Molecular Ecology Resources

WILEY

MOLECULAR ECOLOGY RESOURCES

RESOURCE ARTICLE OPEN ACCESS

Comparative Analyses of Four Reference Genomes Reveal Exceptional Diversity and Weak Linked Selection in the Yellow Monkeyflower (*Mimulus guttatus*) Complex

John T. Lovell[1,2] | Rachel Walstead[1] | Amelia Lawrence[3] | Evan Stark-Dykema[4] | Matthew C. Farnitano[5] | Avril Harder[1] | Tomáš Brůna[2] | Kerrie Barry[2] | David Goodstein[2] | Jerry Jenkins[1] | Anna Lipzen[2] | LoriBeth Boston[1] | Jenell Webber[1] | Mansi Chovatia[2] | Joanne Eichenberger[2] | Jayson Talag[6] | Jane Grimwood[1] | Jeremy Schmutz[1,2] | John K. Kelly[7] | Andrea L. Sweigart[5] | Lila Fishman[4] | John H. Willis[3]

[1]Genome Sequencing Center, HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA | [2]US Department of Energy Joint Genome Institute, Berkeley, California, USA | [3]Department of Biology, Duke University, Durham, North Carolina, USA | [4]Division of Biological Sciences, University of Montana, Missoula, Montana, USA | [5]Department of Genetics, University of Georgia, Athens, Georgia, USA | [6]Arizona Genomics Institute, University of Arizona, Tucson, Arizona, USA | [7]Department of Ecology & Evolution, University of Kansas, Lawrence, Kansas, USA

Correspondence: John T. Lovell (jlovell@hudsonalpha.org) | Lila Fishman (lila.fishman@mso.umt.edu) | John H. Willis (jwillis@duke.edu)

Received: 16 April 2025 | **Revised:** 10 June 2025 | **Accepted:** 1 July 2025

Handling Editor: Joanna Kelley

Funding: The work (proposal: 10.46936/10.25585/60001364) conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a Division of Environmental Biology No. DEB-1856180 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.

Keywords: chromosomal evolution | linked selection | *Mimulus* | nucleotide diversity | recombination | reference genomes

ABSTRACT

Yellow monkeyflowers (*Mimulus guttatus* complex, Phrymaceae) are a powerful system for studying ecological adaptation, reproductive variation, and genome evolution. To initiate pan-genomics in this group, we present four chromosome-scale assemblies and annotations of accessions spanning a broad evolutionary spectrum: two from a single *M. guttatus* population, one from the closely related selfing species *M. nasutus*, and one from a more divergent species *M. tilingii*. All assemblies are highly complete and resolve centromeric and repetitive regions. Comparative analyses reveal such extensive structural variation in repeat-rich, gene-poor regions that large portions of the genome are unalignable across accessions. As a result, this *Mimulus* pan-genome is primarily informative in genic regions, underscoring limitations of resequencing approaches in such polymorphic taxa. We document gene presence–absence, investigate the recombination landscape using high-resolution linkage data, and quantify nucleotide diversity. Surprisingly, pairwise differences at fourfold synonymous sites are exceptionally high—even in regions of very low recombination—reaching ~3.2% within a single *M. guttatus* population, ~7% within the interfertile *M. guttatus* species complex (approximately equal to SNP divergence between great apes and Old World monkeys), and ~7.4% between that complex and the reproductively isolated *M. tilingii*. Genome-wide patterns of nucleotide variation show little evidence of linked selection, and instead suggest that the concentration of genes (and likely selected sites) in high-recombination regions may buffer diversity loss. These assemblies, annotations, and comparative analyses provide a robust genomic foundation for *Mimulus* research and offer new insights into the interplay of recombination, structural variation, and molecular evolution in highly diverse plant genomes.

John K. Kelly, Andrea L. Sweigart, Lila Fishman and John H. Willis contributed equally to this work.

1 | Introduction

The yellow monkeyflowers (*Mimulus* section *Simiolus*; Phrymaceae) are a diverse group of wildflowers centred in western North America that exhibit remarkable ecological, phenotypic, and genomic variation. These plants occupy environments ranging from serpentine soils and mine tailings to geothermal crusts and coastal salt spray, and span elevations from sea-level dunes to subalpine meadows (Vickery 1978; Wu et al. 2008). This adaptive breadth is accompanied by striking variation in life history and mating system, from bee-pollinated outcrossing perennials to highly selfing annuals (Fishman et al. 2002; Friedman and Willis 2013; Kolis et al. 2022; Lowry and Willis 2010; Troth et al. 2018). As such, the group has become a powerful model for studying the genetic basis of adaptation, reproductive isolation, and genome evolution.

A key component of this system is the *Mimulus guttatus* species complex—a taxonomically and biologically diverse assemblage of partially interfertile lineages that includes *M. guttatus*, *M. nasutus*, and numerous other ecologically and morphologically divergent taxa (Nesom 2012; Vickery 1978). While traditionally treated as distinct taxonomic species or subspecies, population genomic studies have revealed extensive shared variation among members of the complex due to both ancestral polymorphism and a history of gene flow (Brandvain et al. 2014; Ivey et al. 2023; Sweigart and Willis 2003; Twyford et al. 2020). Against this general backdrop of historical and ongoing migration, genetic diversity in the flagship species, *M. guttatus*, clusters into two broad groups that roughly correspond to the northern and southern parts of the species range (Brandvain et al. 2014). The well-studied 'Iron Mountain' population of *M. guttatus* (Troth et al. 2018; Willis 1993) clusters with other 'northern' samples, whereas *M. nasutus*, despite its highly selfing mating system and distinctive morphology, is nested entirely within the 'southern' genetic cluster (Brandvain et al. 2014). *Mimulus tilingii*, by contrast, is strongly reproductively isolated from members of the *M. guttatus* complex, primarily due to hybrid seed lethality (Garner et al. 2016; Sandstedt et al. 2021), and is typically considered a distinct biological species. Nevertheless, it shares a relatively recent common ancestor with the complex (Sandstedt et al. 2021). Interestingly, genome-wide divergence between *M. tilingii* and *M. guttatus* is only modestly higher than that observed between some members of the *M. guttatus* complex itself (Brandvain et al. 2014; Sandstedt et al. 2021), underscoring the continuum of variation and blurred species boundaries that characterise this group.

These genomic relationships and the evolutionary richness of *Mimulus* have motivated a larger collaborative effort to generate high-quality reference genomes across the genus, with particular focus on the *M. guttatus* species complex. This comparative framework is designed to capture the full spectrum of genomic variation—both structural and sequence-level—within and among lineages, providing a foundation for evolutionary and ecological genomics across the group. In this paper, we present the first four genome assemblies generated as part of this effort: two from the highly polymorphic Iron Mountain (IM) population of *M. guttatus* (IM62 and IM767), one from the closely related selfing species *M. nasutus* (SF), and one from the more divergent outgroup *M. tilingii* (LVR; Figure 1). Together, these accessions span a broad evolutionary gradient within the *M. guttatus* complex and its relatives, enabling analyses of within-population diversity, species-level divergence, and patterns of recombination and selection across genomes.

The IM population has played a central role in evolutionary genetic studies of *Mimulus*, serving as a model for research on inbreeding depression, mating system evolution, complex trait variation, local adaptation, and chromosomal polymorphism (Kelly 2022; Troth et al. 2018; Veltsos and Kelly 2024; Willis 1993, 1999). In addition to abundant adaptive variation, the IM population segregates for a costly centromeric driver that distorts the transmission of Chromosome 11 through female meiosis (Finseth et al. 2022, 2021; Fishman and Saunders 2008). Our focal genotypes IM62 and IM767 carry functionally distinct MDL11 haplotypes ('driver' and 'resistant', respectively), while SF *M. nasutus* line harbours a third highly 'susceptible' haplotype. Representation of these variants in newly assembled centromeric and pericentromeric regions offers a unique opportunity to study the genomic basis of centromere drive.

Despite the ecological and evolutionary richness of monkeyflowers, genomic studies in the system have been limited by the incompleteness of existing reference assemblies. The widely used *M. guttatus* v2.0 reference (Hellsten et al. 2013), based on IM62, lacks substantial pericentromeric and repetitive sequence and includes mis-assemblies that complicate downstream analyses (Flagel et al. 2019). Moreover, the extreme sequence and structural variation across *Mimulus* genomes presents challenges for alignment, annotation, and comparative analysis.

Here, we address these limitations by generating and analysing chromosome-scale reference genomes for IM62 and IM767 *M. guttatus*, SF *M. nasutus*, and LVR *M. tilingii*. We begin by evaluating improvements in contiguity, completeness, and annotation in the new IM62 assembly relative to v2.0. We then use whole-genome alignments and pan-genome graphs to assess structural and gene content variation across the four genomes. Using high-resolution linkage data, we characterise recombination landscapes in relation to chromosome structure and gene density. Finally, we compare patterns of nucleotide diversity at fourfold degenerate sites across the recombination landscape to test for evidence of linked selection. Together, these assemblies and analyses provide a foundational resource for studying adaptation, conflict, and genome evolution in this exceptionally diverse system.

2 | Methods

2.1 | Biological Materials

We derived the *Mimulus guttatus* 'IM62' and 'IM767' inbred lines through more than 15 generations of single-seed descent from unrelated maternal seed families collected at Iron Mountain, Oregon, USA (44.402217° N, 122.153317° W). See Puzey et al. (2017), Troth et al. (2018), and Willis (1999) for additional details on the IM population and inbred line formation. The *Mimulus nasutus* 'SF' line originated from a naturally inbreeding population at Sherars Falls, Oregon (45.257092744° N, 121.03939758° W), and was inbred through five generations of

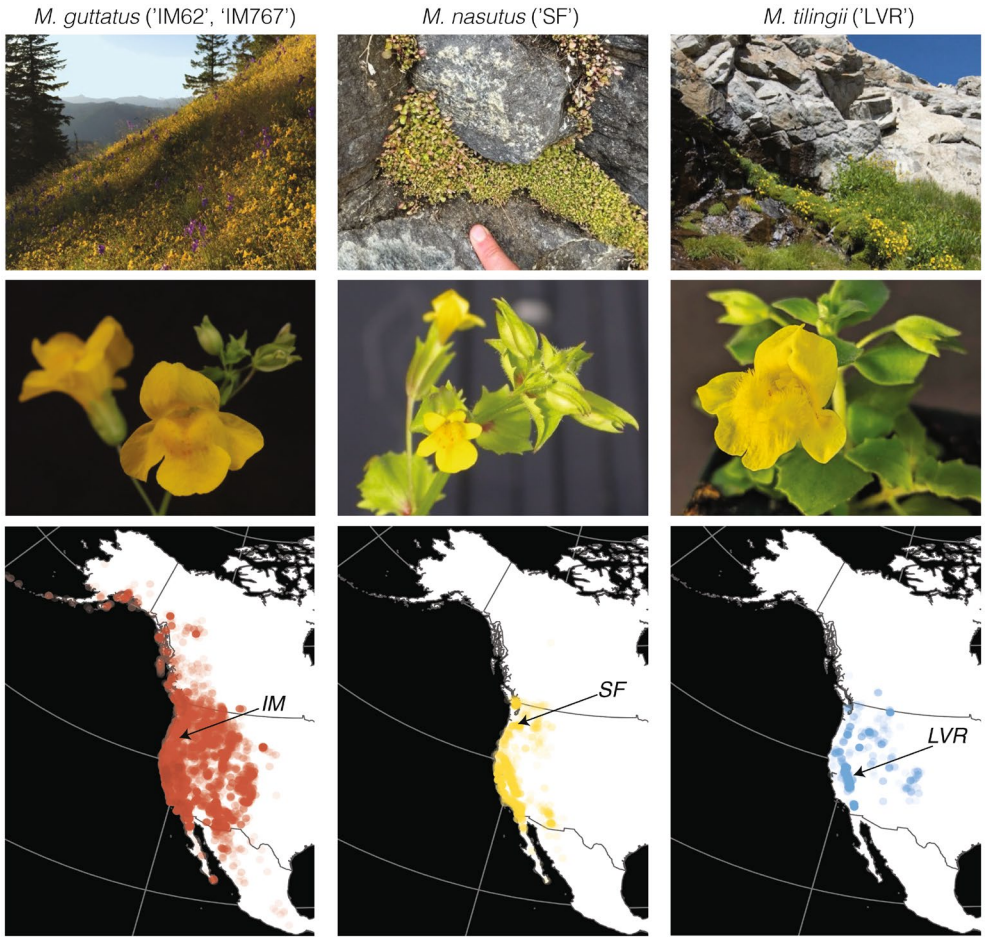

FIGURE 1 | Geographic and phenotypic diversity of yellow monkeyflowers. The three focal species are morphologically and ecologically similar, broadly sympatric, and generally interfertile. Images are courtesy of Dena Grossenbacher (top right), Natalie Gonzalez (bottom right), and herbarium record (downloaded from GBIF) collection locales are displayed on a map of western North America.

single-seed descent (Fishman et al. 2001), followed by over 15 generations of selfing. The *Mimulus tilingii* 'LVR' line was generated through at least eight generations of single-seed descent from a wild seed collected near Lee Vining, California, USA (37.950817° N, 119.225733° W; Garner et al. 2016). All lines are highly inbred (as confirmed by sequencing), and we collected tissues for DNA and RNA extraction from multiple individuals from the same generation.

2.2 | Nucleic Acid Library Preparation

We extracted high-molecular-weight DNA from 5 to 10 g of flash-frozen floral bud tissue at the Arizona Genomics Institute using a modified CTAB-chloroform protocol (Doyle and Doyle 1987). We ground tissues in liquid nitrogen, incubated them at 50°C in 2% CTAB buffer containing proteinase K, PVP-40, and beta-mercaptoethanol for 30–60 min, and extracted DNA with two rounds of 24:1 chloroform: isoamyl alcohol. After adding 1/10 volume of 3 M sodium acetate, we precipitated DNA with isopropanol, centrifuged to collect the DNA, washed it with 70% ethanol, air-dried it, and resuspended it in 10 mM Tris buffer. We treated the DNA with RNase (BioBasic; Markham, ON, Canada) and further cleaned it with magnetic beads. We assessed purity with a Nanodrop, measured DNA concentration with a Qubit HS kit (ThermoFisher Scientific), and validated fragment size with the Agilent Femto Pulse System.

To support protein-coding gene annotation, we extracted RNA from 6 to 8 flash-frozen tissue types per genotype, including young floral bud, open flower, seedling, young leaf, ovary, root tip, and inflorescence stem, with 1–3 biological replicates per tissue. We used Spectrum Total Plant RNA kits (Sigma-Aldrich) and assessed RNA quality with a Nanodrop and Bioanalyzer (Agilent Technologies). Appendix S1 lists all tissues.

2.3 | DNA Sequencing and Genome Assembly

We prepared PacBio, Omni-C, and Illumina libraries at HudsonAlpha using the SMRTbell Prep Kit v3.0 (PacBio), Dovetail Omni-C Kit (Cantata Bio), and TruSeq DNA PCR-free Kit (Illumina), respectively. We sequenced PacBio HiFi reads on SEQUEL II (SF and LVR) and REVIO (IM62 and IM767) platforms, and performed Hi-C and Illumina sequencing on an Illumina NovaSeq. Coverage statistics are in Table 1.

We assembled initial contigs using HiFiAsm+HiC v0.16.1 (Cheng et al. 2021) and polished them with RACON v1.4.10

TABLE 1 | Comparison of v2 (2013) and our 2024 genome assemblies and annotations. Statistics taken from phytozome, calculated from exact k-mer matching, or intersections between gene gff3 files. Rows below the middle horizontal line present assembly and sequencing methodological statistics, including read depth and polishing for the four genomes.

	IM62 (v2)	IM62 (v3)	IM767 (v2)	*M. nasutus* ('sf')	*M. tilingii* ('lvr')
Main genome size (Mb)	312.7	339.7	314.6	312.9	315.38
Unplaced scaffold size (Mb)	19.1	0.68	1.33	0	0
Main genome contig N50 (Mb)	0.047	6.1	10.2	6.7	15.4
Genes (CDS, Mb)	33.00	32.18	31.86	32.19	32.92
Centromeric repeats (Mb)	8.08	41.32	31.83	27.13	29.47
HiFi coverage (X)		70.02	88.35	49.39	85.45
HiFi mean length (bp)		13,991	15,194	16,115	14,655
Omni-C coverage (X)		71.7	133.2	77.7	84.1
Polishing coverage (X)		41.8	47.8	46.7	48.1
Polishable bases (Mb)		341.08	285.12	311.38	297.00
Corrected homozygous SNPs		57	64	7	7
Corrected homozygous INDELS		1360	800	766	506
Heterozygous positions (bp)		5534	15,868	1798	3173

(Vaser et al. 2017). Given the homozygosity of all four lines (Table 1), we represented each genome as a haploid by collapsing adjacent alternative haplotypes using the longest common substring matches. We ordered and oriented contigs using the JUICER v1.8.8 pipeline (Durand et al. 2016), ensuring proper orientation of telomeric sequences. Chromosomes were numbered and oriented to match the IM62 v2 genome (https://phytozome-next.jgi.doe.gov/info/Mguttatus_v2_0). We then mapped Illumina reads to each genome with BWA-MEM v2.2.1 (Li and Durbin 2009) and corrected homozygous sequencing errors using GATK v3.7 UnifiedGenotyper (McKenna et al. 2010).

2.4 | Transcriptome Sequencing

We sequenced stranded paired-end (2×150 bp) RNA-seq libraries on an Illumina NovaSeq and equimolar RNA pools from all tissues on a PacBio Sequel II using Iso-Seq. Illumina libraries were prepared with the TruSeq Poly-A Stranded mRNA HT Kit (Illumina) using 8 amplification cycles and quantified by qPCR (Kapa Biosystems) on a Roche LightCycler 480. Iso-Seq libraries were generated from 500 ng total RNA using NEBNext Single Cell/Low Input cDNA Synthesis (New England Biolabs), PCR-amplified (12–20 cycles), purified with AMPure PB Beads, ligated with barcoded SMRTbell adapters, and sequenced.

We assembled transcriptomes using genome-guided short-read assembly via GSNAP v2019-09-12 (Wu and Nacu 2010) and full-length transcript correction and collapsing from Iso-Seq data using GMAP v2019-09-12 (Wu and Watanabe 2005) and PASA v2.0.2 (Haas et al. 2003). PASA merged short-read and long-read transcript assemblies to generate a comprehensive transcriptome.

2.5 | Genome Annotation

We annotated protein-coding loci in repeat-soft-masked genomes using a dual-alignment strategy. We generated repeat libraries de novo with RepeatModeler2 v2.0.4 (Flynn et al. 2020) and masked them with RepeatMasker. We screened predicted repeats with InterProScan v5.51-85.0 (Jones et al. 2014) and removed those overlapping protein-coding domains.

Gene models were predicted using PERTRAN (Lovell et al. 2018), which aligned ESTs (via EXONERATE v2.4.0) and proteins from 20 diverse plant species. We predicted gene models using FGENESH+, FGENESH_EST, EXONERATE, PASA-ORF, and AUGUSTUS v3.3.3 (Stanke et al. 2006), trained on high-confidence PASA models. We selected the best-supported models using positive criteria (EST and protein support) and a negative factor (repeat overlap). PASA refined gene models, adding UTRs, correcting splicing, and incorporating alternative transcripts. Final models were filtered based on C-score and homology coverage. We removed low-confidence models that lacked start/stop codons, had poor support, or were repetitive with weak homology.

2.6 | Linkage Analysis of Recombination Landscape

To examine how meiotic recombination varies across the genome, we used high-density linkage maps produced by Veltsos and Kelly (2024), generated from crosses of IM767 with nine unrelated Iron Mountain inbred lines. These F2 populations ($n = 1373$ individuals) yielded 33,302 crossover events across 2746 meiotic products, localised using RNA-seq genotyping (Veltsos and Kelly 2024). We translated marker coordinates from an earlier IM767 assembly to the new IM767 v2.1 genome

and approximated crossover locations using the midpoints of marker intervals. The positions of crossovers that could not be refined to a 2 Mb interval or less were excluded from downstream analyses. We interpolated genetic map positions to estimate recombination rates in centiMorgans (cM) and generated a genome-wide linkage map from the crossover locations in R/qtl (Broman et al. 2003), which permitted estimation of the map position for each protein-coding gene in the IM767 v2.1 annotation. Since the recombination rate is determined as the slope of a linear interpolation where the number of crossover events is a function of the physical distance between two adjacent genes, the density of genes has no effect on the accuracy of the recombination rate estimate itself, just the precision of the physical position of that estimate.

To identify genomic regions with extremely low-recombination rates, we estimated recombination rates in non-overlapping 500 kb windows across each chromosome. We defined low-recombining "pericentromeric" regions as stretches of two or more consecutive windows with recombination rates < 1 cM/Mb. In some cases, isolated central windows had > 1 cM/Mb but were flanked by low-recombination regions with large inter-marker intervals (> 2 Mb), suggesting poor localisation; we still classified those as pericentromeric. Conversely, we excluded edge-adjacent windows with > 1 cM/Mb if they had short marker intervals (< 2 Mb), which suggested more accurate crossover localisation.

2.7 | Single Nucleotide Differences Between Reference Genomes

To assess nucleotide-level variation among the four reference genomes, we focused on orthologous coding sites under minimal selective constraint. We used MAFFT v7.520 (Katoh and Standley 2013) and pal2nal v14 (Suyama et al. 2006) to generate ungapped peptide-guided CDS alignments for 19,236 single-copy orthologs shared across all four assemblies. We then used a custom Python script (https://github.com/ahlawrence/fourfold_amino_acid_aligned) to identify and compare fourfold degenerate synonymous sites across each genome pair.

For each pairwise comparison, the script counted the number of fourfold sites with aligned nucleotides ("total"), and the number of nucleotide differences ("variants") at those sites. We calculated nucleotide diversity (π_{4fold}) as variants/total. We use the term π_{4fold}, rather than divergence (e.g., dxy), because most pairwise differences reflect polymorphism segregating within the interfertile *M. guttatus* complex rather than fixed differences between species. We further analysed relationships between π_{4fold}, recombination rate, and gene density using custom R scripts.

To compare standard resequencing-based estimates of nucleotide diversity to values obtained from direct alignment of de novo assemblies, we analysed Illumina data from each of the four *Mimulus* lines, down-sampled to 60 million read pairs per accession. Reads were adapter- and quality-trimmed with Trimmomatic v0.39 (Bolger et al. 2014), then mapped to each of the four reference assemblies using BWA v0.7.17 (Li and Durbin 2009). Alignments were filtered for map quality ≥ 29 and proper pairing using SAMtools v1.16.1 (Li et al. 2009), and duplicates were removed with Picard v2.27.5 ("Picard Toolkit," 2019). We called variants with GATK v4.4.0.0 (McKenna et al. 2010) using both biallelic SNPs and monomorphic sites. Filtering criteria included QD < 2.0, SOR > 3.0, MQ < 40.0 for all sites, plus FS > 60.0, QUAL < 40.0, MQRankSum < −12.5, and ReadPosRankSum < −12.5 or > 12.5 for variants. Four-fold degenerate sites were identified using genome-specific CDS annotations and a custom script (https://github.com/tsackton/linked-selection). We calculated π_{4fold} for each pairwise contrast by dividing the number of variant sites by the total number of four-fold sites with confident genotype calls. Heterozygous sites, which were rare in these inbred lines (0.16%–0.41% of fourfold sites), were excluded. To test the effect of variant-only filtering, we repeated the analysis without filters that exclude invariant sites.

2.8 | Comparative Genomics

Whole-genome synteny was visualised with DEEPSPACE (github.com/jtlovell/DEEPSPACE). Contig maps were visualised with GENESPACE (Lovell et al. 2022) with the following parameters: telomere kmers = CCCGAAA/CCCTAAA, maximum distance between adjacent kmers = 100, minimum telomere size = 400, and minium telomere k-mer density = 0.50. Orthogroups were calculated with OrthoFinder v2.5.2 (Emms and Kelly 2019) called through GENESPACE. Structural variants were detected with SyRI (Goel et al. 2019) applied to minimap2 (Li 2018) alignments. The pan-genome graph was built with minigraph-cactus (Hickey et al. 2023) and visualised with sequenceTubemap (https://github.com/vgteam/sequenceTubeMap; Beyer et al. 2019) and ODGI (Guarracino et al. 2022) with custom modifications when necessary. Centromeric regions were annotated by first running BLAST on the known 728 bp repeat (Fishman and Saunders 2008; Melters et al. 2013), then filtering to contiguous regions > 1200 bp.

3 | Results

3.1 | An Updated Reference Genome for *Mimulus guttatus* var. IM62

We generated an improved *Mimulus guttatus* IM62 reference genome using a whole-genome shotgun assembly approach (see Section 2). We sequenced PacBio HiFi reads (70.02× coverage, mean read length = 14.0 kb) and Omni-C reads (77.73×), assembled polished contigs (N50 = 6.1 Mb; Table 1), and scaffolded them using Hi-C data. Given the extremely low residual heterozygosity in IM62 (0.0016% of callable bases), we produced a haploid assembly. We polished scaffolds using 41.8× Illumina reads and finalised the assembly using custom scripts to join contigs, remove duplicates, and ensure proper orientation relative to the IM62 v2.0 assembly.

Our new v3.1 assembly incorporates 27 Mb of previously missing sequence, improves contiguity nearly 400-fold (Figure 2A; Table 1), and resolves gaps in (peri)centromeric regions that were underrepresented or misassembled in v2.0. We annotated 25,113 genes and 38,393 transcripts based on transcriptomic data (Table 2, Appendix S1).

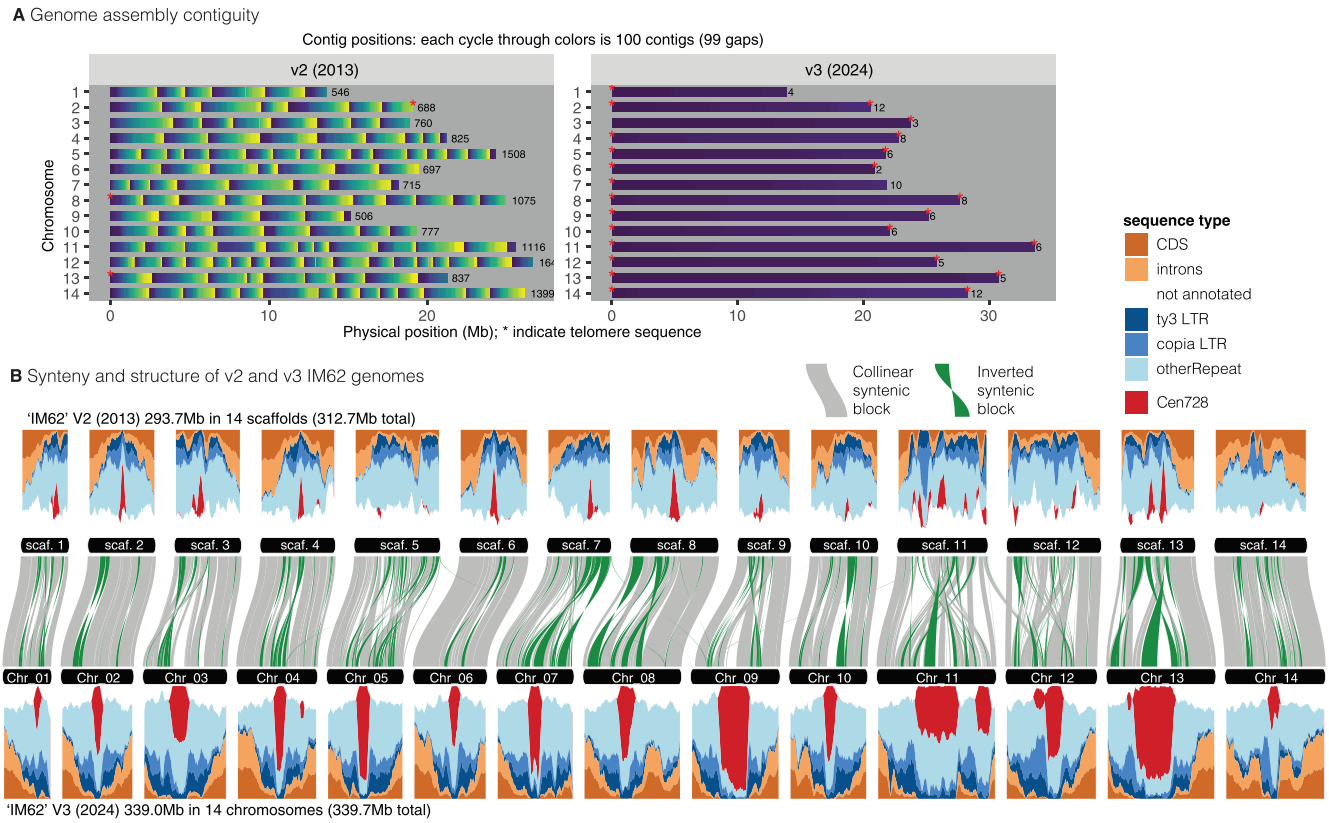

FIGURE 2 | The structure and improvements of the V3 *M. guttatus* IM62 genome. (A) Contiguity maps of the two genome versions for any scaffold > 5 Mb. In each panel, each adjacent contig is visualised as an adjacent colour on a 100-value blue-green-yellow ("viridis") palette. The total number of contigs per chromosome is printed to the right of the chromosome. Positions with significant telomeric repeats are flagged with a red*. (B) Genome architecture of and syntenic relationships between the v2 and v3 IM62 genomes. Top (v2) and bottom (v3) panels visualise 0.9 Mb-overlapping 1 Mb sliding windows that hierarchically classify the genomes by centromeric 15-mers, CDS, ty3, copia, other repeats, and introns. The centre panel connects the physical positions of both sliding window panels by a synteny map constructed using 1 kb non-overlapping windowed alignments between the two genomes. Inverted sequence is highlighted in green. All plots were built with GENESPACE v1.4 using default parameters except where specified.

TABLE 2 | Protein-coding gene annotation and sequencing statistics. Read depth and sequencing for gene model annotations.

	IM62	IM767	Nasutus (SF)	Tilingii (LVR)
RNA-seq tissues (*n*)	6	8	6	6
RNA-seq read pairs (M)	250	419	326	886
Iso-seq bases (Gb)	7.96	11.11	4.72	18.20
Iso-seq reads (M)	4.0	5.5	2.4	11.8
Gene models (*n*)	25,113	25,226	25,116	26,266
Alternative transcripts (*n*)	13,280	13,682	11,878	16,647
BUSCO (% Euk. Genes)	98	99.4	97.7	100

Despite major improvements in contiguity and repeat resolution, GENESPACE remains largely conserved between v2.0 and v3.1. Both assemblies include ~68 Mb of genic sequence, and 93.0% of annotated genes have orthologs in both versions (Appendix S2). This high concordance suggests that previous gene-based studies using v2.0 remain broadly valid. To facilitate conversion between assemblies, we provide a positional map of syntenic blocks (Appendix S3; Figure 2B), protein-coding orthologs (Appendix S2), and a chain file for variant call liftovers (Appendix S4).

In contrast, repeat content and pericentromeric sequence structure differ substantially between assemblies. While gene-rich chromosome arms remain collinear, the v3.1 assembly reveals extensive rearrangement and improved ordering of repetitive pericentromeric regions (Figure 2B). Notably, v3.1 contains 103.9 Mb of centromere-associated "Cent728" repeat sequence—over six times the amount annotated in v2.0 (17.2 Mb). Additionally, 25 of 28 chromosome ends in v3.1 are capped with telomeric repeats, compared to only three in v2.0. Other repetitive elements are also more abundant and better resolved in v3.1. Together, these improvements will enhance fine-scale genomic analyses in this key ecological model.

3.2 | Structural and Repetitive Variation Across the *Mimulus* Pan-Genome

To characterise structural and genic variation across the *Mimulus guttatus* species complex, we assembled and annotated three additional genomes: (1) IM767, an inbred line from the same Iron Mountain population as IM62 but genomically unrelated (Veltsos and Kelly 2024); (2) SF, an accession of self-fertilising *M. nasutus* (Brandvain et al. 2014); and (3) LVR, a *M. tilingii* accession from a distinct clade (Sandstedt et al. 2021; Figure 1). All three genomes were assembled and annotated using the same pipeline as IM62, with minor differences in coverage and curation (Tables 1 and 2, see Section 2). Each genome is slightly smaller (~313–315 Mb) than the IM62 genome (339 Mb).

In all three *M. guttatus* complex genomes, the centromere-associated 728-bp repeat Cent728 (Fishman and Saunders 2008) was concentrated at the centre of each chromosome, confirming it as the primary centromeric satellite across diverse populations (Figure 3). At the female meiotic drive locus (MDL11) on Chromosome 11 (Fishman and Willis 2005), IM62 carries the *D* driving allele, IM767 the D^{-} resistant type, and SF the *d* susceptible allele. Consistent with functional divergence and cytogenetics (Fishman and Saunders 2008), our assemblies reveal that the structurally distinct *D* chromosome uniquely contains two dense Cent728 arrays flanking a gene-rich Cent728-depleted region (Figure 3).

We constructed a whole-genome alignment and pan-genome graph across the four assemblies (.gfa and associated data are hosted on phytozome: https://phytozome-next.jgi.doe.gov/info/Mguttatusvar_IM62_v3_1), enabling interactive visualisation and summary of sequence divergence and gene content. Focusing on syntenic, alignable regions (Figure 3), we documented a total of 14,207 structural variants > 50 bp (13,984 insertion/deletions, 223 inversions; Appendix S5, Table S1). We confirmed the presence of three major inversions: on Chromosome 11 specific to IM62 *M. guttatus* (MDL11; (Flagel et al. 2019)), on Chromosome 10 specific to IM *M. guttatus* (Flagel et al. 2019), and on Chromosome 13 specific to *M. tilingii* (Garner et al. 2016). As expected, all four genomes carried the annual (or ancestral, in the case of *M. tilingii*) orientations of widespread life history-associated inversions on chromosomes 5 and 8 (Coughlan and Willis 2019; Flagel et al. 2019; Lowry and Willis 2010). Gene content also varied across the genomes: of 101,661 total annotated genes, 88,672 (87.2%) were contained in shared orthogroups across all genomes, while each genome also contained over 500 "private" genes within orthogroups specific to only one genome (Appendix S2).

To validate these patterns, we projected CDS sequences from each orthogroup's longest representative gene across all genomes. The presence or absence of full-length matches confirmed that most gene PAVs reflect real sequence differences, not annotation artefacts. For example, we observed deletion-driven absence of three consecutive genes on Chromosome 1 in both IM767 and SF (Appendix S2, Figure 4A). Genome-wide alignment coverage further underscored high structural divergence but also revealed the challenges of applying traditional graph-based pan-genome approaches to the diverse *M. guttatus* genomes. As expected, protein-coding sequences are generally well represented in the pan-genome graph; however, alignments even immediately adjacent to exons can become remarkably complex (i.e., Figure 4B). For example, the IM62 assembly aligned to only ~75% of the IM767 genome, and ~44%–45% of SF and LVR genomes (Figure 4C), combined, less than half of all sites across had physical positions that can be tracked across all four genomes. These numbers are comparable to the breadth of coverage from short-read alignments: IM62 short-read sequence aligns at sufficient coverage to only 63% of the IM767 genome, and to 42%–45% of the SF and LVR genomes (Table S2). While a graph pan-genome will likely improve variant detection and reduce reference bias, these large blocks of unalignable sequence will limit the utility of graph-based pan-genome exploration.

3.3 | Recombination Landscape Is Strongly Shaped by Gene Density

The four genome assemblies—particularly IM767, which is the recurrent parent in our mapping populations—provide

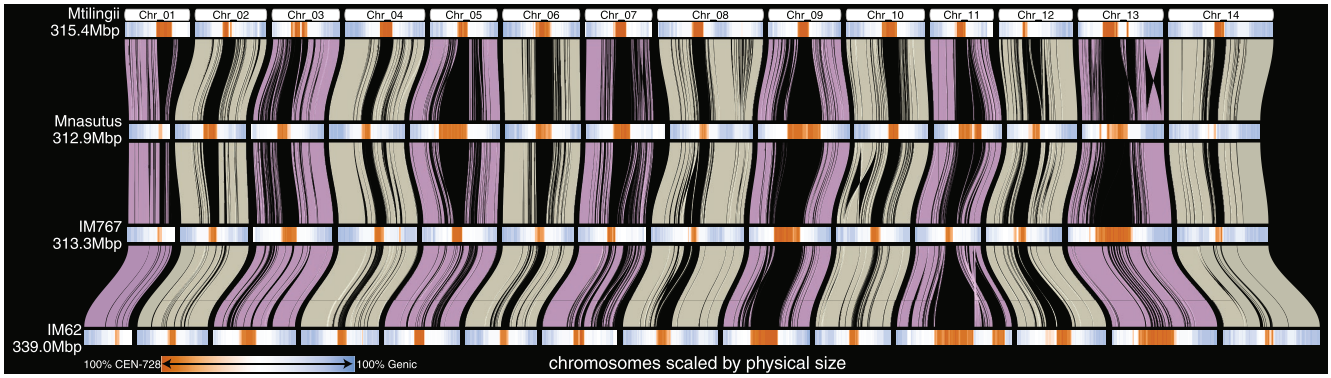

FIGURE 3 | DEEPSPACE synteny map highlighting positions of centromeric repeats. Collinear blocks between the same chromosomes are shown as transparent 'braids' and chromosome segments are visualised as colour-gradient rectangles along an *x*-axis that scales each genome by its physical size. Regions that do not map between chromosomes are visualised as black 'wedges' in the map. These can be due to ineffective unique mapping in highly repetitive centromeres (e.g., Chr 9 IM62-IM767), expansion of centromeric arrays (e.g., Chr 11 IM62-IM767) or sequence presence/absence. The orange-blue colour gradient indicates regions that are gene-rich (blue) or centromeres (orange); fully saturated colours indicate that all sequence in those intervals is attributable to that annotation type. White regions have neither genes nor Cent728 repeats, and are likely repeat-rich pericentromeres.

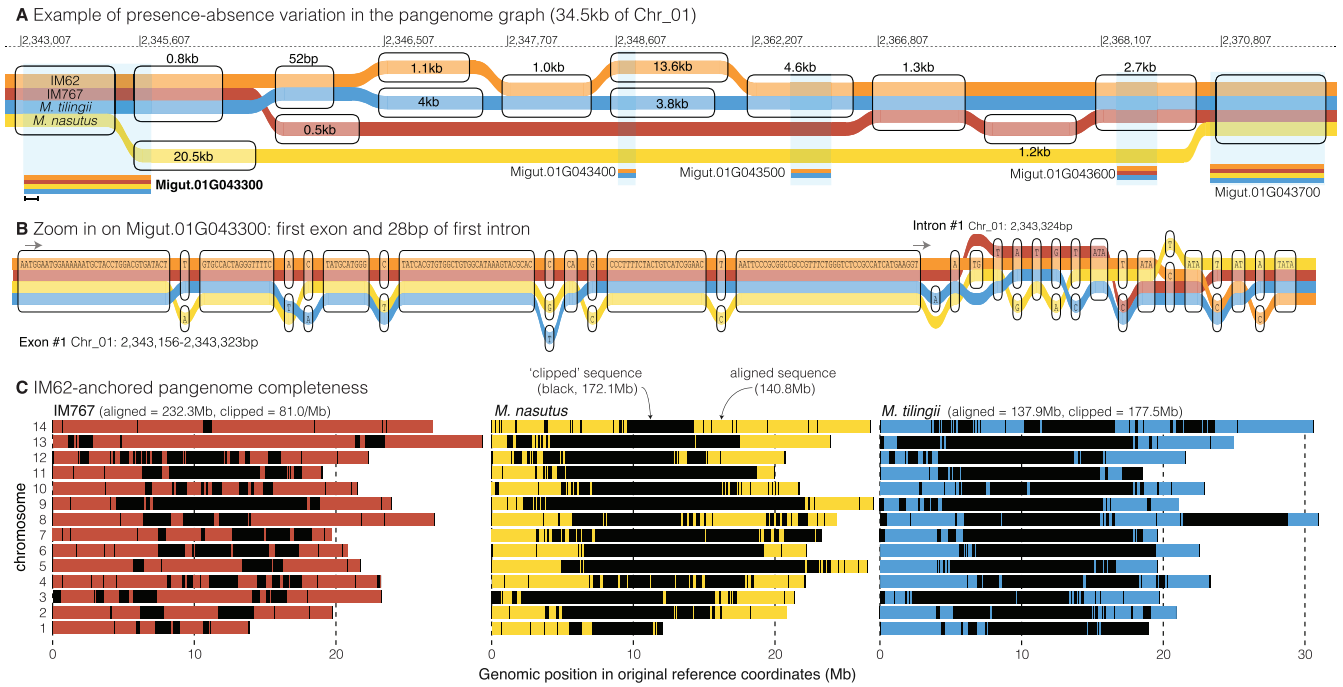

FIGURE 4 | Exploration of the pan-genome graph. (A) 2.34–2.37 Mb on Chromosome 1 is a complex region in the genome that illustrates how the pan-genome graph handles large-scale presence–absence variation. Here, any variants smaller than 52 bp are excluded so that the large insertion deletions in this region are apparent. Haplotypes containing similar sequences are binned in the transparent rounded rectangles ("nodes"), divergent but syntenic/orthologous sequences have stacked nodes, and deleted sequences show the path outside of a node. (B) The first exon and the first 28 bp of the first intron of a gene with family members across all four genomes is shown as an example of how SNPs and INDELs appear at the base-pair level in a sequence graph. (C) The positions of aligned sequence retained in the graph (coloured following the tubes in panels (A, B) and that which is unalignable and clipped (black)) are presented in a pan-genome anchored by IM62.

complete, contiguous coverage of both gene-rich and repetitive regions, allowing for exploration of the causes and consequences of recombination frequency. Using high-resolution linkage maps from 1373 F_2 plants derived from 10 Iron Mountain lines (Veltsos and Kelly 2024), we mapped 33,302 crossovers to the IM767 genome and estimated recombination rates across the genome (Figure 5A,B; Appendix S6).

We localised crossovers using RNAseq-based genotyping that takes advantage of the presence of informative SNPs present in many genes in each of the 10 crosses. As expected, this genotyping strategy yielded high precision in gene-rich chromosome arms but lower precision in gene-poor regions because of variation in the distance between flanking markers. Fortuitously, gene-poor regions also exhibited very low rates of recombination, Figure 5C). For example, only 2.4% of crossovers were placed in intervals > 1 Mb and 0.85% in intervals > 2 Mb. Mean linkage map lengths varied across chromosomes from ~60 cM (Chr 7) to ~130 cM (Chr 14), totalling ~1260 cM across the ~313 Mb IM767 genome—yielding an average recombination rate of ~4 cM/Mb. Recombination rates varied dramatically across chromosomes (Figure 5A). Gene-rich arms exhibited high recombination, while repetitive pericentromeric regions showed extremely low rates (e.g., Figure 6A). We classified pericentromeric regions as contiguous runs of ≥ 2500-kb windows with < 1 cM/Mb recombination and found that in total, pericentromeric regions span 110 Mb (~35%) of the genome. Only 694 of 33,302 crossovers (~2%) occurred in pericentromeric regions, corresponding to an average recombination rate of 0.23 cM/Mb, in contrast to an average rate of 6.2 cM/Mb in the remaining ~65% of the genome.

The recombination rate was tightly associated with gene content. Of 25,226 annotated genes, ~95% were located outside low-recombination regions. Across chromosomes, map length correlated strongly with gene count (Spearman $\rho = 0.965$, $p < 0.0001$). At finer scales, local recombination rate (cM/Mb) was positively correlated with gene density (genes/Mb; Figure 5C), and 50% of genes occurred in regions with recombination rates > 7.5 cM/Mb.

To explore fine-mapping potential, we divided the genome into ~1 cM windows and tallied gene start sites. On average, there were 19.13 gene start sites per cM, with 95% of the map containing < 40 gene start sites per cM (Figure 5D). Thus, with sufficient sample sizes (e.g., several thousand F_2s), forward-genetic mapping is feasible for most genes across the *Mimulus* genome.

3.4 | Patterns of Nucleotide Diversity Across Genomes and Recombination Landscapes

We measured pairwise nucleotide diversity at fourfold synonymous sites (π_{4fold}) among 19,063 single-copy genes shared across the four genomes (Table S3). For example, we observed $\pi_{4fold} = 0.032$ between the two *M. guttatus* accessions (IM62 and IM767), consistent with previous within-population estimates (Puzey et al. 2017). Pairwise differences between either IM accession and the selfing *M. nasutus* SF were slightly higher than twice the within-population value ($\pi_{4fold} = 0.070$), and only slightly higher still for comparisons of all three *M. guttatus* complex genomes to *M. tilingii* LVR ($\pi_{4fold} = 0.073$–0.074).

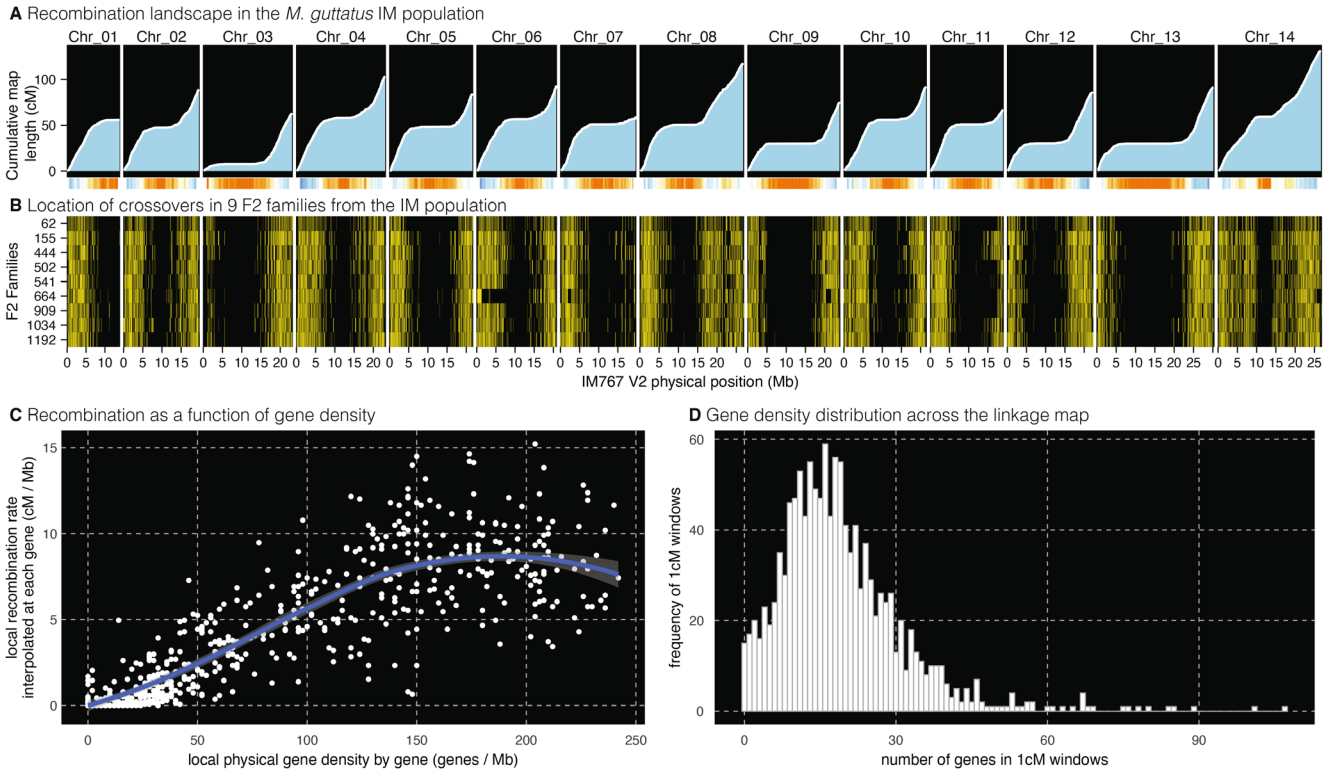

FIGURE 5 | Recombination landscape and impacts in *M. guttatus* IM population. (A) The population-wide mean linkage-physical position map (including a heatmap of mean recombination rate, blue = highest, red = lowest), and (B) the positions of crossover events for individual families in a multiparent *M. guttatus* IM population cross. (C) The relationship between local recombination rate interpolated at each annotated gene in the IM767 genome (cM/Mb) and local gene density (genes/Mb) by gene start sites in non-overlapping 500 kb windows. (D) The distribution of the number of annotated gene start sites in non-overlapping 1 cM windows across the genome.

We note that these π_{4fold} values were consistently 10%–15% lower when calculated using short-read alignments to a single reference genome (Tables S4 and S5) than the values based on comparisons among reference genomes. Under-calling of SNPs due to stringent variant-only filters is one cause of this discrepancy: removing these filters reduced the gap to ~6%–9% (Table S5). Restricting our analysis to only the 6199 genes with identical gene models among the four species closed this gap considerably (with estimates only ~3%–5% lower, Table S5), suggesting that differences in gene structure and annotation cause undercounting of orthologous fourfold sites when a single genome is used as the reference.

Contrary to expectations under linked selection, we found no strong correlation between local recombination rates and π_{4fold}. Even in pericentromeric regions with near-zero recombination, π_{4fold} values remained high (~0.03 for IM62 vs. IM767), nearly matching genome-wide averages (Figure 6A). We tested for finer-scale effects by examining the relationship between local recombination rate and π_{4fold} across individual genes. Across all comparisons (IM62 vs. IM767, IM767 vs. SF, IM767 vs. LVR), correlations were weak (Spearman $\rho = -0.015$ to 0.029, Figure 6B,C). We hypothesised that gene density variation might confound detection of linked selection, since selection targets are more concentrated in high-recombination regions. However, gene density (genes per cM) also exhibits only weak correlations with π_{4fold} ($\rho = 0.18$ within IM; -0.09 to -0.12 between species; Figure 6D,E). Thus, at least when recombination is assessed at the kilobase-scale of linkage mapping studies, genome-wide diversity in this group of yellow monkeyflowers does not appear to be strongly shaped in ways expected by classic models of linked selection. The positive association between gene density and recombination may roughly equalise the density of selected sites per cM, buffering the effect of linked selection on neutral diversity at all but the most local scales. More work is needed to examine the potential for linked selection to impact neutral diversity at much finer base-pair scales.

4 | Discussion

This study provides a new comparative genomic framework for the yellow monkeyflowers (*Mimulus* section *Simiolus*), a classic ecological and evolutionary model system. Using long-read sequencing and deeply supported integrative annotation, we generated chromosome-scale reference genome resources for three inbred lines from the *M. guttatus* species complex and a member of the outgroup species *M. tilingii*. These assemblies resolve large-scale structural variation, complex repetitive regions, and gene content variation, enabling new insights into recombination, sequence diversity, and the genomic consequences of selection.

A key contribution of this work is the substantial improvement of the *M. guttatus* IM62 reference genome. Compared to the widely used v2.0 assembly (Hellsten et al. 2013), our v3.1 version adds 27 Mb of sequence, improves contiguity nearly 400-fold, and resolves previously collapsed or misassembled pericentromeric

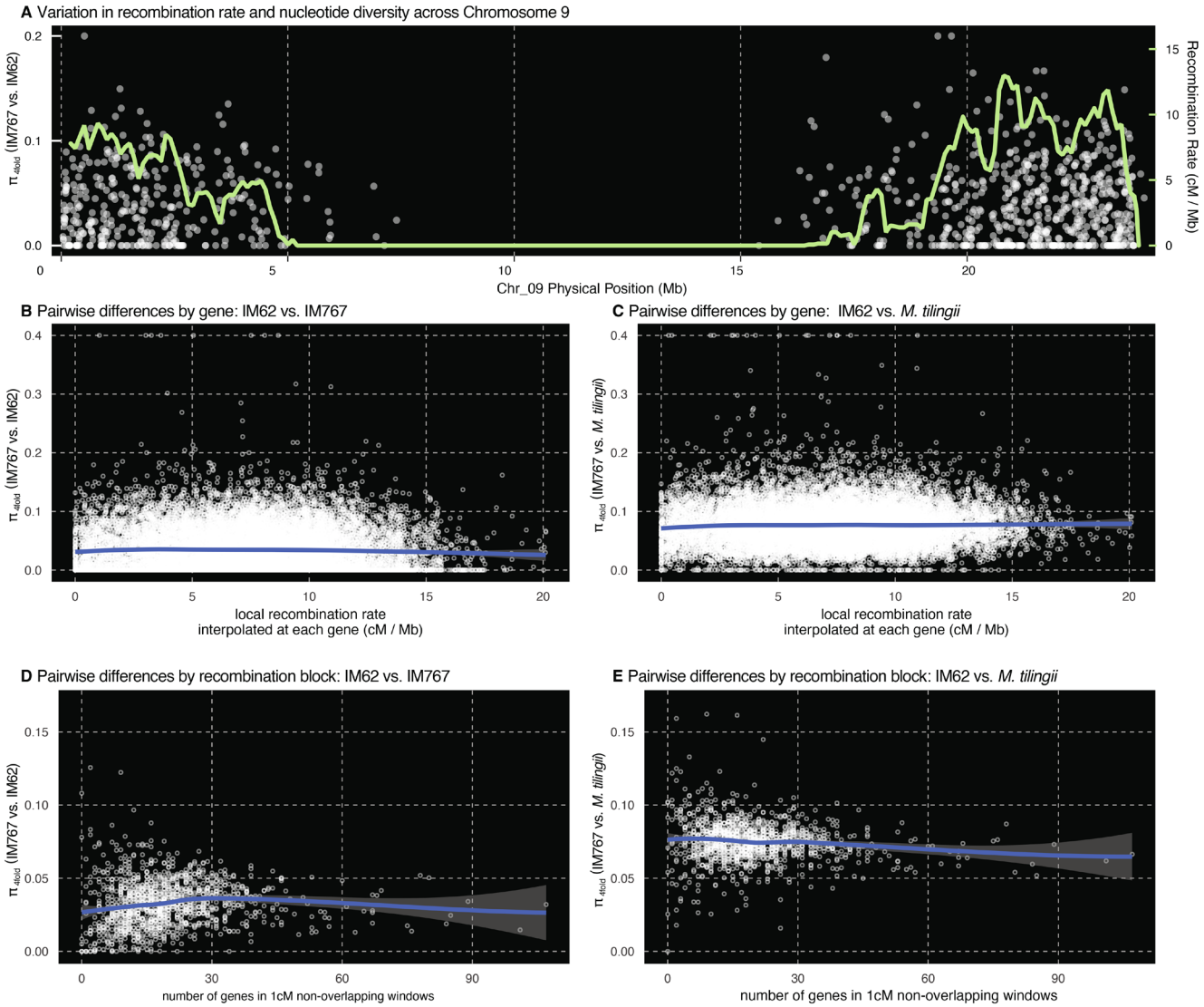

FIGURE 6 | Relationships between recombination and diversity. (A) The pattern of recombination and diversity along Chromosome 9 in IM767 is typical of the rest of the genome: Nearly all recombination (recombination rate in the green line) occurs on the chromosome arms where most of the genes are also found (white points indicate π_{4fold} at individual genes). Variation in local recombination rates (cM/Mb) do not predict variation in pairwise nucleotide differences (π_{4fold}) at genes within the IM population (B) or between IM767 and *M. tilingii* (C). Variation among genes in nucleotide differences (π_{4fold}) within the IM population (D) and between species (E) are also not related to local gene start site densities in 1 cM non-overlapping windows across the genome.

regions. These improvements stem from our use of PacBio HiFi reads and Omni-C scaffolding, which circumvent the limitations of recombination-based assembly in low-crossover regions. Despite large gains in assembly quality, the v3.1 gene set remains highly concordant with v2.0, preserving the utility of previous gene-based studies.

Across the four genomes, we identified extensive structural variation and gene presence–absence differences. Several large inversions and other structural variants previously implicated in genetic conflict, adaptive divergence, and/or species barriers (Fishman and Saunders 2008; Flagel et al. 2019; Zuellig and Sweigart 2018) were confirmed by our genome alignments. Further, the new assemblies fully resolve the structurally complex, highly repetitive pericentromeric and centromeric regions. Homogeneous arrays of the putatively centromeric Cent728 satellite repeat identified in *M. guttatus* (Fishman and Saunders 2008) define metacentric regions on all chromosomes across all three species, similar to centromeric satellites in other generally less diverse plant systems like *Arabidopsis thaliana* (Naish and Henderson 2024) (Wlodzimierz et al. 2023). Further, this continuity of *M. guttatus* centromere repeats provides a strong platform for analyses of centromere and transposable element evolution across monkeyflowers and for testing how structural and gene content variation contribute to functional differences among MDL11 alleles engaged in genetic conflict. The pan-genome analysis further revealed hundreds of genes unique to each genome, and alignment coverage between genomes dropped sharply outside of genic regions, emphasising the limits of pan-genomes for highly diverse groups.

One particularly striking feature of the recombination landscape is the near-complete absence of crossover events in the large

pericentromeric regions of each chromosome. Consistent with many other angiosperm genomes, recombination was extremely rare across all gene-poor pericentromeric regions of each monkeyflower chromosome (Figure 5A,B); even the few crossovers mapped to these regions likely occurred in or near islands of gene content at their edges, where gene-based marker density was higher (Veltsos and Kelly 2024). Such pericentromeric recombination suppression appears to be a general feature of plant genomes. Unlike model mammals with PRDM9-directed recombination hotspots (Paigen and Petkov 2018), plants and fungi tend to exhibit meiotic double-strand breaks and crossover enrichment in open chromatin regions, particularly around gene promoters (Choi et al. 2018; Pan et al. 2011; Yelina et al. 2012). Accordingly, our previous work in *Mimulus guttatus* found that crossovers were highly concentrated in genic regions and near transcription start sites (Hellsten et al. 2013). This bias in crossover locations likely accounts for the exceptionally strong positive correlation between local recombination rate and physical gene density (genes per Mb), and constrains the number of genes per centimorgan to a narrow range (~20 genes/cM) across both gene-rich and gene-poor regions. These features of the recombination landscape are important for interpreting genome-wide patterns of nucleotide diversity and the expected impact of linked selection.

Nucleotide diversity at fourfold degenerate synonymous sites ($\pi_{4\text{fold}}$) is unusually and uniformly high across monkeyflower genomes, with intra-population, intra-complex, and inter-complex levels of pairwise differences rivalling or exceeding the most polymorphic known animal and plant species (reviewed by Leffler et al. 2012; Corbett-Detig et al. 2015; Roberts and Josephs 2024; Romiguier et al. 2014, but see Dey et al. 2013). Within the IM population, diversity between IM62 and IM767 is ~3.2%, which is similar to the genome-wide divergence between human and orangutan genomes (Yoo et al. 2025; Locke et al. 2011). Pairwise differences between IM *M. guttatus* and *M. nasutus* (members of the same species complex) are 7%, and differences between any of those three genomes and outgroup *M. tilingii* reach 7.4%, values that are comparable with SNP divergence between great apes and Old World monkeys (Locke et al. 2011; Rhesus et al. 2007).

In contrast to classic expectations from linked selection theory (Begun and Aquadro 1992; Corbett-Detig et al. 2015; Cutter and Payseur 2013), local recombination rate is not positively correlated with nucleotide diversity in comparisons of yellow monkeyflower genomes. Even in near-zero recombination regions, $\pi_{4\text{fold}}$ remains comparable to high genome-wide averages. Furthermore, we fail to find compelling evidence of linked selection when we compare gene density and nucleotide diversity in 1 cM windows. The tight relationship between gene density and recombination rate in yellow monkeyflowers provides one possible explanation, as the genomic distribution of genic targets of selection mirrors recombination rate. Thus, at the centimorgan scale of recombination rates measured with high-resolution linkage mapping, the density of selected sites may be relatively constant, buffering against the effects of linked selection. These results challenge general expectations that diversity should be depleted in low-recombination regions and raise questions about the generality of linked selection as a dominant force structuring diversity in large, gene-dense plant genomes.

While classical models of linked selection were inspired by the positive correlations between recombination rate and nucleotide diversity observed in many animal genomes (e.g., *Drosophila*; Begun and Aquadro 1992), these *Mimulus* analyses add to a growing number of exceptions from plant systems (Slotte 2014). Early population genomic analyses in *Arabidopsis thaliana* reported only a weak relationship between recombination and diversity (Cao et al. 2011), and more recent work in *Arabidopsis* confirms that genes in low-recombination pericentromeric regions actually harbour higher diversity than genes on chromosome arms (Fernandes et al. 2024). Similar to the *Mimulus* species investigated here, no significant correlation between recombination and synonymous diversity was also reported in short-read based analyses of multiple *M. guttatus* and *M. nasutus* accessions (Brandvain et al. 2014), a result that contrasts those from the distantly related *M. aurantiacus* species complex (Stankowski et al. 2019). These results suggest that the link between recombination and diversity may be more complex—and weaker—than predicted by standard models, at least in highly polymorphic plant genomes. In *Mimulus*, highly effective population sizes, frequent gene flow, and high recombination in gene-rich regions reduce the genome-wide footprint of selection on linked neutral variation at cM scales. It remains to be seen whether linked selection impacts neutral diversity at much finer scales of recombination than measured experimentally, for example, at scales that reflect the impact of long-term population genetic processes on patterns of base-pair level variation in recombination and linkage disequilibrium.

The remarkably high diversity and gene density (averaging ~20 genes per cM genome-wide) of yellow monkeyflowers presents both advantages and challenges for researchers in this (and similar) plant systems. Importantly, high-recombination rates and high diversity in gene-dense regions make forward-genetic and fine-mapping approaches an effective tool for dissecting the genetic basis of ecologically and evolutionarily important traits. At the same time, high repeat content and extensive structural variation, in genes as well as intergenic regions and gene-poor pericentromeres, create serious challenges for sequence alignment (Figure 4). Thus, short-read resequencing analysis pipelines in such systems must account for low mappability and paralogous alignment (e.g., by only calling variants in genes or even just coding exons; Puzey et al. 2017).

Notably, the nucleotide diversity values among the four *Mimulus* reference genomes presented here exceed previously published estimates based on short-read alignment to the IM62 v2.0 reference (e.g., Brandvain et al. 2014; Garner et al. 2016; Sandstedt et al. 2021). In this study, we also find that, using standard short-read resequencing pipelines, $\pi_4$fold values were consistently 10%–15% lower than those derived from codon-aware alignment and comparisons among our four de novo reference assemblies. This discrepancy reflects two primary sources of bias. First, mapping short reads to a single reference genome underestimates the number of orthologous fourfold degenerate sites due to differences in annotation and gene structure among genomes. Second, variants observed in genome-to-genome alignments sometimes appear as no-calls in resequencing-based SNP datasets—a form of differential dropout that arises from the sensitivity of short-read variant callers (such as GATK) to alignment confidence and software-specific filtering thresholds. These

issues are exacerbated in highly polymorphic genomes like these yellow monkeyflowers, which differ substantially from the human-based systems for which these tools were originally designed. Still, the complexities of alignment in π_{4fold} analyses, which are restricted to genic regions, pale in comparison to the difficulties of comparing regions outside of genes, where short-read methods often fail to even produce alignments. These findings highlight the importance of pan-genome-aware approaches and genome-specific annotations when estimating diversity or functional variation from resequencing data.

The exceptionally high nucleotide diversity in these yellow monkeyflowers points to massive long-term effective population sizes and ancient variants at all taxonomic and geographic scales, creating further theoretical and methodological challenges for population genomics. Indeed, the expected age of common neutral sequence variants may be hundreds of thousands or even millions of years old (Kimura and Ohta 1973), and with a mutation rate of $\sim 7 \times 10^{-9}$ per bp (as it is in Arabidopsis; Weng et al. 2019), the expected time to the common ancestor of two random sequences from within the *M. guttatus* species complex (differing at ~7% of neutral sites) is 5 million generations. One potential challenge of population genetic analyses in this system stems from our observation that recombination in gene-rich regions ($r = \sim 6.2 \times 10^{-8}$ per bp, on average) may often exceed the per-base mutation rate. If per-nucleotide recombination rates are uniformly high across large chromosomal regions, then the effectiveness of population genomic approaches such as ancestral recombination graph (ARG)–based methods may be limited because high recombination can erode local genealogical signal (Ishigohoka and Liedvogel 2025). Thus, the *M. guttatus* complex presents a novel testing ground for the development of models and tools suited to high-diversity, high-recombination genomes.

The four monkeyflower high-quality reference genomes presented here offer robust scaffolds for trait mapping, population genomics, and functional studies in a system with remarkable diversity at phenotypic and genomic levels. Comparative genome analyses reveal extensive structural variation and divergence in repeat-rich, gene-poor regions, such that large portions of the genome are unalignable across accessions. As a result, the *Mimulus* pan-genome is primarily informative in genic regions, underscoring the limitations of reference-based resequencing in highly polymorphic taxa. Nonetheless, they provide a foundation for understanding the complex interactions among genome structure, recombination, and natural selection. Their dramatic sequence diversity and structural complexity, even at the within-population scale, further underscores the need for multiple high-quality references to fully capture plant genome variation.

Acknowledgements

The work (proposal: 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. Support for generation of the IM62 *M. guttatus* and SF *M. nasutus* tissues for assembly and annotation was provided by NSF OIA-1736249 and DEB-2344468 to L.F. and by the University of Montana ECOR and Genomics Core Facilities. Support for the generation of the LVR *M. tilingii* tissues for assembly and annotation was provided by NSF DEB-1856180 to A.L.S. The recombination mapping was supported by MCB-1940785 to J.K.K.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Non-monetary benefits from this work arise from the sequences themselves and the diversity presented herein. As such, the reference genome assemblies, annotations, orthologs, and other comparative resources are available on the JGI's Plant Genomics Platform: Phytozome (https://phytozome-next.jgi.doe.gov/). Sequencing reads, assemblies, and annotations are also available on NCBI under bioprojects: PRJNA1112459 (IM62), PRJNA1112461 (IM767), PRJNA1112462 (SF), PRJNA1112463 (LVR).

References

Begun, D. J., and C. F. Aquadro. 1992. "Levels of Naturally Occurring DNA Polymorphism Correlate With Recombination Rates in *D. melanogaster*." *Nature* 356, no. 6369: 519–520.

Beyer, W., A. M. Novak, G. Hickey, et al. 2019. "Sequence Tube Maps: Making Graph Genomes Intuitive to Commuters." *Bioinformatics* 35, no. 24: 5318–5320.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics (Oxford, England)* 30, no. 15: 2114–2120.

Brandvain, Y., A. M. Kenney, L. Flagel, G. Coop, and A. L. Sweigart. 2014. "Speciation and Introgression Between *Mimulus nasutus* and *Mimulus guttatus*." *PLoS Genetics* 10, no. 6: e1004410.

Broman, K. W., H. Wu, S. Sen, and G. A. Churchill. 2003. "R/Qtl: QTL Mapping in Experimental Crosses." *Bioinformatics (Oxford, England)* 19, no. 7: 889–890.

Cao, J., K. Schneeberger, S. Ossowski, et al. 2011. "Whole-Genome Sequencing of Multiple *Arabidopsis thaliana* Populations." *Nature Genetics* 43, no. 10: 956–963.

Cheng, H., G. T. Concepcion, X. Feng, H. Zhang, and H. Li. 2021. "Haplotype-Resolved de Novo Assembly Using Phased Assembly Graphs With Hifiasm." *Nature Methods* 18, no. 2: 170–175.

Choi, K., X. Zhao, A. J. Tock, et al. 2018. "Nucleosomes and DNA Methylation Shape Meiotic DSB Frequency in *Arabidopsis thaliana* Transposons and Gene Regulatory Regions." *Genome Research* 28, no. 4: 532–546.

Corbett-Detig, R. B., D. L. Hartl, and T. B. Sackton. 2015. "Natural Selection Constrains Neutral Diversity Across a Wide Range of Species." *PLoS Biology* 13, no. 4: e1002112.

Coughlan, J. M., and J. H. Willis. 2019. "Dissecting the Role of a Large Chromosomal Inversion in Life History Divergence Throughout the *Mimulus guttatus* Species Complex." *Molecular Ecology* 28, no. 6: 1343–1357.

Cutter, A. D., and B. A. Payseur. 2013. "Genomic Signatures of Selection at Linked Sites: Unifying the Disparity Among Species." *Nature Reviews Genetics* 14, no. 4: 262–274.

Dey, A., C. K. W. Chan, C. G. Thomas, and A. D. Cutter. 2013. "Molecular Hyperdiversity Defines Populations of the Nematode Caenorhabditis Brenneri." *Proceedings of the National Academy of Sciences of the United States of America* 110, no. 27: 11056–11060.

Doyle, J. J., and J. L. Doyle. 1987. "A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue." *Phytochemical Bulletin*. https://worldveg.tind.io/record/33886/.

Durand, N. C., M. S. Shamim, I. Machol, et al. 2016. "Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments." *Cell Systems* 3, no. 1: 95–98.

Emms, D. M., and S. Kelly. 2019. "OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics." *Genome Biology* 20, no. 1: 238.

Fernandes, J. B., M. Naish, Q. Lian, et al. 2024. "Structural Variation and DNA Methylation Shape the Centromere-Proximal Meiotic Crossover Landscape in Arabidopsis." *Genome Biology* 25, no. 1: 30.

Finseth, F., K. Brown, A. Demaree, and L. Fishman. 2022. "Supergene Potential of a Selfish Centromere." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 377, no. 1856: 20210208. https://doi.org/10.1098/rstb.2021.0208.

Finseth, F. R., T. C. Nelson, and L. Fishman. 2021. "Selfish Chromosomal Drive Shapes Recent Centromeric Histone Evolution in Monkeyflowers." *PLoS Genetics* 17, no. 4: e1009418.

Fishman, L., A. J. Kelly, E. Morgan, and J. H. Willis. 2001. "A Genetic Map in the *Mimulus guttatus* Species Complex Reveals Transmission Ratio Distortion due to Heterospecific Interactions." *Genetics* 159, no. 4: 1701–1716.

Fishman, L., A. J. Kelly, and J. H. Willis. 2002. "Minor Quantitative Trait Loci Underlie Floral Traits Associated With Mating System Divergence in Mimulus." *Evolution* 56, no. 11: 2138–2155.

Fishman, L., and A. Saunders. 2008. "Centromere-Associated Female Meiotic Drive Entails Male Fitness Costs in Monkeyflowers." *Science* 322, no. 5907: 1559–1562.

Fishman, L., and J. H. Willis. 2005. "A Novel Meiotic Drive Locus Almost Completely Distorts Segregation in Mimulus (Monkeyflower) Hybrids." *Genetics* 169, no. 1: 347–353. https://doi.org/10.1534/genetics.104.032789.

Flagel, L. E., B. K. Blackman, L. Fishman, P. J. Monnahan, A. Sweigart, and J. K. Kelly. 2019. "GOOGA: A Platform to Synthesize Mapping Experiments and Identify Genomic Structural Diversity." *PLoS Computational Biology* 15, no. 4: e1006949.

Flynn, J. M., R. Hubley, C. Goubert, et al. 2020. "RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families." *Proceedings of the National Academy of Sciences of the United States of America* 117, no. 17: 9451–9457.

Friedman, J., and J. H. Willis. 2013. "Major QTLs for Critical Photoperiod and Vernalization Underlie Extensive Variation in Flowering in the *Mimulus guttatus* Species Complex." *New Phytologist* 199, no. 2: 571–583.

Garner, A. G., A. M. Kenney, L. Fishman, and A. L. Sweigart. 2016. "Genetic Loci With Parent-of-Origin Effects Cause Hybrid Seed Lethality in Crosses Between *Mimulus* Species." *New Phytologist* 211, no. 1: 319–331.

Goel, M., H. Sun, W.-B. Jiao, and K. Schneeberger. 2019. "SyRI: Finding Genomic Rearrangements and Local Sequence Differences From Whole-Genome Assemblies." *Genome Biology* 20, no. 1: 277.

Guarracino, A., S. Heumos, S. Nahnsen, P. Prins, and E. Garrison. 2022. "ODGI: Understanding Pangenome Graphs." *Bioinformatics* 38, no. 13: 3319–3326.

Haas, B. J., A. L. Delcher, S. M. Mount, et al. 2003. "Improving the Arabidopsis Genome Annotation Using Maximal Transcript Alignment Assemblies." *Nucleic Acids Research* 31, no. 19: 5654–5666.

Hellsten, U., K. M. Wright, J. Jenkins, et al. 2013. "Fine-Scale Variation in Meiotic Recombination in *Mimulus* Inferred From Population Shotgun Sequencing." *Proceedings of the National Academy of Sciences of the United States of America* 110, no. 48: 19478–19482.

Hickey, G., J. Monlong, J. Ebler, et al. 2023. "Pangenome Graph Construction From Genome Alignments With Minigraph-Cactus." *Nature Biotechnology* 42: 663–673. https://doi.org/10.1038/s41587-023-01793-w.

Ishigohoka, J., and M. Liedvogel. 2025. "High-Recombining Genomic Regions Affect Demography Inference Based on Ancestral Recombination Graphs." *Genetics* 229, no. 3: iyaf004. https://doi.org/10.1093/genetics/iyaf004.

Ivey, C. T., N. M. Habecker, J. P. Bergmann, J. Ewald, M. E. Frayer, and J. M. Coughlan. 2023. "Weak Reproductive Isolation and Extensive Gene Flow Between Mimulus *glaucescens* and *M. guttatus* in Northern California." *Evolution* 77, no. 5: 1245–1261.

Jones, P., D. Binns, H.-Y. Chang, et al. 2014. "InterProScan 5: Genome-Scale Protein Function Classification." *Bioinformatics* 30, no. 9: 1236–1240.

Katoh, K., and D. M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30, no. 4: 772–780.

Kelly, J. K. 2022. "The Genomic Scale of Fluctuating Selection in a Natural Plant Population." *Evolution Letters* 6, no. 6: 506–521.

Kimura, M., and T. Ohta. 1973. "The Age of a Neutral Mutant Persisting in a Finite Population." *Genetics* 75, no. 1: 199–212.

Kolis, K. M., C. S. Berg, T. C. Nelson, and L. Fishman. 2022. "Population Genomic Consequences of Life-History and Mating System Adaptation to a Geothermal Soil Mosaic in Yellow Monkeyflowers." *Evolution* 76, no. 4: 765–781.

Leffler, E. M., K. Bullaughey, D. R. Matute, et al. 2012. "Revisiting an Old Riddle: What Determines Genetic Diversity Levels Within Species?" *PLoS Biology* 10, no. 9: e1001388.

Li, H. 2018. "Minimap2: Pairwise Alignment for Nucleotide Sequences." *Bioinformatics* 34, no. 18: 3094–3100.

Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment With Burrows-Wheeler Transform." *Bioinformatics* 25, no. 14: 1754–1760.

Li, H., B. Handsaker, A. Wysoker, et al. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25, no. 16: 2078–2079.

Locke, D. P., L. W. Hillier, W. C. Warren, et al. 2011. "Comparative and Demographic Analysis of Orang-Utan Genomes." *Nature* 469, no. 7331: 529–533. https://doi.org/10.1038/nature09687.

Lovell, J. T., J. Jenkins, D. B. Lowry, et al. 2018. "The Genomic Landscape of Molecular Responses to Natural Drought Stress in *Panicum hallii*." *Nature Communications* 9, no. 1: 5213.

Lovell, J. T., A. Sreedasyam, M. E. Schranz, et al. 2022. "GENESPACE Tracks Regions of Interest and Gene Copy Number Variation Across Multiple Genomes." *eLife* 11: e78526. https://doi.org/10.7554/eLife.78526.

Lowry, D. B., and J. H. Willis. 2010. "A Widespread Chromosomal Inversion Polymorphism Contributes to a Major Life-History Transition, Local Adaptation, and Reproductive Isolation." *PLoS Biology* 8, no. 9: e1000500.

McKenna, A., M. Hanna, E. Banks, et al. 2010. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data." *Genome Research* 20, no. 9: 1297–1303.

Melters, D. P., K. R. Bradnam, H. A. Young, et al. 2013. "Comparative Analysis of Tandem Repeats from Hundreds of Species Reveals Unique Insights into Centromere Evolution." *Genome Biology* 14: R10. https://doi.org/10.1186/gb-2013-14-1-r10.

Naish, M., and I. R. Henderson. 2024. "The Structure, Function, and Evolution of Plant Centromeres." *Genome Research* 34, no. 2: 161–178.

Nesom, G. L. 2012. "Taxonomy of Erythranthe Sect. Simiola (Phrymaceae) in the USA and Mexico." *Phyton* 40, no. 1: 1–123.

Paigen, K., and P. M. Petkov. 2018. "PRDM9 and Its Role in Genetic Recombination." *Trends in Genetics: TIG* 34, no. 4: 291–300.

Pan, J., M. Sasaki, R. Kniewel, et al. 2011. "A Hierarchical Combination of Factors Shapes the Genome-Wide Topography of Yeast Meiotic Recombination Initiation." *Cell* 144, no. 5: 719–731.

Picard toolkit. 2019. *Broad Institute, GitHub Repository*. Broad Institute. https://broadinstitute.github.io/picard/.

Puzey, J. R., J. H. Willis, and J. K. Kelly. 2017. "Population Structure and Local Selection Yield High Genomic Variation in *Mimulus guttatus*." *Molecular Ecology* 26, no. 2: 519–535.

Rhesus Macaque Genome Sequencing and Analysis Consortium, R. A. Gibbs, J. Rogers, et al. 2007. "Evolutionary and Biomedical Insights from the Rhesus Macaque Genome." *Science* 316, no. 5822: 222–234. https://doi.org/10.1126/science.1139247.

Roberts, M., and E. B. Josephs. 2024. "Previously Unmeasured Genetic Diversity Explains Part of Lewontin's Paradox in a k-Mer-Based Meta-Analysis of 112 Plant Species." *bioRxiv.org*: The Preprint Server for Biology. https://doi.org/10.1101/2024.05.17.594778.

Romiguier, J., P. Gayral, M. Ballenghien, et al. 2014. "Comparative Population Genomics in Animals Uncovers the Determinants of Genetic Diversity." *Nature* 515, no. 7526: 261–263.

Sandstedt, G. D., C. A. Wu, and A. L. Sweigart. 2021. "Evolution of Multiple Postzygotic Barriers Between Species of the *Mimulus tilingii* Complex." *Evolution* 75, no. 3: 600–613.

Slotte, T. 2014. "The Impact of Linked Selection on Plant Genomic Variation." *Briefings in Functional Genomics* 13, no. 4: 268–275.

Stanke, M., O. Schöffmann, B. Morgenstern, and S. Waack. 2006. "Gene Prediction in Eukaryotes With a Generalized Hidden Markov Model That Uses Hints From External Sources." *BMC Bioinformatics* 7, no. 1: 62.

Stankowski, S., M. A. Chase, A. M. Fuiten, M. F. Rodrigues, P. L. Ralph, and M. A. Streisfeld. 2019. "Widespread Selection and Gene Flow Shape the Genomic Landscape During a Radiation of Monkeyflowers." *PLoS Biology* 17, no. 7: e3000391. https://doi.org/10.1371/journal.pbio.3000391.

Suyama, M., D. Torrents, and P. Bork. 2006. "PAL2NAL: Robust Conversion of Protein Sequence Alignments Into the Corresponding Codon Alignments." *Nucleic Acids Research* 34, no. Web Server issue: W609–W612.

Sweigart, A. L., and J. H. Willis. 2003. "Patterns of Nucleotide Diversity in Two Species of Mimulus Are Affected by Mating System and Asymmetric Introgression." *Evolution* 57, no. 11: 2490–2506.

Troth, A., J. R. Puzey, R. S. Kim, J. H. Willis, and J. K. Kelly. 2018. "Selective Trade-Offs Maintain Alleles Underpinning Complex Trait Variation in Plants." *Science* 361, no. 6401: 475–478.

Twyford, A. D., E. L. Y. Wong, and J. Friedman. 2020. "Multi-Level Patterns of Genetic Structure and Isolation by Distance in the Widespread Plant *Mimulus guttatus*." *Heredity* 125, no. 4: 227–239.

Vaser, R., I. Sović, N. Nagarajan, and M. Šikić. 2017. "Fast and Accurate de Novo Genome Assembly From Long Uncorrected Reads." *Genome Research* 27, no. 5: 737–746.

Veltsos, P., and J. K. Kelly. 2024. "The Quantitative Genetics of Gene Expression in *Mimulus guttatus*." *PLoS Genetics* 20, no. 4: e1011072.

Vickery, R. K., Jr. 1978. "Case Studies in the Evolution of Species Complexes in *Mimulus*." In *Evolutionary Biology*, 405–507. Springer US.

Weng, M.-L., C. Becker, J. Hildebrandt, et al. 2019. "Fine-Grained Analysis of Spontaneous Mutation Spectrum and Frequency in *Arabidopsis thaliana*." *Genetics* 211, no. 2: 703–714.

Willis, J. H. 1993. "Partial Self-Fertilization and Inbreeding Depression in Two Populations of *Mimulus guttatus*." *Heredity* 71, no. 2: 145–154.

Willis, J. H. 1999. "The Role of Genes of Large Effect on Inbreeding Depression in *Mimulus guttatus*." *Evolution* 53, no. 6: 1678–1691.

Wlodzimierz, P., F. A. Rabanal, R. Burns, et al. 2023. "Cycles of Satellite and Transposon Evolution in Arabidopsis Centromeres." *Nature* 618, no. 7965: 557–565.

Wu, C. A., D. B. Lowry, A. M. Cooley, K. M. Wright, Y. W. Lee, and J. H. Willis. 2008. "Mimulus Is an Emerging Model System for the Integration of Ecological and Genomic Studies." *Heredity* 100, no. 2: 220–230.

Wu, T. D., and S. Nacu. 2010. "Fast and SNP-Tolerant Detection of Complex Variants and Splicing in Short Reads." *Bioinformatics* 26, no. 7: 873–881.

Wu, T. D., and C. K. Watanabe. 2005. "GMAP: A Genomic Mapping and Alignment Program for mRNA and EST Sequences." *Bioinformatics* 21, no. 9: 1859–1875.

Yelina, N. E., K. Choi, L. Chelysheva, et al. 2012. "Epigenetic Remodeling of Meiotic Crossover Frequency in *Arabidopsis thaliana* DNA Methyltransferase Mutants." *PLoS Genetics* 8, no. 8: e1002844.

Yoo, D., A. Rhie, P. Hebbar, et al. 2025. "Complete Sequencing of Ape Genomes." *Nature* 641, no. 8062: 401–418. https://doi.org/10.1038/s41586-025-08816-3.

Zuellig, M. P., and A. L. Sweigart. 2018. "Gene Duplicates Cause Hybrid Lethality Between Sympatric Species of *Mimulus*." *PLoS Genetics* 14, no. 4: e1007130.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.