a point estimate instead of a range for breakpoints, we accepted the Syri estimate whenever it fell within the interval predicted by the two previous methods. When there was no estimate from Syri, we chose the center of the range from exon mapping as the putative inversion breakpoint.

For a subset of loci, we used the long read data to confirm that putative inversions were not due to contig assembly error. Using Minimap2, we mapped the original PacBio long reads from several of the lines, first to the IM767 reference genome and then to the alternate lines carrying the alternative orientation. With IM767 as the mapping reference, we confirmed that the long reads from IM767 DNA aligned in fully colinear way to the IM767 assembly across the predicted inversion breakpoints. We confirmed that long reads from the alternative line did not, either because they did not map at all or because only one end the long read mapped (the remainder ‘soft clipped’ and not aligned). For these partially mapped sequences, we split the original long read (usually 10–50 kb in length) into two fasta files, breaking the sequence beyond the mapped portion. We then remapped the two ends as distinct sequences. Routinely, we were able get the split reads to map, each to one end of the inversion with the predicted change in alignment direction. We repeated this procedure with the long-read mappings to the alternative genome assembly. Here, we obtained the same outcome except that splitting was necessary on the long reads from the IM767 DNA.

We polarized inversion orientations by inspecting inverted regions in genome assemblies from more distantly related *Mimulus* genomes. We used AnchorWave and exon mapping to align *M. tilingii*, *M. nasutus*, *M. cardinalis* and *M. lewisii* to IM767. *M. tilingii* and *M. nasutus* are close relatives of IM *M. guttatus*, each about twice as sequence divergent from IM as IM sequences are to each other. *M. cardinalis* and *M. lewisii* are more distantly related<sup>75</sup>. For most inverted regions, we could identify the homologous sequence region in multiple outgroups and the ancestral orientation was unambiguous: All outgroup species shared the same orientation and it matched either IM767 or the alternative line. In some cases, the orientation of *M. tilingii* or *M. nasutus* differed from the other groups, suggesting that (a) the inversion segregated within the common ancestor of IM and the outgroup species, or that (b) the inversion was produced by mutation within one of the descendant lineages but has since introgressed into the other lineage. In these cases, we assigned orientation based on the most distantly related, aligned outgroup (*M. cardinalis* and/or *M. lewisii*). Ten of the 64 inversions could not be polarized because the close relatives differed from each other in orientation and extensive structural changes made the alignments to *M. cardinalis* and *M. lewisii* unintelligible.

### Nucleotide diversity calculations

For nucleotide diversity estimation (and testing), we distill observations from each inversion into a single collection of values ( $\pi_{AA}$ ,  $\pi_{AD}$ , and  $\pi_{DD}$ ) and then averaged over genes within the inversion. We then average estimates across inversions, which guarantees an equal contribution of each inversion to the overall results (distinct inversions are the units of replication). We used Mummer 3.0<sup>76</sup> and SVMU<sup>77</sup> to call SNP variants across positions that were reliably aligned between IM767 and the other genome assemblies. Using custom python scripts, we determined  $\pi$  between each pair of lines for each gene (from start to stop codon) across the genome. Genes were classified as (a) in the interior of inversions, (b) within the inversion but nearest the breakpoints, (c) nearest to breakpoints but outside an inversion, or (d)

residing in the remainder of the (colinear) genome. For the 54 of 64 inversion that could be polarized as ancestral or derived, we assigned the intra-orientation nucleotide diversity estimates as  $\pi_{AA}$  or  $\pi_{DD}$ . The inter-orientation diversity,  $\pi_{AD}$ , could be assigned to all inversion related genes.

We tested for differences in diversity statistics using a two factor ANOVA with location-relative-to-inversion (interior, inside breakpoint, or outside breakpoint) as one factor and contrast ( $\pi_{AA}$ ,  $\pi_{AD}$ , or  $\pi_{DD}$ ) as the second factor (Fig. 1b). The genomic background estimate for  $\pi$  is reported as a constant (the confidence interval is miniscule because it is based on thousands of genes). We used a general linear model with contrast as a categorical factor and either derived frequency (Fig. 1c) or Log-transformed inversion size (Fig. 1d) as a continuous predictor. This model included an interaction to test for slope heterogeneity. Given significant heterogeneity for both, we performed separate linear regressions to estimate the distinct responses of  $\pi_{AA}$ ,  $\pi_{AD}$ , and  $\pi_{DD}$  to each predictor. We also calculated  $\pi_{AA}$ ,  $\pi_{AD}$ , and  $\pi_{DD}$  in 2 kb windows across each inversion without consideration of gene boundaries. These traces were initiated 20 kb upstream of the left breakpoint of each inversion and terminated 20 kb downstream of the right flank (Fig. 2, Supplementary Figs. 1–63). These calculations could only be performed where homologous sites could be confidently aligned between genomes. Consequently,  $\pi$  trajectories through intergenic regions routinely have gaps.

### Gene expression calculations

For the first analysis, we directly extracted estimates for the genetic variance in expression per gene, and the cis-component of that variance, from Supplementary Table 4 of the eQTL paper<sup>53</sup>. This analysis was limited to the 12,987 genes that could be ascertained in all ten genomes and that passed other filters. We reanalyzed the data here to include more genes, all with estimates from at least 5 of the 9 crosses. For determining the average level of expression of each genotype in each F<sub>2</sub> panel (homozygous for the IM767 allele, heterozygote, or homozygous for the alternative line allele), we started with the raw RNAseq read counts for each gene<sup>53</sup>. We then normalized for the total number of RNAseq reads per plant, which gives expression levels at each gene in Counts per Million (CPM). We excluded genes with an average CPM < 0.5. Next, we standardized each gene by dividing individual plant expression levels by the CPM of that gene so that the average is 1.0 across all F<sub>2</sub> plants. We applied linear regression to each gene within each cross with standardized expression as the dependent variable and the alternative allele count (0, 1, or 2) as the independent variable. The regression slope is the average effect<sup>56</sup> of the alternative allele in that cross for that gene. Given our standardization, an average effect of 0.1 means that heterozygotes have  $\approx$  10% higher expression than the IM767 homozygote. Applied to all ascertained genes, we find that the mean average effect is close to zero (a mixture of positive and negative values across genes and crosses). The mean absolute value for the average effect (horizontal dotted line in Fig. 2B) was 0.182 (S.E.M. = 0.00056).

The genetic variance in expression,  $V_g$ , and the cis-component of that variance,  $V_{g(cis)}$ , were treated as the dependent variable in an ANOVA with gene location (outside flanking, inside flanking, inversion interior, or background genome) as the predictor (Supplementary Fig. 64). We performed a similar test on the reanalysis of the data on the effect estimates (regression slopes) for each gene within each cross. For this, we calculated the average pairwise difference between estimates for each gene and tested whether this was affected by gene location (Supplementary Fig. 64). Next, we considered only genes within inversions and partitioned pairwise differences statistics within and between cross types (Fig. 3c). For the subset of inversions with more than one line carrying both the Ancestral and Derived orientations, we partitioned the within-cross type statistic into Ancestral and Derived components. To test for differences (within versus between

and Ancestral-within versus Derived-within) we permuted inversion orientation versus genetic effect estimates within genes. We then compared the observed differences (Fig. 3c) to the distribution of differences for each contrast across 10000 permuted datasets (Fig. 3d, Supplementary Fig. 65).

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Eight of the ten genome assemblies (IM155, IM444, IM502, IM541, IM664, IM909, IM1034, and IM1192) used in this study have been deposited in both NCBI under accession [PRJNA1183383](https://www.ncbi.nlm.nih.gov/PRJNA1183383) and Mimubase [<http://mimubase.org/FTP/Genomes/>]. The other two assemblies from Iron Mountain, IM767 and IM62, were created by the Joint Genome Institute and are used here with permission. Source data are provided with this paper.

### Code availability

All python programs written for analysis of these data are available through our Github page [<https://doi.org/10.5281/zenodo.14013030>].

### References

- Dobzhansky, T. & Sturtevant, A. H. Inversions in the Chromosomes of *Drosophila Pseudoobscura*. *Genetics* **23**, 28–64 (1938).
- Ford, E. B. *Ecological genetics*. 3rd edn, (Chapman and Hall, 1971).
- Schwander, T., Libbrecht, R. & Keller, L. Supergenes and Complex Phenotypes. *Curr. Biol.* **24**, R288–R294 (2014).
- Joron, M. et al. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**, 203–206 (2011).
- Hammer, M. F., Schimenti, J. & Silver, L. M. Evolution of mouse chromosome 17 and the origin of inversions associated with haplotypes. *Proc. Natl Acad. Sci. USA* **86**, 3261–3265 (1989).
- Zanders, S. E. et al. Genome rearrangements and pervasive meiotic drive cause hybrid infertility in fission yeast. *eLife* **3**, e02630 (2014).
- Fishman, L. & Saunders, A. Centromere-Associated Female Meiotic Drive Entails Male Fitness Costs in Monkeyflowers. *Science* **322**, 1559–1562 (2008).
- Larracuente, A. M. & Presgraves, D. C. The Selfish Segregation Distorter Gene Complex of *Drosophila melanogaster*. *Genetics* **192**, 33–53 (2012).
- Pardo-Manuel de Villena, F. & Sapienza, C. Female meiosis drives karyotypic evolution in mammals. *Genetics* **159**, 1179–1189 (2001).
- Kirkpatrick, M. & Barton, N. Chromosome Inversions, Local Adaptation and Speciation. *Genetics* **173**, 419–434 (2006).
- Barney, B. T., Munkholm, C., Walt, D. R. & Palumbi, S. R. Highly localized divergence within supergenes in Atlantic cod (*Gadus morhua*) within the Gulf of Maine. *BMC Genomics* **18**, 271 (2017).
- Jones, F. C. et al. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
- Love, R. R. et al. Chromosomal inversions and ecotypic differentiation in *Anopheles gambiae*: the perspective from whole-genome sequencing. *Mol. Ecol.* **25**, 5889–5906 (2016).
- Roesti, M., Kueng, B., Moser, D. & Berner, D. The genomics of ecological vicariance in threespine stickleback fish. *Nat. Commun.* **6**, 8767 (2015).
- Wellenreuther, M., Rosenquist, H., Jaksons, P. & Larson, K. W. Local adaptation along an environmental cline in a species with an inversion polymorphism. *J. Evol. Biol.* **30**, 1068–1077 (2017).
- Ayala, D., Ullastres, A. & González, J. Adaptation through chromosomal inversions in *Anopheles*. *Front. Genet.* **5**, <https://doi.org/10.3389/fgene.2014.00129> (2014).
- Bush, G. L., Case, S. M., Wilson, A. C. & Patton, J. L. Rapid speciation and chromosomal evolution in mammals. *Proc. Natl. Acad. Sci. USA* **74**, 3942–3946 (1977).
- Rieseberg, L. H. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**, 351–358 (2001).
- Noor, M. A. F., Grams, K. L., Bertucci, L. A. & Reiland, J. Chromosomal inversions and the reproductive isolation of species. *Proc. Natl Acad. Sci. USA* **98**, 12084–12088 (2001).
- McGaugh, S. E. & Noor, M. A. F. Genomic impacts of chromosomal inversions in parapatric *Drosophila* species. *Philos. Trans. R. Soc. B: Biol. Sci.* **367**, 422–429 (2012).
- Dagilis, A. J. & Kirkpatrick, M. Prezygotic isolation, mating preferences, and the evolution of chromosomal inversions. *Evolution* **70**, 1465–1472 (2016).
- Feulner, P. G. D. & De-Kayne, R. Genome evolution, structural rearrangements and speciation. *J. Evol. Biol.* **30**, 1488–1490 (2017).
- Ragland, G. J. et al. A test of genomic modularity among life-history adaptations promoting speciation with gene flow. *Mol. Ecol.* **26**, 3926–3942 (2017).
- Lohse, K., Clarke, M., Ritchie, M. G. & Etges, W. J. Genome-wide tests for introgression between cactophilic *Drosophila* implicate a role of inversions during speciation. *Evolution; Int. J. Org. Evolution* **69**, 1178–1190 (2015).
- Kozak, G. M. et al. A combination of sexual and ecological divergence contributes to rearrangement spread during initial stages of speciation. *Mol. Ecol.* **26**, 2331–2347 (2017).
- Zouros, E. On the Role Of Chromosomal Inversions In Speciation. *Evolution* **36**, 414–416 (1982).
- Kirkpatrick, M. & Barrett, B. Chromosome inversions, adaptive cassettes and the evolution of species' ranges. *Mol. Ecol.* **24**, 2046–2055 (2015).
- Kirubakaran, T. G. et al. Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Mol. Ecol.* **25**, 2130–2143 (2016).
- Tuttle, Elaina M. et al. Divergence and Functional Degradation of a Sex Chromosome-like Supergene. *Curr. Biol.* **26**, 344–350 (2016).
- Kupper, C. et al. A supergene determines highly divergent male reproductive morphs in the ruff. *Nat. Genet.* **48**, 79–83 (2016).
- Charlesworth, D., Charlesworth, B. & Marais, G. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95**, 118–128 (2005).
- Johnson, N. A. & Lachance, J. The genetics of sex chromosomes: evolution and implications for hybrid incompatibility. *Ann. N. Y. Acad. Sci.* **1256**, E1–E22 (2012).
- Wright, A. E., Dean, R., Zimmer, F. & Mank, J. E. How to make a sex chromosome. *Nat. Commun.* **7**, 12087 (2016).
- Caputo, B. et al. The last bastion? X chromosome genotyping of *Anopheles gambiae* species pair males from a hybrid zone reveals complex recombination within the major candidate 'genomic island of speciation'. *Mol. Ecol.* **25**, 5719–5731 (2016).
- Faria, R., Johannesson, K., Butlin, R. K. & Westram, A. M. Evolving Inversions. *Trends Ecol. Evol.* **34**, 239–248 (2019).
- Charlesworth, D. & Charlesworth, B. Mimicry: The Hunting of the Supergene. *Curr. Biol.* **21**, R846–R848 (2011).
- Lamichhaney, S. et al. Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat. Genet.* **48**, 84–88 (2016).
- Wang, J. et al. A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**, 664–668 (2013).
- Dyer, K. A., Charlesworth, B. & Jaenike, J. Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. *Proc. Natl Acad. Sci. USA* **104**, 1587–1592 (2007).
- Fishman, L. & Kelly, J. K. Centromere-associated meiotic drive and female fitness variation in *Mimulus*. *Evolution* **69**, 1208–1218 (2015).



41. Krimbas, C. B. & Powell, J. R. *Drosophila Inversion Polymorphism*. (Taylor & Francis, 1992).
42. Navarro, A., Betrán, E., Barbadilla, A. & Ruiz, A. Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. *Genetics* **146**, 695–709 (1997).
43. Korunes, K. L. & Noor, M. A. F. Pervasive gene conversion in chromosomal inversion heterozygotes. *Mol. Ecol.* **28**, 1302–1315 (2019).
44. Berdan, E. L. et al. How chromosomal inversions reorient the evolutionary process. *J. Evol. Biol.* **36**, 1761–1782 (2023).
45. Guerrero, R. F., Rousset, F. & Kirkpatrick, M. Coalescent patterns for chromosomal inversions in divergent populations. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 430–438 (2012).
46. Wallace, A. G., Detweiler, D. & Schaeffer, S. W. Molecular population genetics of inversion breakpoint regions in *Drosophila pseudoobscura*. *G3 (Bethesda)* **3**, 1151–1163 (2013).
47. Kelly, J. K. & Wade, M. J. Molecular evolution near a two-locus balanced polymorphism. *J. Theor. Biol.* **204**, 83–101 (2000).
48. Kreitman, M. Molecular population genetics. *Oxf. Surv. Evol. Biol.* **4**, 38–60 (1987).
49. Corbett-Detig, R. B. & Hartl, D. L. Population Genomics of Inversion Polymorphisms in *Drosophila melanogaster*. *PLoS Genet.* **8**, e1003056 (2012).
50. Willis, J. H. Partial self fertilization and inbreeding depression in two populations of *Mimulus guttatus*. *Heredity* **71**, 145–154 (1993).
51. Kelly, J. K. The genomic scale of fluctuating selection in a natural plant population. *Evol. Lett.* **6**, 506–521 (2022).
52. Lee, Y. W., Fishman, L., Kelly, J. K. & Willis, J. H. A Segregating Inversion Generates Fitness Variation in Yellow Monkeyflower (*Mimulus guttatus*). *Genetics* **202**, 1473–1484 (2016).
53. Veltsos, P. & Kelly, J. K. The quantitative genetics of gene expression in *Mimulus guttatus*. *PLoS Genet.* **20**, e1011072 (2024).
54. Coughlan, J. M. & Willis, J. H. Dissecting the role of a large chromosomal inversion in life history divergence throughout the *Mimulus guttatus* species complex. *Mol. Ecol.* **28**, 1343–1357 (2019).
55. Leslie, M. K., Lauren, E. S., Sunil Kumar Kenchanmane, R., David, B. L. & Chad, E. N. The role of breakpoint mutations, supergene effects, and ancient nested rearrangements in the evolution of adaptive chromosome inversions in the yellow monkey flower, *Mimulus guttatus*. *bioRxiv*, 2023.2012.2006.570460, <https://doi.org/10.1101/2023.12.06.570460> (2023).
56. Lynch, M. & Walsh, B. *Genetics and analysis of quantitative characters*. (Sinauer associates, 1998).
57. Fishman, L., Kelly, A. J., Morgan, E. & Willis, J. H. A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. *Genetics* **159**, 1701–1716 (2001).
58. Lowry, D. B. & Willis, J. H. A Widespread Chromosomal Inversion Polymorphism Contributes to a Major Life-History Transition, Local Adaptation, and Reproductive Isolation. *PLoS Biol.* **8**, e1000500 (2010).
59. Hall, M. C. & Willis, J. H. Divergent selection on flowering time contributes to local adaptation in *Mimulus guttatus* populations. *Evolution* **60**, 2466–2477 (2006).
60. Lowry, D. B., Rockwood, R. C. & Willis, J. H. Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*. *Evolution* **62**, 2196–2214 (2008).
61. Monnahan, P. J. & Kelly, J. K. The Genomic Architecture of Flowering Time Varies Across Space and Time in *Mimulus guttatus*. *Genetics*, <https://doi.org/10.1534/genetics.117.201483> (2017).
62. Puzey, J. R., Willis, J. H. & Kelly, J. K. Population structure and local selection yield high genomic variation in *Mimulus guttatus*. *Mol. Ecol.* **26**, 519–535 (2017).
63. Twyford, A. D. & Friedman, J. Adaptive divergence in the monkey flower *Mimulus guttatus* is maintained by a chromosomal inversion. *Evolution* **69**, 1476–1486 (2015).
64. Scoville, A., Lee, Y. W., Willis, J. H. & Kelly, J. K. Contribution of chromosomal polymorphisms to the G-matrix of *Mimulus guttatus*. *N. Phytologist* **183**, 803–815 (2009).
65. Connallon, T. & Olito, C. Natural selection and the distribution of chromosomal inversion lengths. *Mol. Ecol.* **31**, 3627–3641 (2022).
66. Lavington, E. & Kern, A. D. The Effect of Common Inversion Polymorphisms In(2L)t and In(3R)Mo on Patterns of Transcriptional Variation in *Drosophila melanogaster*. *G3 (Bethesda)* **7**, 3659–3668 (2017).
67. Colicchio, J. et al. Individualized mating system estimation using genomic data. *Mol. Ecol. Resour.* **20**, 333–347 (2020).
68. Troth, A., Puzey, J. R., Kim, R. S., Willis, J. H. & Kelly, J. K. Selective trade-offs maintain alleles underpinning complex trait variation in plants. *Science* **361**, 475–478 (2018).
69. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**, 722–736 (2017).
70. Nordberg, H. et al. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res* **42**, D26–31 (2014).
71. Song, B. et al. AnchorWave: Sensitive alignment of genomes with high sequence diversity, extensive structural polymorphism, and whole-genome duplication. *Proc. Natl Acad. Sci.* **119**, e2113075119 (2022).
72. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
73. Zhou, Y. et al. Pan-genome inversion index reveals evolutionary insights into the subpopulation structure of Asian rice. *Nat. Commun.* **14**, 1567 (2023).
74. Goel, M., Sun, H., Jiao, W.-B. & Schneeberger, K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* **20**, 277 (2019).
75. Nelson, T. C. et al. Ancient and recent introgression shape the evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation (*Mimulus* section *Erythranthe*). *PLoS Genet.* **17**, e1009095 (2021).
76. Kurtz, S. et al. Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
77. Chakraborty, M., Emerson, J. J., Macdonald, S. J. & Long, A. D. Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. *Nat. Commun.* **10**, 4872 (2019).

## Acknowledgements

Sequencing was done by the KU Genome Sequencing Core which is supported by the National Institute of General Medical Sciences (NIGMS/NIH) under award number P30GM145499. We thank the Department of Energy Joint Genome Institute and collaborators Lila Fishman and John Willis for pre-publication access and use of the IM62 and IM767 genomes from the ongoing *Mimulus* pan-genome project. The work (proposal: <https://doi.org/10.46936/10.25585/60001364>) conducted by the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. J.K.K. received support from NSF grants MCB-1940785 and 2421689 (FAIN).

## Author contributions

P.V. and J.K.K. conceived the experiments. P.V. performed the molecular laboratory work. P.V., L.J.M., and J.K.K. wrote and implemented computer programs to analyze the data. L.J.M. made the figures and took the photo in Fig. 3a. J.K.K. wrote the paper with editorial contributions from P.V. and L.J.M.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-54534-1>.

**Correspondence** and requests for materials should be addressed to John K. Kelly.

**Peer review information** *Nature Communications* thanks Jianquan Liu, Claire Mérot and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024