PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Finseth F, Brown K, Demaree A, Fishman L. 2022 Supergene potential of a selfish centromere. *Phil. Trans. R. Soc. B* **377**: 20210208.
https://doi.org/10.1098/rstb.2021.0208

Received: 6 January 2022
Accepted: 5 April 2022

One contribution of 15 to a theme issue 'Genomic architecture of supergenes: causes and evolutionary consequences'.

Subject Areas:
evolution, genomics

Keywords:
female meiotic drive, gene drive, supergene, monkeyflower, *Mimulus*

Authors for correspondence:
Findley Finseth
e-mail: ffinseth@kecksci.claremont.edu
Lila Fishman
e-mail: lila.fishman@mso.umt.edu

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5980562.

THE ROYAL SOCIETY PUBLISHING

Supergene potential of a selfish centromere

Findley Finseth[1], Keely Brown[2], Andrew Demaree[3] and Lila Fishman[3]

[1]W.M. Keck Science Department, Claremont McKenna, Scripps, and Pitzer Colleges, Claremont, CA 91711, USA
[2]Department of Botany and Plant Sciences, University of California Riverside, Riverside, CA 92521, USA
[3]Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA

FF, 0000-0002-5040-9627; KB, 0000-0002-5371-5830; LF, 0000-0002-7297-9049

Selfishly evolving centromeres bias their transmission by exploiting the asymmetry of female meiosis and preferentially segregating to the egg. Such female meiotic drive systems have the potential to be supergenes, with multiple linked loci contributing to drive costs or enhancement. Here, we explore the supergene potential of a selfish centromere (*D*) in *Mimulus guttatus*, which was discovered in the Iron Mountain (IM) Oregon population. In the nearby Cone Peak population, *D* is still a large, non-recombining and costly haplotype that recently swept, but shorter haplotypes and mutational variation suggest a distinct population history. We detected *D* in five additional populations spanning more than 200 km; together, these findings suggest that selfish centromere dynamics are widespread in *M. guttatus*. Transcriptome comparisons reveal elevated differences in expression between driving and non-driving haplotypes within, but not outside, the drive region, suggesting large-scale *cis* effects of *D*'s spread on gene expression. We use the expression data to refine linked candidates that may interact with drive, including Nuclear Autoantigenic Sperm Protein ($NASP^{SIM3}$), which chaperones the centromere-defining histone CenH3 known to modify *Mimulus* drive. Together, our results show that selfishly evolving centromeres may exhibit supergene behaviour and lay the foundation for future genetic dissection of drive and its costs.

This article is part of the theme issue 'Genomic architecture of supergenes: causes and evolutionary consequences'.

1. Introduction

Supergenes are clusters of multiple functional genes that segregate as a simple genetic polymorphism [1–3]. Drastically reduced recombination in supergene heterozygotes, often due to differences in chromosomal structure (e.g. inversions), generates tight linkage among jointly advantageous or co-adapted variants. Supergenes are a common genetic basis for complex alternative strategies segregating within populations [4–6]. Intriguingly, the same properties that make supergenes important in organismal adaptation also make them common contributors to selfish evolution within populations and species. Indeed, many of the classic killer meiotic drive systems are poison-antidote or killer-target systems where distinct functional components of the system are locked together in inversions [reviewed in 7–10]. When maintained within species, these selfish supergenes also tend to accumulate genetic load on top of the intrinsic fitness costs of their killing mechanism [11]. However, while the selfishness of many killer meiotic drive systems absolutely requires interactions between multiple linked genes, the role of supergenes in other forms of meiotic drive remains less clear.

By contrast to killer meiotic drive systems, which gain a relative transmission advantage by destroying alternative products of meiosis, female meiotic drive (FMD) occurs via over-representation in a fixed pool of gametes [8,11]. In FMD, variant chromosomes capitalize on the inherent asymmetry of female meiosis common in plants and animals to preferentially segregate to

the lone egg cell [12–14]. Because a variant chromosome can manipulate the outcome simply through biased structural interactions with the asymmetric spindle apparatus rather than through the action of linked genes [15], such FMD need not involve linked genes. Nonetheless, FMD detected in natural populations may often be tied up in supergenes for two reasons. First, both genic and structural factors may be functionally necessary to effect or enhance drive. Second (and non-exclusively), FMD polymorphisms may only persist on the landscape long enough to be detected if they carry genetic baggage generated by suppression of recombination within a supergene. Both of these factors play a role in the classic neocentromeric Ab10 driver in maize, one of the best characterized FMD systems. Ab10 drive requires both novel meiotic motors (tandem arrays of kinesin genes; 16,17) and a vehicle (heterochromatic knobs of repetitive DNA; reviewed in 18) to subvert Meiosis II and physically drag Ab10 and other knobbed chromosomes to the outside poles (one of which will become the egg) in heterozygotes. The Ab10 driver thus functions as a selfish supergene with linked structural and genic components. Because the large region of suppressed recombination necessary for Ab10 drive also captures hundreds of other genes, hitchhiking deleterious mutations also likely contribute to its substantial costs and polymorphism across the range of maize [19–22].

Centromeres, which can distort segregation completely in Meiosis I, theoretically need no driving motor beyond the ordinary asymmetric machinery of female meiosis. Indeed, a series of elegant studies in mouse oocytes demonstrates that variant 'strong' centromeres with expanded arrays of satellite DNA and larger kinetochores predictably take advantage of transient meiotic polarity to drive in hybrids [23–26 reviewed in 26]. However, molecular work also reveals that the opportunity for drive is exquisitely dependent on the cellular environment and can be accentuated or attenuated by manipulation of the genic background [27]. This direct evidence for the context-dependence of centromeric FMD, along with karyotypic evidence that the cellular arena defining centromere 'strength' is evolutionarily labile across mammals [28], suggests that genes may matter greatly in creating opportunities for centromeric drive. Thus, a structural variant (e.g. inversion, translocation, transposition) that generates tight linkage between an enhancer of drive (e.g. gene variants that create greater spindle or kinetochore asymmetry) and a centromere could, like Ab10, both maximize its own spread as a motor-vehicle supergene [29] and alter the competitive context for other chromosomes (as Ab10 does for unlinked knobs; 22). Furthermore, as with driving Robertsonian translocations in mice [23,28], centromeric drive may be directly precipitated by chromosomal rearrangements that alter centromere size and position. If such structural mutations simultaneously capture genes that can act as linked enhancers of drive the conditions for supergene evolution may be set. Further, if suppression of recombination associated with the birth and spread of a strong centromere also generates linkage with deleterious variants (or the opportunity for new deleterious mutations to accumulate; 30), it may be prevented from rapid fixation and thus be more likely to be detected. However, because there are few empirical cases of centromeric drive, little is known about the role of linked genes in the evolution of selfish centromeres.

Here, we investigate the potential for supergene behaviour in one of the few selfish centromeres known to segregate in wild populations, *D* (the driving allele of the meiotic drive locus on chromosome 11 (MDL11)) in the yellow monkeyflower *Mimulus guttatus*. *D* was initially discovered in an interspecific cross with the closely related selfing species *M. nasutus* (*d*), where it drives nearly 100 : 0 via female function [31,32]. Such strong FMD is only possible by completely distorting segregation in Meiosis I, which only centromeres can do [33]. The centromeric function of *D* is confirmed by its association with massively expanded arrays of the *M. guttatus* centromere-specific satellite repeat Cent728 [32]. Within the Iron Mountain (Oregon) annual population in which *D* was initially discovered, *D* drives approximately 60:40 against conspecific MDL11 genotypes (D^-; [32]). At Iron Mountain, *D* has no detectable effects on fertility when heterozygous (i.e. when drive occurs), but causes approximately 20% reduction in pollen viability (wild plants and controlled crosses) and female fecundity (seedset/fruit measured in the field) in *DD* homozygotes [32,34]. These recessive costs predict an equilibrium frequency for *D* close to the observed 35–40%, generate unexpectedly high genetic load in this large outcrossing population [35], and maintain costly polymorphism that selects for suppression or resistance to drive [34].

Population genomic analyses reveal that *D* is not just a centromere, but a massive and uniform haplotype (containing greater than 350 genes as well as satellite DNA arrays; greater than 12 Mb). Two factors likely cause this large extended haplotype. First, *D* is located within a structural variant (probably an inversion) on LG11 that shows suppressed recombination [36]. Second, *D* has swept through the IM population within the past 1500 years (assuming one generation per year, as is typical in annual wildflower populations), and selective sweeps can cause strong, long-range linkage disequilibrium [37]. The recent spread of the drive haplotype has precipitated a concurrent sweep of one of *M. guttatus*'s two copies of the key kinetochore protein Centromeric Histone 3 (CenH3A; [38]), which also acts as an unlinked modifier (suppressor) of drive in experimental hybrids [37]. However, genetic mapping also shows that structural and/or genic variation within the MDL11 locus itself is the primary source of the dramatic difference in the strength of *D* drive against D^- (conspecific) and *d* (heterospecific) competitors [37]. These features make *D* an ideal model for understanding centromeric drive and the formation of selfish supergenes, but the lack of recombination between *D* and either D^-/*d*, and its striking homogeneity within our focal IM population, have limited genetic dissection. Thus, our joint aims here are (1) to characterize the distribution, variation and effects of *D* beyond the IM population, as a platform for genetic dissection of drive and associated costs, and (2) to investigate patterns of gene expression within MDL11 (i.e. *D* versus D^- and *d*) to identify candidate linked modifiers of drive and/or contributors to its substantial reproductive costs.

For the first aim, we capitalize on whole-genome sequences from the nearby Cone Peak population of *M. guttatus* to broaden our understanding of the population dynamics of centromeric drive. The Cone Peak (CP) and IM populations are less than 2 km apart and only weakly differentiated across their nuclear genomes, but have functionally and evolutionarily distinct mitochondria and different histories of selection on the matched nuclear restorer locus [39]. Thus, their generally low differentiation likely reflects very large effective population sizes sampling from a shared pool of regional

variation rather than high current rates of CP-IM pollen and seed migration [40]. Population genomic analyses suggested that *D* is not unusually divergent from D^- at the gene scale (i.e. not the product of long-range dispersal or interspecific introgression) and could even have been derived by structural mutation alone within IM [37]. Furthermore, although *D* is known to be present at a young (less than 40 years) disturbed site near IM and CP showing admixture between annual and perennial *M. guttatus* ecotypes [41,42], its range is unknown. Thus, we do not yet know whether the sequence variation, costs and dynamics of *D* at IM are unique or representative of a widespread balanced polymorphism. From the CP sequences, we also generate new markers to assess the male fertility (pollen viability) effects of *D* and its local frequency in a large, wild-collected sample from that population. We then conducted a first-pass scan for *D* genotypes in plants collected (as seeds) from 11 additional populations from a transect across western Oregon [43].

For the second aim, we analyze two RNASeq datasets to assess gene expression divergence associated with the recent spread of the driving *D* haplotype at IM. Genes that show shifts in expression in one or both datasets, in addition to those in regions of gene content differences, are potential candidates as enhancers of *D* and/or the source of its male and female fertility costs. First, we examine patterns of differential expression between *M. guttatus* IM62 (*D*) and *M. nasutus* (*d*) stamen (male) and carpel (female) tissues for genes across MDL11 (and flanking control regions). These data were previously used as controls in a study of interspecific hybrid incompatibility, which caused abundant differences in genome-wide gene expression between fertile and sterile hybrid siblings [44]. Thus, these samples capture key reproductive stages when we might also expect differences in expression relevant to drive and its costs. Second, we used a large eGWAS dataset ($N = 151$ lines; [45]) to investigate patterns of expression divergence between *D* and D^- haplotypes within IM. This large dataset from young floral buds (which are a mix of tissues) do not precisely capture tissue-specific divergence in gene expression, but is ideally suited for characterizing the direct consequences of *D*'s sweep for constitutive expression variation at the population scale. Because there is little evidence of *trans*- effects of common alleles on gene expression on other chromosomes [45], we focus on loci within MDL11 and flanking regions. While linkage disequilibrium within MDL11 makes it impossible to rule out (or in) any gene in the interval as a contributor to drive and costs, significant shifts in expression (particularly those consistent across datasets) may help identify functional candidates important in the biology of this selfish supergene.

2. Methods

(a) Cone peak genome sequence, alignment, SNP calling and data filtering

Whole-genome re-sequence data (fastqs, Illumina reads) were obtained from the Sequence Read Archive for eight lines from CP [39]. We also generated new sequence data for three CP lines, as described in [37,39]. In brief, DNA was extracted via a modified CTAB-chloroform extraction protocol [31], libraries were prepared using the Nextera tagmentation protocol, and prepped libraries were sequenced on a NextSeq platform (Illumina NextSeq paired-end, 150 bp reads; Ilumina Inc., San Diego, USA). Information regarding samples, sequence data, MDL11 haplotype call and source are detailed in electronic supplementary material, table S1.

Sequencing filtering, mapping and variant calling are as in [37]. To summarize, Trimmomatic v. 0.35 was used to trim low-quality sequences and adapters from raw sequences [46]. Cleaned sequences were aligned to the *M. guttatus* v2 reference genome (https://phytozome-next.jgi.doe.gov/) using bwa mem v. 0.7.15 with default parameters [47]. We used SAMtools v. 1.3 [48] to remove reads with mapping qualities less than 29 and Picard tools v. 1.119 (http://broadinstitute.github.io/picard) to remove duplicate reads. The Genome Analysis Toolkit (GATK) was used to re-align around indels and call variant sites with the Unified Genotyper tool [49,50]. We restricted sites to biallelic SNPs within genes using vcftools v. 0.1.12b, and also converted sites covered by less than three reads to missing data [51]. Because the inbred lines from CP are not fully homozygous (mean H_{OBS} = per individual = 0.123, s.d. = 0.02), we retained heterozygous sites and randomly kept either the reference or alternative allele. For population genomic analyses, sites with genotype calls from at least seven (CP) individuals were retained and genes with fewer than 150 retained sites were removed or coded as missing data. For visualizations of chromosome 11, we used the re-ordered collinear MDL11 map generated in [37].

We used vcftools v. 1.12a to calculate linkage disequilibrium (i.e. the squared correlation coefficient between genotypes = r^2) for SNP pairs [51]. To accommodate the heterozygosity present in CP, SNPs were subsampled to greater than 1000 bp apart prior to calculating r^2 ($N = 4{,}073$ SNPs at CP). Average r^2 across all polymorphic sites was then calculated for each gene pair ($N = 1475$ genes). We explored haplotype structure by calculating the proportion of SNPs per gene on chromosome 11 that matched the reference genome (IM62, *D*) for each line. For the haplotype structure analyses, we removed CP5 due to heterozygosity across MDL11 and coded genes with fewer than seven polymorphic sites genotyped as missing data (total $N = 1064$ genes). To build haplotype networks and compare *D* with IM haplotypes, we identified exonic mutations in CP lines as described for IM in [37]. We excluded heterozygous sites and entire genes with more than five heterozygous sites, which generally indicate incorrect alignments. We used SplitsTree4 [52] to generate haplotype networks of the SNP variation for all CP and IM *D* lines.

(b) Marker development and genotyping

In the IM *M. guttatus* population, a single 276 bp amplicon at the marker lb5a indicates *D*, while diverse other alleles indicate D^-; this simple diagnostic allowed accurate characterization of *D*'s frequency and fitness costs even in diverse outbred wild plants [32,34]. However, lb5a includes a small non-genic region as well as the 5′UTR and part of the first exon of Migut.K00858, and visual inspection of the CP alignments in Integrative Genome Viewer (IGV 2.10; Broad Institute, Boston, MA) indicated that multiple indels in the repetitive upstream region led to a loss of informativeness at CP, with some *D* and D^- individuals convergently sharing predicted amplicon sizes. We developed a new marker (mK0858, referred to here as 'mK858'; electronic supplementary material, table S2) flanking a small microsatellite (4–5 3b repeats) within the coding sequence of Migut.K00858; the reference allele shares a 3 bp insertion with all other *D* lines relative to all D^- lines from CP (electronic supplementary material, table S1) and IM [37] and was also distinct in predicted amplicon length from range-wide (non-*IM*) *M. guttatus* examined ($N = 22$; see [40] for the list of non-IM accessions examined in IGV). Genotypes at mK858 (which amplifies only two alleles in all samples in this study) were highly but imperfectly correlated with lb5a genotypes in our outbred Cone Peak

samples ($n = 169$; $r^2 = 0.58$) and the mismatches were consistent with under-calling of *D* alleles at lb5a versus mK858. Therefore, we substitute mK858 for lb5a as our *D*-diagnostic marker in this study. We also genotyped CP individuals at mK1229 (electronic supplementary material, table S2), which is diagnostic of the approximately 45-gene MDL11 subregion with consistent gene content differences between *D* (present) and D^- (absent) lines at IM and includes an internal control [37].

For genotyping, DNA was extracted from leaf or bud tissue using a 96-well CTAB chloroform extraction protocol (http://dx.doi.org/10.17504/protocols.io.bgv6jw9e), quantified with a spectrofluorometer, and diluted to a target concentration of 2–10 ng μl^{-1}. The fluorescently labelled markers were amplified singly or in multiplex using a standard touch-down PCR protocol, fragments were separated and sized using an ABI 3130 Genetic Analyzer, and genotypes visualized and scored using GeneMapper software, as previously described [34].

(c) CP and Oregon transect populations

To test whether the frequency and male fertility effects at CP paralleled those at IM, we grew up (in 2021) one progeny each from outbred maternal seedsets wild-collected at Cone Peak in 2013 (total $N = 231$). Seeds were stratified (one week at 4°C) and germinated in Petri dishes on wet sand, and then a single seedling per family was transplanted into 2″ pots filled with Pro-mix #4. Plants were grown to flowering in a Percival PGC-40 growth chamber with high-intensity LED lights and a 16/8 hr (15/20°C) day/night cycle. We collected all four anthers from one flower of each plant on its day of anthesis and assayed pollen viability by lactophenol aniline blue staining of pollen grains (fertile = blue, sterile = unstained) counted on a hemocytometer, following established protocols [53]. We obtained pollen fertility data from 180 CP individuals (mean proportion viable pollen = 0.77).

To assay the incidence of *D* further afield, we grew up (under the same culture conditions as above) wild-collected progeny of maternal plants from 11 annual *M. guttatus* populations collected from a transect across western Oregon (electronic supplementary material, table S3; [43]). Few maternal families per population germinated, so we retained all individuals (total $N = 77$, $n = 1$–6 moms and 2–19 progeny per population). Four wild individuals from IM were included in the runs as positive controls.

(d) Differential gene expression analyses

To examine patterns of gene expression across MDL11 haplotypes, we used publicly available RNASeq reads from two datasets. Specifically, we analysed (1) young bud transcriptomes from an eGWAS study of inbred lines from the IM population of *M. guttatus* that were either *D* or D^- homozygotes ([45] SRA project number PRJNA736440) and (2) carpel and stamen transcriptomes collected from three replicates each of a *M. guttatus DD* individual (IM62) and an *M. nasutus dd* individual ([44] Biosample project # SAMN13979871- SAMN13979871).

For the intraspecific datasets, RNASeq data from 151 IM lines (*D* or D^- homozygotes) were initially treated as described in [45]. MDL11 haplotype calls followed [54] and five MDL11 heterozygotes (125, 294, 1133, Z368 and Z99) were dropped. Mapping, variant calling, and filtering are described in detail in [45]. In brief, reads were trimmed with bbuk from BBTools 38.86 (https://sourceforge.net/projects/bbmap/) and aligned to the *M. guttatus* v2 reference genome (https://phytozome-next.jgi.doe.gov/) with STAR 2.5.0a [55]. Gene counts were generated by htseq-count 0.11.2 [56] and normalized using the estimateSizeFactors function in DESeq2.1.28.1 [57]. For differential expression analyses, genes were first filtered for minimal expression, read counts were standardized by library, the effect of cohort was regressed out and each gene was Box-Cox normalized in preparation for linear model fits, as in [45]. Two models were then fit to ask if gene expression varied according to MDL11 haplotype (*D* versus D^-). First, significant differential expression between *D* and D^- was determined using one-way ANOVAs with a Bonferroni corrected threshold ($p < 4.4 \times 10^{-5}$). Second, we examined whether gene expression level was a function of MDL11 genotype as a fixed effect, with line as a nested random effect, using the lme4 package in R [58]. Because lme4 does not calculate *p*-values, we considered genes significantly differentially expressed using the equivalent *F*-statistic that accompanied a *p*-value of 4.4×10^{-5}. We examined whether proportions of differentially expressed genes within MDL11 versus flanking regions on chromosome 11 (268 and 862 genes, respectively) varied according to MDL11 haplotype by computing pairwise two proportion *z*-tests in R. A principal components analysis (PCA) was conducted on the Box-Cox normalized read counts using the factoextra package in R [59].

For the interspecific datasets (IM62 carpel, IM62 stamen, SF carpel, SF stamen), mapping, variant calling, read counting and filtering was performed following [44]. In transcriptomic studies, interspecific divergence can create mapping biases when aligning reads from one focal species (i.e. *M. nasutus*) to a reference from the other (i.e. *M. guttatus*). One strategy to reduce this bias is to map to a diploid pseudogenome that represents nucleotide variation present in both species [44]. To account for divergence of *M. nasutus* from the IM62 *v2* reference genome (*M. guttatus*-*M. nasutus* $d_s = 4.94\%$; [60]), reads were mapped to a *M. guttatus* IM62-*M. nasutus* SF diploid pseudogenome generated from the IM62 reference and high-quality SF genomic SNPs, as in [44]. The diploid pseudogenome was constructed by first mapping whole genome SF5 data (NCBI Sequence Read Archive database: SRR400478) to the *M. guttatus v2.0* reference genome using BWA-MEM [47,48]. Variants were called with GATK HaplotypeCaller to identify biallelic SNPs and a SF5 pseudoreference was made using GATK FastaAlternateReferenceMaker [49,61]. A diploid *M.guttatus* IM62-*M.nasutus* SF5 pseudoreference genome was made by merging the two genomes into a single file, while retaining allelic identifiers (e.g. IM62 or SF5). RNA-Seq reads were filtered using Trimmomatic v. 0.35 [46], mapped to the pseudoreference genome using STAR 2.5.0a [55] and read counts were generated with htseq-count 0.11.2 [56]. Additional details regarding pseudogenome construction and RNA-Seq alignment are as in [44]. Read counts were obtained from a total of 28 140 genes, but our analyses here focus on MDL11 and flanking regions from chromosome 11.

For the interspecific comparisons, we determined differential expression in edgeR [62], following a similar workflow to [44]. To retain only those genes with expression in at least one treatment, we required that genes have at least three counts per million for two or more individuals (21 409 genes retained). The remaining data were normalized with the calcNormFactor function using default values. We examined within-subject correlations of gene expression using the duplicateCorrelation function in the limma R package [63]. Because the correlation was low (approx. 0.07), we built a design matrix that examined the fixed effects of our treatments (tissue, line) but did not pair by subject. This allowed us to make the full suite of comparisons both within and between subjects. A differential expression analysis was performed using the glmTreat function in edgeR [62], which applied a stringent test requiring a log2-fold change greater than 1.25 for assigning significant differential expression. Genes with a false discovery rate less than 0.05 were considered differentially expressed. We looked for differential expression of genes (1) between carpel and stamen tissue within each line (IM62, SF; 'Carp-Stam IM'; 'Carp-Stam SF'), (2) between IM62 versus SF within each tissue (carpel, stamen; 'IM-SF Carpel'; 'IM-SF Stamen') and (3) showing evidence of a significant line * tissue interaction (Line * Tissue). We examined

whether the proportion of differentially expressed genes inside versus outside MDL11 on chromosome 11 (242 and 912 genes, respectively) varied according to treatment by computing pairwise two proportion *z*-tests in R. Tests with a false discovery rate less than 0.05 were considered significant.

For both interspecific and intraspecific RNASeq datasets, we repeated all proportion tests after removing a 45-gene block missing in D^-/d haplotypes relative to *D* [37]. We also examined MDL11 and flanking regions for (1) the proportion of up- and downregulated genes that were differentially expressed in each tissue and (2) the proportion of shared up- and downregulated differentially expressed genes in the intra- and interspecific datasets. All analyses were done in R v. 4.0.4 [64] and RStudio v. 1.4.1106 [65], unless stated otherwise.

3. Results

(a) *D* exhibits similar signatures of selection at CP, but has a distinct evolutionary history and more variation

Inbred lines from the nearby Cone Peak (CP) population exhibit parallel elevated linkage disequilibrium across MDL11 (figure 1*a*), as was also seen at IM. Consistent with pervasive drive, *D* experienced a selective sweep to intermediate frequency at CP (figure 1*b*); half of our partially inbred CP lines carry the swept IM62-reference like *D* haplotype (5 *DD* homozygotes and 1 DD^- heterozygote; $N = 11$ total; electronic supplementary material, table S1) across a core MDL11 region of hundreds of genes. However, one of the five homozygous *D* lines (CP24) exhibits a shorter *D* haplotype than other CP lines, specifically shifting to non-reference genotypes over an approximately 45-gene region lacking coverage in IM D^- lines (Migut.K01214-Migut.K01259; see [37]), as well as flanking genes present in all lines (27 and 4 genes on exterior and interior sides, respectively; approx. 2.9 Mb shorter than *D* in IM). Otherwise strong haplotype structure across MDL11 also breaks down for those distal 27 genes (but not the 45-gene insertion) in other CP lines, generating an overall shorter core MDL11 block in the CP population (figure 1*a*,*b* Migut K01047-Migut.K01259; approx. 1.4 Mb shorter than IM) relative to our inferences from the recently swept IM population [37]. The highly repetitive and abundant nature of the centromere-associated repeats linked to *D* [32,37] make the MDL11 region particularly challenging for physical sequence assembly and we rely on a previous work that re-ordered chromosome 11 to better match linkage mapping data [36,37]. Thus, the size and breakpoints of the haplotypes described here are likely to shift as improved assemblies become available, and both structural variation (and resultant suppression of recombination) and selection contribute to the inferred haplotype boundaries of MDL11 at IM [36,37]. The shorter core haplotype at CP may reflect an older sweep, with time for additional recombination to break down associations outside the functional core that distinguishes *D* and D^-, or a more complex history potentially including the evolution of a larger region of structural difference and suppression of recombination at IM.

To further characterize within-*D* variation, we scored CP genotype at nine exonic SNP variants previously used to estimate the age of MDL11 sweep at IM and identified novel (invariant at IM) SNPs in the five CP *D* lines (including CP24 but excluding sites outside the core *D* haplotype at CP; electronic supplementary material, table S4). The core MDL11 haplotypes (spanning approx. 10 Mb) are nearly identical in CP and IM, with only eight exonic SNPs unique to CP (figure 1*b*), suggesting that the two haplotypes share a common origin. Mutations on this shared core show that the most common (and inferred ancestral) *D* sequence at IM is not present at CP; instead, all five of our CP lines share three SNPs found in only in one of the 13 IM *D* lines previously assayed (IM1145; figure 2*a*). This suggests that the selective sweeps at CP and IM, while likely parallel in their drive dynamics and local equilibrium (see below), had independent trajectories. Although we do not have direct crossing evidence that the shorter *D* haplotype found at CP drives through female meiosis, a parallel sweep to similar frequency (and similar costs; see below) suggest that the common *D* haplotype at CP behaves similarly to the longer IM haplotype. CP24, unlike the other CP lines (which each had less than two mutations versus the inferred IM1145/CP49-like CP ancestor), had a total of six variants. Although CP24 also has a uniquely short *D* haplotype, these SNPs are all singletons at the gene scale, suggesting independent mutation rather than recombination or gene conversion with distinct D^- haplotypes. This contrasts with the star-like structure of *D* variation (figure 2*a*) within IM (and other CP lines) and suggests that CP24 may represent a deeper/older pool of *D* variation outside IM (especially given our small sample). Because this excess accumulation of mutations in one lineage violates the assumption of simple mutation-counting models for ageing sweeps [66], and because the edge of the *D* haplotype is generally less clear, we do not attempt to use the number of new mutations to age the spread of *D* at CP, as we did at IM [37].

CP24's possession of a distinct long-range swept *D* haplotype, minus the 45-gene insertion characteristic of other *D* lines, may be due to rare recombination during/after a sweep by longer haplotypes, but also raises the intriguing possibility that it represents an older *D* haplotype. Under such a scenario, short CP24-like haplotypes drive but have been largely displaced at CP and IM by structurally distinct *D* haplotypes containing the Migut.K01212–K01259 insertion. Because this *D*-specific insertion includes at least one cyclin (Migut.K01228/1229), one kinesin (Migut.K01254) and several copies of a histone h3k9 demethylase (Migut.K01225, Migut.K01216) involved in regulation of heterochromatin (IBM1; [67]), it is also tempting to hypothesize a genic enhancer role for this region. However, because current *M. guttatus* genome assemblies are either *D* but misassembled across MDL11 (V2; https://phytozome-next.jgi.doe.gov) [36,37] or well-assembled but non-*D* (V4 and V5; https://phytozome-next.jgi.doe.gov), we cannot yet accurately reconstruct the structural events and genic divergence associated with *D*'s emergence and spread. Nonetheless, greater within-*D* variation at CP (and elsewhere; see below) provides a potentially greater time-frame over which we can infer its evolution, and also allows crosses between divergent *D* haplotypes to investigate variation in the strength and costs of drive.

(b) Parallel frequency and male fertility costs at CP, as well as evidence of *D* in other Oregon populations, suggest centromeric drive and its costs may be widespread in *M. guttatus*

In outbred CP plants, the allele frequency of *D* was 0.29 (N genotyped at mK858 = 188; electronic supplementary material,

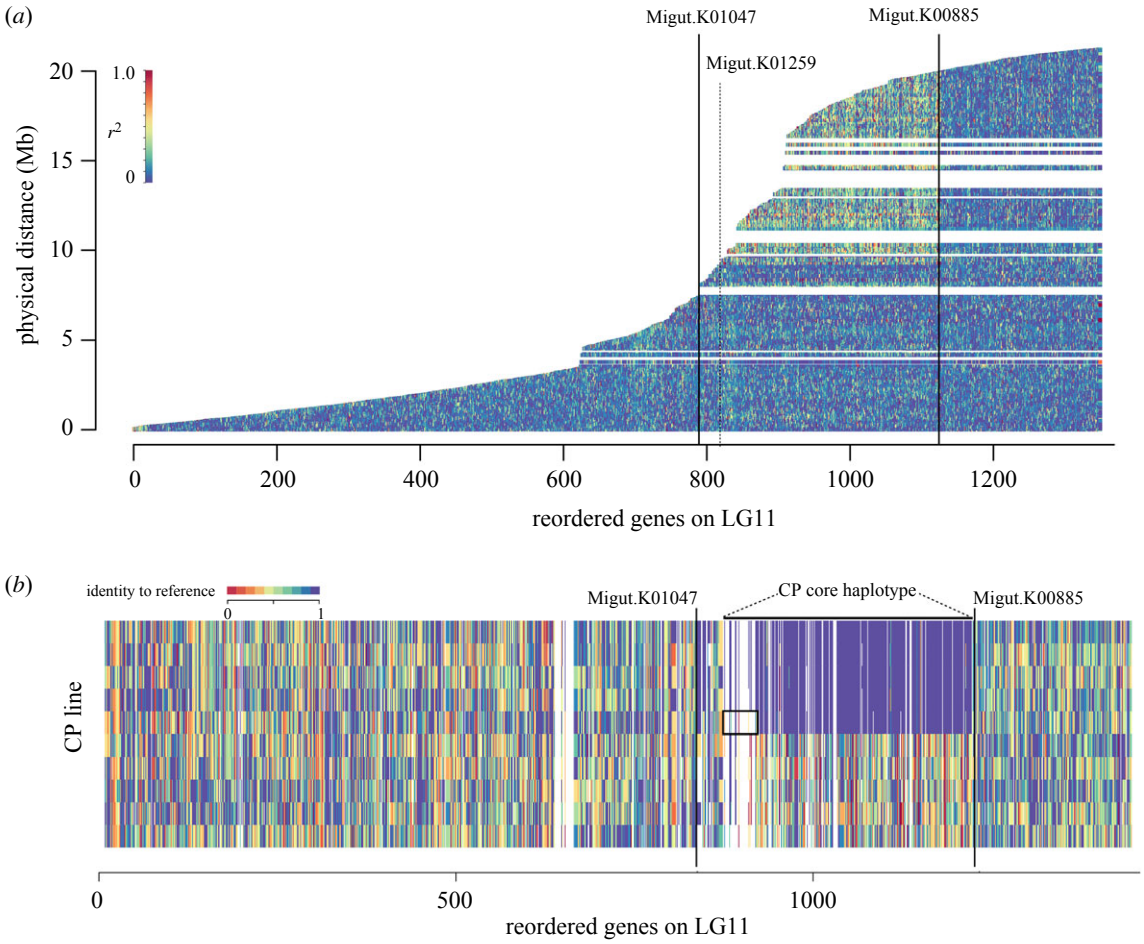

Figure 1. The centromeric driver (*D*) is a large, extended and swept haplotype in Cone Peak (CP) *M. guttatus*, but distinctly shorter than in the IM population in which it was discovered. (*a*) Linkage disequilibrium and (*b*) haplotype structure are shown across chromosome 11 in inbred lines from CP, a population approximately 1.7 km from IM with low nuclear differentiation [39]. Solid vertical lines bound the MDL11 region in IM [37]. (*a*) Pairwise estimates of linkage disequilibrium (r^2) per gene, for all genes on chromosome 11. The region between Migut.K01259 and Migut.K01047 is part of MDL11 in IM, but missing from the core MDL11 haplotype in CP. The horizontal axis displays ordered genes on chromosome 11, while the vertical axis scales shows genes along with estimated physical positions on the chromosome from the re-ordered map in [37]. (*b*) Proportion of SNPs in a gene that match IM62, the driving reference haplotype ($N = 1328$ genes with sufficient data; genes with insufficient data are coded in white). The long horizontal line shows the core MDL11 haplotype in CP, while the vertical lines represent the genes bounding the MDL11 haplotype in IM (Migut.K0147-Migut.K0885). The rectangle shows the region that is found in the core CP MDL11 haplotype but excluded from the CP24 haplotype. Rows represent individual inbred lines isolated from CP, from top to bottom: CP48, CP49, CP44, CP47, CP24, CP19, CP37, CP27, CP50, CP36.

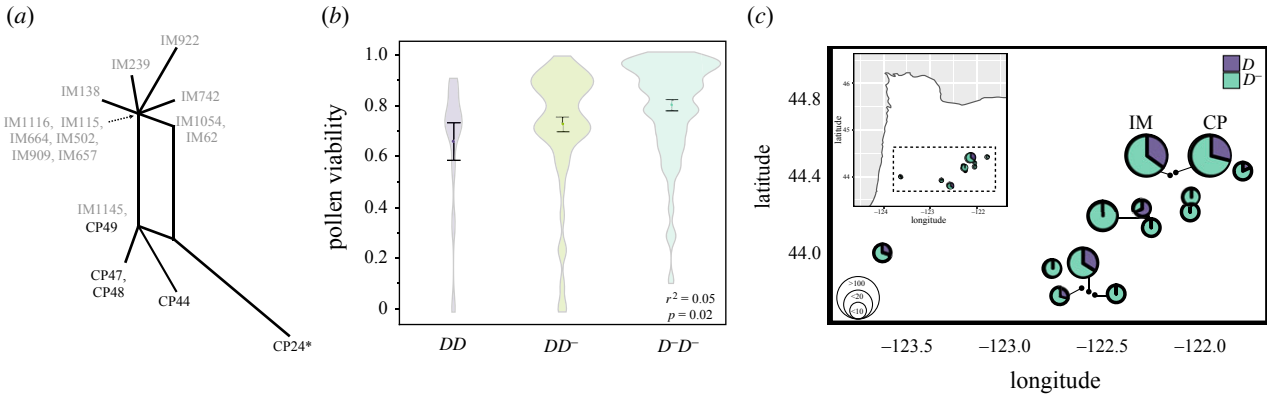

Figure 2. The centromeric driver (D) has distinct haplotype structure and male fertility costs at CP, as well as a widespread distribution across Oregon. (*a*) Haplotype network of the MDL11 region for *DD* individuals at IM (grey) and CP (black). The shorter and distinct haplotype of CP24 is marked with an asterisk (*). (*b*) Violin plots of pollen viability at CP for *DD*, DD^- and D^-D^- genotypes as determined by marker mK858 ($N = 146$). Points are means and error bars are +/− one standard error. (*c*) Estimates of *D* frequency for 11 populations across Oregon. Inset shows Oregon zoomed out for orientation purposes and circles represent sample sizes (electronic supplementary material table S3).

table S5), comparable to estimates from IM (0.35–40.0; [34]). To assess the presence of CP24-like recombinant haplotypes, we also genotyped the CP plants at mK1229, which segregates as a presence-absence polymorphism diagnostic of the 45-gene indel (with an internal control; [37]). If we recode mK858 to match the dominance of mK1229 (present = *DD* or DD^-), the markers are highly congruent, as expected from the sequenced lines; however, 13/155 (8%) individuals

appeared recombinant. Notably, only 2/75 (less than 3%) of individuals scored as D^- homozygotes at the highly diagnostic mK858 were carriers of the *D*-associated insertion including mK1229. By contrast, 11/80 (14%) individuals carrying the *D* haplotype at mK858 lacked a mK1229 band and were scored as D^- homozygotes there (i.e. had a CP24-like genotype). Because mK858 appears diagnostic of the core long-range *D* haplotype (and its expected costs; see below), this asymmetric mismatch supports the scenario outlined above, in which the approximately 45-gene insertion characterizes a novel and successful subset of *D* haplotypes fixed at IM and predominant at CP. However, it is also possible that these individuals represent recent recombination outside the structural bounds of *D*, and the association of this insertion with *D* at IM is transitory; new long-read assemblies of *D* and D^- haplotypes and additional sequencing of variant lines will be necessary to distinguish these scenarios.

MDL11 genotype (as assayed at mK858) significantly affected male fertility (N =146, $p = 0.02$, $r^2 = 0.05$); *DD* plants exhibited a 17.5% reduction in pollen viability relative to D^-D^- plants (0.66 ± 0.06 versus 0.81 ± 0.02; figure 2*b*) comparable to male fertility costs at IM [32]. However, heterozygotes were intermediate in fertility (0.73 ± 0.03), whereas *D*'s male fertility costs at IM and in interspecific hybrids were strictly recessive [32]. Further, there was no significant relationship between mK1229 genotype (45-gene *D* insertion present versus not) and pollen viability ($n = 124$, $p = 0.21$). Lower power at this dominant marker may be a factor, but potential recombination with the core *D* haplotype may also reduce the association with genic and/or structural sources of the costs. Because the predicted frequency of *D* depends strongly on the magnitude and dominance of both male and female fertility effects [34], additional experiments (particularly controlled crosses where the sample size of *D* homozygotes can be maximized) will be necessary to fully characterize its dynamics outside of IM.

Finally, we identified potential *D* alleles (as indicated by mK858 genotype) in five of the 11 additional Oregon populations screened (figure 2*c*), including the easternmost (MWL, 30 km east of IM), and westernmost (WC, 125 km west in the Coast Range). We do not place much weight on the calculated allele frequencies (figure 2*c*), as the sample sizes are small, not all samples within a population were from independent maternal plants, and mK858 may not be perfectly diagnostic in these populations. However, the 0.34 frequency of *D* in the largest sample (MO; $N = 19$ progeny from 6 moms) is strikingly similar to the values observed at CP and IM (electronic supplementary material, table S3). This suggests that centromere-associated drive and costs and the dynamics characterized in detail at IM may be a common feature of populations in this region, creating a rich opportunity to explore the mechanism and dynamics of selfish supergene evolution across a complex metapopulation [68].

(c) MDL11 genes exhibit elevated interspecific expression divergence in reproductive tissues and intra-population shifts in expression reveal candidate modifiers

Analyses of young floral buds (from a panel of 151 IM *D* and D^- lines) and developing female (carpels) and male (stamens) reproductive tissues (from *DD* (IM62 *M. guttatus*) and *dd* (SF *M. nasutus*) lines) suggest that meiotic drive has altered patterns of gene expression in *M. guttatus* (figure 3*a–d* and electronic supplementary material, figure S1; tables S6 and S7). In conspecific comparisons (*D* versus D^-) of developing bud transcriptomes, proportionally more genes are differentially expressed in the MDL11 region than flanking regions (figure 3*a*). ANOVAs (FDR-adjusted $p < 0.0001$) and mixed models (FDR-adjusted $p < 0.0001$) show consistent patterns. Further, results are qualitatively similar after removing a 45-gene block that is missing in D^- haplotypes in IM, suggesting that copy number variation alone does not explain the differences in expression profiles ($p < 0.0001$ in both comparisons). Indeed, a large proportion of the genes within MDL11 are differentially expressed (30–35%), while, as expected, almost no differential expression occurs outside the MDL11 region (approx. 1%; figure 3*a*). In comparisons between *DD M. guttatus* and *dd M. nasutus*, where interspecific divergence is expected genome-wide, the proportion of differentially expressed genes is still elevated within the MDL11 versus flanking regions in female tissues (IM-SF Carpel; $p = 0.00249$) and is marginal in male tissues (IM-SF Stamen; $p \sim 0.063$; figure 3*b*). While between-tissue comparisons of differential expression (Carpel-Stamen IM; Carpel-Stamen SF) show the highest proportion of differentially expressed genes, they do not differ within versus outside the MDL11 region (figure 3*b*; $p > 0.05$ in all cases). Likewise, region does not impact the proportion of differentially expressed genes showing an interaction between line and tissue (figure 3*b*; $p > 0.05$). After the missing block of 45 genes is removed, results are qualitatively similar with the exception that the within MDL11 region versus flanking region comparisons in IM-SF Stamen are no longer marginally significant ($p = 0.15$). Together, these results suggest elevated expression differences in drivers versus non-drivers specifically within the MDL11 region (and not flanking regions), and that female tissues drive this pattern.

To further explore variation in intrapopulation gene expression related to drive, we performed a PCA of gene expression patterns in the MDL11 and flanking regions on chromosome 11 (figure 3*c*,*d*). We find striking contrasts between the two regions. Within MDL11, individuals cluster according to MDL11 haplotype and confidence interval ellipses around the means of *D* and D^- are completely non-overlapping (figure 3*c*). In the flanking region, *D* and D^- individuals show no such pattern and have overlapping means (figure 3*d*). Dimension 1 of the MDL11 region explains a higher proportion of the variance (approx. 14%) than dimension 1 of the flanking region (approx. 10%), and only dimension 1 of MDL11 distinctly separates *D* and D^- individuals (figure 3*c*,*d* and electronic supplementary material, figure S2*ab*). Thirty-four genes (electronic supplementary material, table S8), including three from the 45-gene insertion in *D* individuals, make contributions greater than 1 to dimension 1. Most of these 34 show elevated levels of expression in *D* versus D^- individuals, suggesting that upregulation of MDL11 genes in *D* individuals underlies major differences in gene expression among the haplotypes (positive dimension 1 values on electronic supplementary material, figure S2*c*). We find that gene expression patterns within, but not outside, MDL11 cluster individuals according to drive haplotype and that several (largely upregulated) genes drive the overall pattern.

One intriguing candidate gene that contributed significantly to dimension 1 and is outside the 45-gene indel is Migut.K00842, the sole *Mimulus* homologue of the Nuclear

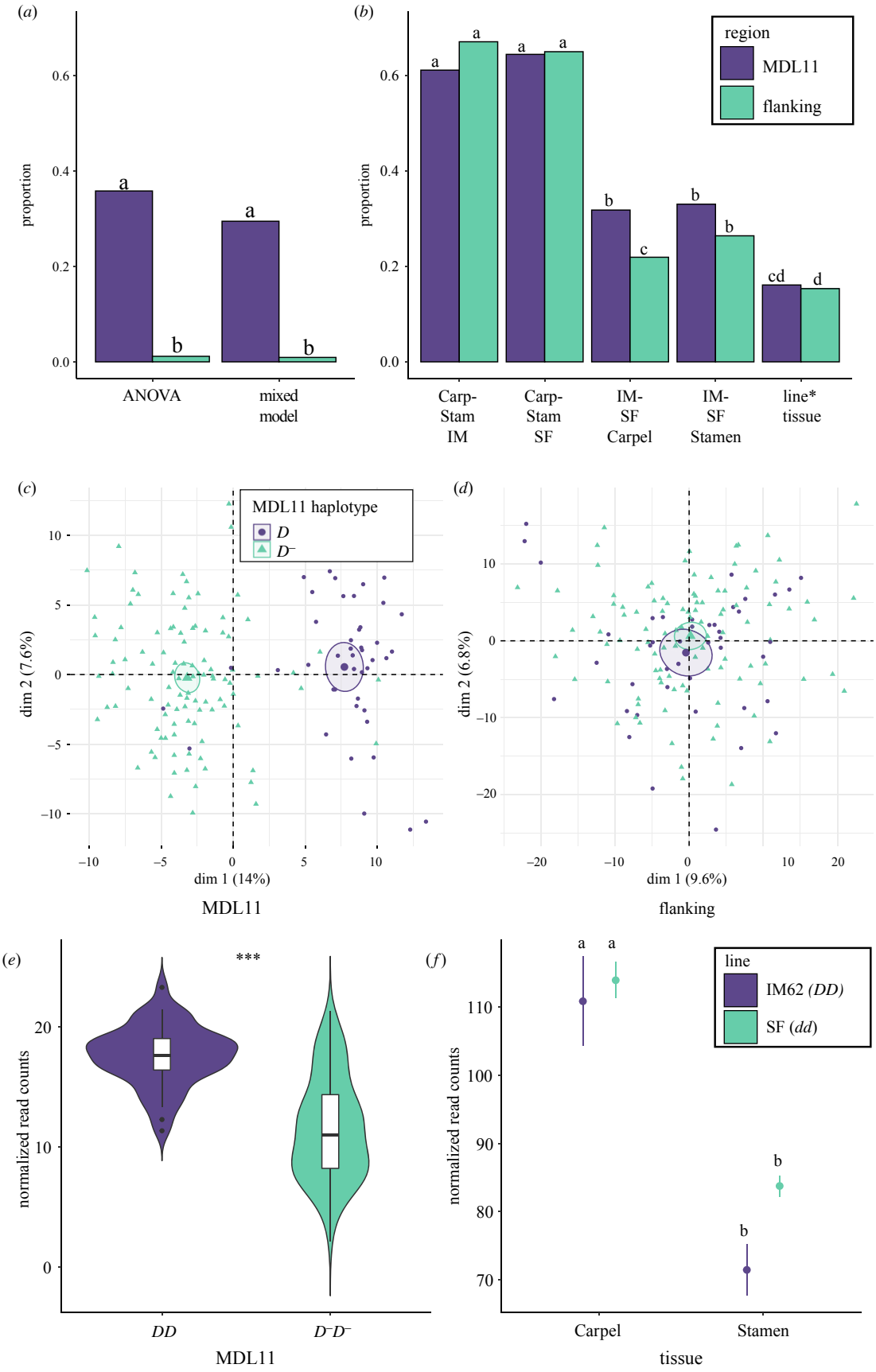

Figure 3. (*Caption continued.*)

Autoantigenic Sperm Protein (NASP-related). The *Arabidopsis* orthologue of NASP and *Schizosaccharomyces pombe* histone chaperone Sim3 (NASP^{SIM3}) binds the key kinetochore protein CenH3 [69]. Migut.K00842 is expressed at higher levels in *DD* versus $D^{-}D^{-}$ individuals within IM (ANOVA $F = 125.86$, $p < 0.0001$; figure 3*e*; electronic supplementary material, table S6). In the interspecific comparisons, Migut.K00842 is expressed at higher levels in the carpels relative to stamens of both species (IM $p = 0.00089$; SF $p = 0.045$), but it is not differentially expressed in IM versus SF (i.e. *D* vs

Figure 3. (*Overleaf.*) Striking differences in gene expression patterns for driving (*D*) and non-driving (D^- and *d*) haplotypes across the MDL11 region and in a candidate gene (Migut.K00842). (*a*) In developing bud transcriptomes of M. guttatus, a large proportion of genes are differentially expressed in *DD* versus D^-D^- individuals within the MDL11 region (purple), but almost no genes are differentially expressed in the flanking region of chromosome 11 (green). Results are similar whether considering models using ANOVAs or mixed models with line as a random effect ($N = 151$ IM lines). (*b*) The proportion of differentially expressed genes in comparisons between *M. guttatus* (IM62, *DD*) and *M. nasutus* (SF, *dd*) is significantly higher inside versus outside the MDL11 region in female tissues (IM-SF Carpel; $p < 0.05$) and marginally different in male tissues (IM-SF Stamen; $p \sim 0.06$; $N = 3$ replicates per treatment). Comparisons of male and female tissues within species and the interaction of species and tissue do not differ across the MDL11 region. (*c*,*d*) Principal components analyses based on gene expression within IM show clear clustering of driving (purple circles) and non-driving (green triangles) individuals (*c*) within the MDL11, but not (*d*) in the flanking regions of chromosome 11. Large points are means and ellipses are 95% confidence intervals around the mean. (*e*,*f*) Normalized read counts for Migut.K00842, a candidate gene that interacts with the centromere-defining histone CenH3, for (e) whole buds isolated from *DD* and D^-D^- IM lines ($N = 151$) and (*f*) developing carpel and stamen tissue isolated from IM62 and SF ($N = 3$ per treatment). In (*e*), boxplots show interquartile ranges and read counts were Box-Cox normalized (***$p < 0.0001$). In (*f*), points are means, error bars are +/− one standard deviation, and read counts represent counts per million following normalization by the trimmed mean of M-values method [62]. Results with unique letters in (*a*), (*b*), and (*f*) are significantly different (FDR adjusted *p*-value < 0.05). Legends for (*a*) and (*b*), as well as (*c*) and (*d*), are the same.

d) comparisons for either tissue (electronic supplementary material, table S7; figure 3*f*).

In addition to comparing patterns of differential expression broadly across the MDL11 region, we also parsed differentially expressed genes by whether they were up- or downregulated (electronic supplementary material, figure S1). We find two interesting patterns. First, more genes are upregulated than downregulated in the MDL11 region of developing bud transcriptomes ($D \times D^-$; $p < 0.0001$), but this is not true of carpel ($D \times d$; $p = 0.3336$) or stamen ($D \times d$; $p = 1$) tissues. The flanking region shows similar patterns, with the exception that carpel tissues go in the opposite direction with more genes down- versus upregulated (bud: $p = 0.038$; stamen: $p = 0.11$; carpel: $p < 0.0001$). Second, we observe more upregulated genes shared among all three tissues in the MDL11 than shared up- or downregulated genes in the flanking regions ($p = 0.01791$ and $p = 0.00078$, respectively). There was no difference in the proportion of shared genes up- versus downregulated within MDL11 or any other pairwise comparison ($p > 0.05$ in all cases). Eight genes were upregulated in the MDL11 region of drivers relative to non-drivers in all three comparisons (electronic supplementary material, table S9; figure S1*a*). Four of these genes also contributed significantly to dimension 1 of the within-IM PCA: Migut.K00780 (protein of unknown function, SEL-1 like protein), Migut.K00832 (methyl esterase 10), Migut.K00911 (LisH and RanBPM domains containing protein), and Migut.K01223 (galacturonosyltransferase 9). This comparison reinforces the finding that genes within the drive haplotype are primarily upregulated in drivers relative to non-drivers and points towards a small number of candidate genes that are upregulated in drivers in both intra- and inter-specific comparisons.

4. Discussion

Overall, our results demonstrate that a driving centromere in yellow monkeyflowers (*D*) has the potential to exhibit supergene behaviour across a wide range. In a neighbouring population (Cone Peak; CP) to the Oregon site in which it was initially discovered, *D*'s parallel selective sweep, intermediate frequency and fertility costs suggest that it may generally evolve as a balanced polymorphism, with the strength of drive offset by linked deleterious variation. However, evidence of shorter haplotypes, increased molecular variation, and distinct haplotype structure at CP suggest that *D* experiences somewhat different evolutionary dynamics at CP than at IM. This variation, along with detection of *D*-associated marker genotypes in a least five additional populations, sets the stage for genetic dissection of drive strength, resistance, and associated costs. Driving and non-driving haplotypes also show striking and consistent differences in gene expression profiles across diverse reproductive and developing tissues. These differences support the possibility that genic variation (as well as massive differences in centromeric DNA structure; [32]) contribute to drive or its linked costs, and refine the list of candidate genes underlying these phenomena.

The key evolutionary feature that makes supergenes so fascinating—suppression of recombination among numerous functional components—also creates barriers to dissecting their genetics. This challenge is particularly acute for selfish meiotic drivers, where strong environmentally independent selection may transiently extend haplotype structure even beyond the bounds of the supergene itself. For example, sex ratio drivers in *Drosophila* can exhibit continent-scale selective sweeps (and the evolution of suppression) over decadal timescales [70,71]. Nonetheless, some selfish supergenes have intrinsic costs that maintain them on the landscape as balanced polymorphisms [72], providing time for the accumulation of novel variation (including enhancers and deleterious variants) as well as partial homogenization (through gene conversion and/or recombination) with alternative haplotypes. Such variation within the classic SD sperm-killer system in *Drosophila* has been instrumental in identifying its essential functional components [73], as well as reconstructing the complex evolutionary history of its spread and maintenance [74]. Similarly, the neocentromeric Ab10-knob system in maize has diversified cytogenetically [75] in ways that appear to affect the balance of selfish and natural selection across altitudinal gradients [21]. Compared to these long-studied systems, our understanding of centromeric drive in *Mimulus* is in its infancy; however, the confirmation of drive dynamics and discovery of novel recombinational/mutational variation in one new population (CP), plus evidence for drive haplotypes in five others, is a key step toward both evolutionary genomic reconstruction and functional/genetic refinement.

The gene content variation and overall length differences within *D* haplotypes is particularly informative; although we cannot yet causally connect recombinational/length variation

to differences in costs or drive, such experiments are now feasible. The shorter *D* haplotype of CP24 excludes several potential functional candidate enhancers, most notably the cyclin Migut.K01228/1229, which was previously confirmed to be a *D*-specific duplicate of the unlinked gene Migut.J00575 (which is present in all *M. guttatus* assayed) and Mini-Chromosome Maintenance 2 (MCM2; Migut.K01043, which is within four genes of the end of the extended *D* haplotype found at IM) [37]. It is possible that these genic hitchhikers are unrelated, as they are clearly separable from the core *D* haplotype at CP. Intriguingly, despite apparent additivity of *D* fertility costs assayed at the core mK858 *D* marker at CP (figure 2*b*), we did not detect any significant association of the 45-gene indel genotype (assayed at the mK1229 *D*-dominant marker) with pollen inviability. Because it appears common (in 13% of genotyped wild individuals) and breaks off gene content variation novel to *D*, the shortest CP24-like haplotype is an ideal target for crossing experiments examining variation in the strength and costs of drive. In addition, population genomic analyses (which will also be facilitated by forthcoming long-read genome assemblies of *D* and D^- lines) can target individuals and populations variable in these regions. Although *D* has swept recently in both CP and IM, with profound effects of fitness variation, this work also suggests that it may have informative strata of genomic and functional variation, like other selfish supergenes.

Like other selfish supergenes, *D*'s spread has also profoundly affected patterns of gene expression, potentially contributing to its drive and costs. As in several other systems [76–79], transcriptome comparisons show an enrichment of differentially expressed genes between driving and non-driving haplotypes specifically in the region responsible for drive (figure 3*a–d*). This result is expected given known effects of linkage, low recombination and structural rearrangements on gene expression patterns [80], as well as the lack of common *trans* variants affecting gene expression at IM [45]. However, driving inversions need not only lead to divergence in gene expression patterns within the supergene itself and linked allelic variation is important for large-scale changes in gene expression [81]. In studies involving the *t*-haplotype of mice (which has extremely high fitness costs beyond its sperm-killing), there is some evidence that driving haplotypes have proportionally larger *trans* than *cis* effects on gene expression profiles. However, the relative importance of *cis* versus *trans* effects are sensitive to sampling scheme and much of the apparent large *trans* effect is due to duplication onto the *t*-haplotype [8,77,82]. In addition, *D* haplotypes show a consistent enrichment of upregulated genes relative to non-driving haplotypes in both interspecific and intra-population comparisons (electronic supplementary material, figures S1 and S2*c*). Similarly in *D. pseudoobscura*, proportionally more genes are upregulated in selfish (*SR*) relative to standard (*ST*) X-chromosomes [78]. Intriguingly, the *t* haplotype of mice displays conflicting patterns; *trans* effects of the *t*-haplotype were associated with upregulation, whereas *cis* effects resulted in an enrichment of downregulated genes [83]. Copy number variation (CNV) could also partially explain our results, as seen in the *t* haplotype of mice [77]. While removing the known 45-gene indel distinguishing *D* and D^- at IM results in qualitatively similar results, local mis-mapping or mis-assembly may obscure additional CNVs in the MDL11 region. Female meiotic drive is associated with genic CNV (as well as satellite DNA arrays) in diverse systems [16,84], and may also contribute in *Mimulus*.

Finally, while gene expression differences between driving and non-driving haplotypes do not necessarily cause the phenotypic effects of drive, they help refine the list of functional candidates for drive costs or enhancers. One particularly intriguing differentially expressed gene in this analysis is Migut.K00842, the sole *M. guttatus* homologue of the plant CENH3 chaperone $NASP^{SIM3}$ [69]. In addition to being the core component of the kinetochore predicted to coevolve with driving centromeres [15], CenH3A acts as a modifier QTL of *D* drive in heterospecific hybrids and shows evidence of a selective sweep following *D*'s spread through the IM population [37]. As one of the few well-characterized plant interactors with CenH3 [85], $NASP^{SIM3}$ is a strong *a priori* candidate for interacting as a linked modifier of drive or mediator of its deleterious effects. In the within-IM comparison, *D* lines showed strong elevated Migut.K00842 expression relative to (more diverse) D^- lines in young floral buds (figure 3*e*). This could be a plastic response to the cellular environment of cell division in *DD* homozygotes (which also show 20% male and female fertility costs), or it could reflect evolutionary divergence in *cis*-regulation, potentially under selection to enhance conditions for drive. These alternatives are not separable in inbred line comparisons but could be potentially ruled in or out in controlled crosses segregating for *D* (as any effects of Migut.K00842 expression on drive would need to manifest in heterozygotes). However, lack of strong differential expression for Migut.K00842 in IM *D* versus SF *d* also suggests that divergence in the expression of this and other MDL11 genes could also reflect selection for resistance on D^- haplotypes linked in repulsion with *D*; in that case, we would expect to see locus-specific signatures of selection and associated shifts in expression within the population of D^- genotypes. Alternatively, the relative expression of Migut.K00842 could be sensitive to developmental timing or tissue, with drive/non-drive differences emerging at later stages (whole floral buds in intraspecific comparisons versus developing stamens and carpels in interspecific ones). Two additional MDL11 genes annotated with cell cycle functions (Migut.K00775, a D2 cyclin and Migut.K0762, Cyclin-dependent kinase inhibitor 6) show intriguing differences in expression only in the interspecific comparison; both are more highly expressed in carpels than stamens, on average, and both exhibit strongly elevated expression in IM versus SF carpels (electronic supplementary material, table S7). Transcriptome sequencing across finer-scale reproductive tissues and developmental stages may, as in [79,83], help pinpoint strong candidates for meiotic differentiation among MDL11 chromosomal types. Furthermore, long-range genomic assemblies currently in progress, combined with joint examination of costs, drive strength and genic variation across diverse MDL11 haplotypes, will greatly facilitate the investigation of origin and functional effects of the supergene surrounding the centromeric driver *D*.

Data accessibility. Whole-genome sequence data from Cone Peak are available through the short read archive Biosample nos. SAMN12995508–SAMN12995518. RNA-Seq data were previously published and are available via SRA project no. PRJNA736440 and Biosample nos. SAMN13979871- SAMN13979871. All other data are provided in electronic supplementary material [86].

Authors' contributions. F.F.: conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, validation, visualization, writing—original draft, writing—review and editing; K.B.: formal analysis, methodology, writing—review and editing; A.D.: formal analysis, writing—review and editing; L.F.: conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, validation, visualization, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. Funding for the research was provided by NSF grant nos. OIA-1736249, DEB-1457763 and DEB-0846089 to L.F., as well as start-up funds to F.F.

Acknowledgements. We are grateful to G. Murgola and K. Baesen for laboratory and plant care assistance, A. Stathos and J. Colicchio for seed collections, and S.R. Miller, A.L. Sweigart and R. Kerwin for help with genomic and transcriptomic analyses. We would also like to thank the staff of the University of Montana Genomics Core for help with sequencing and genotyping, as well as the University of Montana's Environmental Control for Organismal Research facility for help with growing plants.

References

1. Thompson MJ, Jiggins CD. 2014 Supergenes and their role in evolution. *Heredity* **113**, 1–8. (doi:10.1038/hdy.2014.20)
2. Schwander T, Libbrecht R, Keller L. 2014 Supergenes and complex phenotypes. *Curr. Biol.* **24**, R288–R294. (doi:10.1016/j.cub.2014.01.056)
3. Gutiérrez-Valencia J, Hughes PW, Berdan EL, Slotte T. 2021 The genomic architecture and evolutionary fates of supergenes. *Genome Biol. Evol.* **13**, evab057. (doi:10.1093/gbe/evab057)
4. Huu CN, Keller B, Conti E, Kappel C, Lenhard M. 2020 Supergene evolution via stepwise duplications and neofunctionalization of a floral-organ identity gene. *Proc. Natl Acad. Sci. USA* **117**, 23 148–23 157. (doi:10.1073/pnas.2006296117)
5. Joron M *et al.* 2011 Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**, 203–U102. (doi:10.1038/nature10341)
6. Pearse DE *et al.* 2019 Sex-dependent dominance maintains migration supergene in rainbow trout. *Nat. Ecol. Evol.* **3**, 1731–1742. (doi:10.1038/s41559-019-1044-6)
7. Burt A, Trivers R. 2006 *Genes in conflict*. Cambridge, MA: Belknap Press.
8. Lindholm AK *et al.* 2016 The ecology and evolutionary dynamics of meiotic drive. *Trends Ecol. Evol.* **31**, 315–326. (doi:10.1016/j.tree.2016.02.001)
9. Bravo Núñez MA, Nuckolls NL, Zanders SE. 2018 Genetic villains: killer meiotic drivers. *Trends Genet.* **34**, 424–433. (doi:10.1016/j.tig.2018.02.003)
10. McLaughlin RN, Malik H. 2017 Genetic conflicts: the usual suspects and beyond. *J. Exp. Biol.* **220**, 6–17. (doi:10.1242/jeb.148148)
11. Zanders SE, Unckless RL. 2019 Fertility costs of meiotic drivers. *Curr. Biol.* **29**, R512–R520. (doi:10.1016/j.cub.2019.03.046)
12. Pardo-Manuel de Villena F, Sapienza C. 2001 Nonrandom segregation during meiosis: the unfairness of females. *Mamm. Genome* **12**, 331–339. (doi:10.1007/s003350040003)
13. Henikoff S, Malik H. 2002 Selfish drivers. *Nature* **417**, 227. (doi:10.1038/417227a)
14. Zwick ME, Salstrom JL, Langley CH. 1999 Genetic variation in rates of nondisjunction: associations of two naturally occurring polymorphisms in the chromokinesin nod with increased rates of nondisjunction in *Drosophila melanogaster*. *Genetics* **152**, 1605–1614. (doi:10.1093/genetics/152.4.1605)
15. Malik H, Henikoff S. 2002 Conflict begets complexity: the evolution of centromeres. *Curr. Opin. Genet. Dev.* **12**, 711–718. (doi:10.1016/s0959-437x(02)00351-9)
16. Dawe R *et al.* 2018 A Kinesin-14 motor activates neocentromeres to promote meiotic drive in maize. *Cell* **173**, 1–30. (doi:10.1016/j.cell.2018.03.009)
17. Swentowsky KW, Gent JI, Lowry EG, Schubert V, Ran X, Tseng K-F, Harkess AE, Qiu W, Dawe R. 2020 Distinct kinesin motors drive two types of maize neocentromeres. *Genes Dev.* **34**, 1239–1251. (doi:10.1101/gad.340679.120)
18. Dawe R, Hiatt EN. 2004 Plant neocentromeres: fast, focused, and driven. *Chromosome Res.* **12**, 655–669. (doi:10.1023/B:CHRO.0000036607.74671.db)
19. Higgins DM, Lowry EG, Kanizay LB, Becraft PW, Hall DW, Dawe R. 2018 Fitness costs and variation in transmission distortion associated with the abnormal chromosome 10 meiotic drive system in maize. *Genetics* **208**, 297–305. (doi:10.1534/genetics.117.300060)
20. Hall DW, Dawe R. 2018 Modeling the evolution of female meiotic drive in maize. *G3* **8**, 123–130. (doi:10.1534/g3.117.300073)
21. Bilinski P *et al.* 2018 Parallel altitudinal clines reveal trends in adaptive evolution of genome size in Zea mays. *PLoS Genet.* **14**, e1007162. (doi:10.1371/journal.pgen.1007162)
22. Buckler IV ES, Phelps-Durr TL, Buckler CSK, Dawe R, Doebley JF, Holtsford TP. 1999 Meiotic drive of chromosomal knobs reshaped the maize genome. *Genetics* **153**, 415–426. (doi:10.1093/genetics/153.1.415)
23. Chmátal L, Gabriel SI, Mitsainas GP, Martínez-Vargas J, Ventura J, Searle JB, Schultz RM, Lampson MA. 2014 Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in mice. *Curr. Biol.* **24**, 2295–2300. (doi:10.1016/j.cub.2014.08.017)
24. Akera T, Chmátal L, Trimm E, Yang K, Aonbangkhen C, Chenoweth DM, Janke C, Schultz RM, Lampson MA. 2017 Spindle asymmetry drives non-Mendelian chromosome segregation. *Science* **358**, 668–672. (doi:10.1126/science.aan0092)
25. Iwata-Otsubo A *et al.* 2017 Expanded satellite repeats amplify a discrete CENP-A nucleosome assembly site on chromosomes that drive in female meiosis. *Curr. Biol.* **27**, 2365–2373.e8. (doi:10.1016/j.cub.2017.06.069)
26. Akera T, Trimm E, Lampson MA. 2019 Molecular strategies of meiotic cheating by selfish centromeres. *Cell* **178**, 1132–1144.e10. (doi:10.1016/j.cell.2019.07.001)
27. Kumon T, Ma J, Akins RB, Stefanik D, Nordgren CE, Kim J, Levine MT, Lampson MA. 2021 Parallel pathways for recruiting effector proteins determine centromere drive and suppression. *Cell* **184**, 4904–4918.e11. (doi:10.1016/j.cell.2021.07.037)
28. Pardo-Manuel de Villena F, Sapienza C. 2001 Female meiosis drives karyotypic evolution in mammals. *Genetics* **159**, 1179–1189. (doi:10.1093/genetics/159.3.1179)
29. Thomson GJ, Feldman MW. 1974 Population genetics of modifiers of meiotic drive. II. Linkage modification in the segregation distortion system. *Theor. Popul. Biol.* **5**, 155–162. (doi:10.1016/0040-5809(74)90038-0)
30. Berdan EL, Blanckaert A, Butlin RK, Bank C. 2021 Deleterious mutation accumulation and the long-term fate of chromosomal inversions. *PLoS Genet.* **17**, e1009411. (doi:10.1371/journal.pgen.1009411)
31. Fishman L, Willis JH. 2005 A novel meiotic drive locus almost completely distorts segregation in *Mimulus* (monkeyflower) hybrids. *Genetics* **169**, 347–353. (doi:10.1534/genetics.104.032789)
32. Fishman L, Saunders A. 2008 Centromere-associated female meiotic drive entails male fitness costs in monkeyflowers. *Science* **322**, 1559–1562. (doi:10.1126/science.1161406)
33. Malik H. 2005 Mimulus finds centromeres in the driver's seat. *Trends Ecol. Evol.* **20**, 151–154. (doi:10.1016/j.tree.2005.01.014)
34. Fishman L, Kelly JK. 2015 Centromere-associated meiotic drive and female fitness variation in Mimulus. *Evolution* **69**, 1208–1218. (doi:10.1111/evo.12661)
35. Scoville AG, Lee YW, Willis JH, Kelly JK. 2009 Contribution of chromosomal polymorphisms to the G-matrix of Mimulus guttatus. *New Phytol.* **183**, 803–815. (doi:10.1111/j.1469-8137.2009.02947.x)
36. Flagel LE, Blackman BK, Fishman L, Monnahan PJ, Sweigart A, Kelly JK. 2019 GOOGA: a platform to synthesize mapping experiments and identify genomic structural diversity. *PLoS Comput. Biol.* **15**, e1006949. (doi:10.1371/journal.pcbi.1006949)

37. Finseth FR, Nelson TC, Fishman L. 2021 Selfish chromosomal drive shapes recent centromeric histone evolution in monkeyflowers. *PLoS Genet.* **17**, e1009418. (doi:10.1371/journal.pgen.1009418)
38. Finseth FR, Dong Y, Saunders A, Fishman L. 2015 Duplication and adaptive evolution of a key centromeric protein in mimulus, a genus with female meiotic drive. *Mol. Biol. Evol.* **32**, 2694–2706. (doi:10.1093/molbev/msv145)
39. Case AL, Finseth FR, Barr CM, Fishman L. 2016 Selfish evolution of cytonuclear hybrid incompatibility in *Mimulus*. *Proc. R. Soc. B* **283**, 20161493. (doi:10.1098/rspb.2016.1493)
40. Puzey JR, Willis JH, Kelly JK. 2017 Population structure and local selection yield high genomic variation in *Mimulus guttatus*. *Mol. Ecol.* **26**, 519–535. (doi:10.1111/mec.13922)
41. Monnahan PJ, Colicchio J, Kelly JK. 2015 A genomic selection component analysis characterizes migration-selection balance. *Evolution* **69**, 1713–1727. (doi:10.1111/evo.12698)
42. Colicchio J, Monnahan PJ, Wessinger CA, Brown K, Kern JR, Kelly JK. 2020 Individualized mating system estimation using genomic data. *Mol. Ecol. Resour.* **20**, 333–347. (doi:10.1111/1755-0998.13094)
43. Colicchio J. 2017 Transgenerational effects alter plant defence and resistance in nature. *J. Evol. Biol.* **30**, 664–680. (doi:10.1111/jeb.13042)
44. Kerwin RE, Sweigart AL. 2020 Rampant misexpression in a *Mimulus* (monkeyflower) introgression line caused by hybrid sterility, not regulatory divergence. *Mol. Biol. Evol.* **31**, 166–2098. (doi:10.1093/molbev/msaa071)
45. Brown KE, Kelly JK. 2021 Genome-wide association mapping of transcriptome variation in *Mimulus guttatus* indicates differing patterns of selection on cis- versus trans-acting mutations. *Genetics* **220**, iyab189. (doi:10.1093/genetics/iyab189)
46. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. (doi:10.1093/bioinformatics/btu170)
47. Li H, Durbin R. 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760. (doi:10.1093/bioinformatics/btp324)
48. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009 The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079. (doi:10.1093/bioinformatics/btp352)
49. McKenna A *et al.* 2010 The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303. (doi:10.1101/gr.107524.110)
50. DePristo MA *et al.* 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498. (doi:10.1038/ng.806)
51. Danecek P *et al.* 2011 The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158. (doi:10.1093/bioinformatics/btr330)
52. Huson DH, Bryant D. 2006 Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267. (doi:10.1093/molbev/msj030)
53. Fishman L, Willis JH. 2001 Evidence for Dobzhansky-Muller incompatibilities contributing to the sterility of hybrids between *Mimulus guttatus* and *M. nasutus*. *Evolution* **55**, 1932–1942. (doi:10.1111/j.0014-3820.2001.tb01311.x)
54. Troth A, Puzey JR, Kim RS, Willis JH, Kelly JK. 2018 Selective trade-offs maintain alleles underpinning complex trait variation in plants. *Science* **361**, 475–478. (doi:10.1126/science.aat5760)
55. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013 STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21. (doi:10.1093/bioinformatics/bts635)
56. Anders S, Pyl PT, Huber W. 2014 HTSeq—a Python framework to work with high throughput sequencing data. *Bioinformatics* **31**, 166–169. (doi:10.1093/bioinformatics/btu638)
57. Love MI, Huber W, Anders S. 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. (doi:10.1186/s13059-014-0550-8)
58. Bates D, Mächler M, Bolker B, Walker S. 2014 Fitting linear mixed-effects models using lme4. *J. Stat. Soft.* **67**, 1–48. (doi:10.18637/jss.v067.i01)
59. Kassambara A, Mundt F. 2016 Factoextra: extract and visualize the results of multivariate data analyses. *R package version* **1**, 337–354.
60. Brandvain Y, Kenney AM, Flagel L, Coop G, Sweigart AL. 2014 Speciation and introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PLoS Genet.* **10**, e1004410. (doi:10.1371/journal.pgen.1004410)
61. Poplin, R. *et al*. 2018 Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*. (doi:10.1101/201178)
62. Robinson MD, McCarthy DJ, Smyth GK. 2009 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140. (doi:10.1093/bioinformatics/btp616)
63. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015 limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47. (doi:10.1093/nar/gkv007)
64. R Core Team. 2021 *R: A language and environment for statistical computing*.
65. RStudio Team. 2020 RStudio | Open source & professional software for data science teams.
66. Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW. 2000 Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc. Natl Acad. Sci. USA* **97**, 7360–7365. (doi:10.1073/pnas.97.13.7360)
67. Saze H, Shiraishi A, Miura A, Kakutani T. 2008 Control of genic DNA methylation by a jmjC domain-containing protein in *Arabidopsis thaliana*. *Science* **319**, 462–465. (doi:10.1126/science.1150987)
68. Colicchio JM, Hamm LN, Verdonk HE, Kooyers NJ, Blackman BK. 2021 Adaptive and nonadaptive causes of heterogeneity in genetic differentiation across the *Mimulus guttatus* genome. *Mol. Ecol.* **30**, 6486–6507. (doi:10.1111/mec.16087)
69. Le Goff S *et al.* 2020 The H3 histone chaperone NASPSIM3 escorts CenH3 in Arabidopsis. *Plant J.* **101**, 71–86. (doi:10.1111/tpj.14518)
70. Bastide H, Cazemajor M, Ogereau D, Derome N, Hospital F, Montchamp-Moreau C. 2011 Rapid rise and fall of selfish sex-ratio X chromosomes in *Drosophila simulans*: spatiotemporal analysis of phenotypic and molecular data. *Mol. Biol. Evol.* **28**, 2461–2470. (doi:10.1093/molbev/msr074)
71. Dyer KA, Charlesworth B, Jaenike J. 2007 Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. *Proc. Natl Acad. Sci. USA* **104**, 1587–1592. (doi:10.1073/pnas.0605578104)
72. Price TAR, Verspoor R, Wedell N. 2019 Ancient gene drives: an evolutionary paradox. *Proc. R. Soc. B.* **286**, 20192267. (doi: 10.1098/rspb.2019.2267)
73. Brand CL, Larracuente AM, Presgraves DC. 2015 Origin, evolution, and population genetics of the selfish Segregation Distorter gene duplication in European and African populations of *Drosophila melanogaster*. *Evolution* **69**, 1271–1283. (doi:10.1111/evo.12658)
74. Navarro-Dominguez B, Chang CH, Brand CL, Muirhead CA, Presgraves DC, Larracuente AM. 2022 Epistatic selection on a selfish Segregation Distorter supergene: drive, recombination, and genetic load. *biorxiv.org*. (doi:10.1101/2021.12.22.473781)
75. Kanizay LB, Pyhäjärvi T, Lowry EG, Hufford MB, Peterson DG, Ross-Ibarra J, Dawe R. 2013 Diversity and abundance of the abnormal chromosome 10 meiotic drive complex in Zea mays. *Heredity* **110**, 570–577. (doi:10.1038/hdy.2013.2)
76. Reinhardt JA, Brand CL, Paczolt KA, Johns PM, Baker RH, Wilkinson GS. 2014 Meiotic drive impacts expression and evolution of x-linked genes in stalk-eyed flies. *PLoS Genet.* **10**, e1004362. (doi:10.1371/journal.pgen.1004362)
77. Kelemen RK, Vicoso B. 2018 Complex history and differentiation patterns of the t-haplotype, a mouse meiotic driver. *Genetics* **208**, 365–375. (doi:10.1534/genetics.117.300513)
78. Fuller ZL, Koury SA, Leonard CJ, Young RE, Ikegami K, Westlake J, Richards S, Schaeffer SW, Phadnis N. 2020 Extensive recombination suppression and epistatic selection causes chromosome-wide differentiation of a selfish sex chromosome in *Drosophila pseudoobscura*. *Genetics, Genetics.* **216**, 205–226. (doi:10.1534/genetics.120.303460)
79. Pieper KE, Unckless RL, Dyer KA. 2018 A fast-evolving X-linked duplicate of importin-α2 is overexpressed in sex-ratio drive in *Drosophila neotestacea*. *Mol. Ecol.* **27**, 5165–5179. (doi:10.1111/mec.14928)

80. Huang W, Carbone MA, Magwire MM, Peiffer JA, Lyman RF, Stone EA, Anholt RRH, Mackay TFC. 2015 Genetic basis of transcriptome diversity in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **112**, E6010–9. (doi:10.1073/pnas.1519159112)
81. Said I, Byrne A, Serrano V, Cardeno C, Vollmers C, Corbett-Detig R. 2018 Linked genetic variation and not genome structure causes widespread differential expression associated with chromosomal inversions. *Proc. Natl Acad. Sci. USA* **115**, 5492–5497. (doi:10.1073/pnas.1721275115)
82. Kelemen RK, Elkrewi M, Lindholm AK, Vicoso B. 2022 Novel patterns of expression and recruitment of new genes on the t-haplotype, a mouse selfish chromosome. *Proc. Biol. Sci.* **289**, 20211985. (doi:10.1098/rspb.2021.1985)
83. Lindholm A, Sutter A, Künzel S, Tautz D, Rehrauer H. 2019 Effects of a male meiotic driver on male and female transcriptomes in the house mouse. *Proc. Biol. Sci.* **286**, 20191927. (doi:10.1098/rspb.2019.1927)
84. Didion JP *et al.* 2015 A multi-megabase copy number gain causes maternal transmission ratio distortion on mouse chromosome 2. *PLoS Genet.* **11**, e1004850. (doi:10.1371/journal.pgen.1004850)
85. Keçeli BN, Jin C, Van Damme D, Geelen D. 2020 Conservation of centromeric histone 3 interaction partners in plants. *J. Exp. Bot.* **71**, 5237–5246. (doi:10.1093/jxb/eraa214)
86. Finseth F, Brown K, Demaree A, Fishman L. 2022 Supergene potential of a selfish centromere. Figshare. (doi:10.6084/m9.figshare.c.5980562)