Molecular Biology and Evolution, 2025, **42**, 1–18
https://doi.org/10.1093/molbev/msaf304
Advance access publication 24 December 2025
Discoveries

MBE OXFORD

Recombination Suppression Drives Expansion of the *Drosophila* Dot Chromosome

Timothy J. Stanek [1], Wilson Leung [2], Christopher D. Shaffer [2], The Genomics Education Partnership[†], Ishtar Olaveja [3], Annabelle Laughlin [2,3,7], Jaquelyn Hester [1,3], Darwin Garrido [3], Emily K. Oh [2,3], Maria Volski [1,3,8], Nistha Panda [2,3,9], Mia Mo [2,3], Ethan Cordes [2,10], Martin Dalling [2,3,11], Kacie Kershaw [3,4], Malcolm Arnott [3,5,12], Stephen Daly [2,3,13], Silvia Garcia Valenzuela [3], Paige Thompson [1,3], Kayla L. Hastert [3], Destiny Sabb [3], Kathryn Karpinski [3,4,14], Meher Naaz Arora [2,15], Nicholas Rius [3], Larissa LoBello [2,3], Sebastian Jaramillo [3], Omkar Sonavane [1,3], Alice Herrmann [2,3,16], Laura K. Reed [6], Sarah C.R. Elgin [2], Cindy Arrigo [3], Christopher E. Ellison *[1]

[1]Department of Genetics, Human Genetics Institute of New Jersey, Rutgers University, Piscataway, NJ, USA
[2]Department of Biology, Washington University in St. Louis, St. Louis, MO, USA
[3]Department of Biology, New Jersey City University, Jersey City, NJ, USA
[4]Department of Biology, Saint Joseph's University, Philadelphia, PA, USA
[5]Department of Biology, Lewis & Clark College, Portland, OR, USA
[6]Department of Biological Sciences, The University of Alabama, Tuscaloosa, AL, USA
[7]Present address: Cooper Medical School of Rowan University, Camden, NJ, USA
[8]Present address: Temple Lewis Katz School of Medicine, Bethlehem, PA, USA
[9]Present address: University of Wisconsin - Madison, Madison, WI, USA
[10]Present address: Weill Cornell Medicine, New York, NY, USA
[11]Present address: Stanford University, Stanford, CA, USA
[12]Present address: University of Delaware, Newark, DE, USA
[13]Present address: University of Cincinnati College of Medicine, Cincinnati, OH, USA
[14]Present address: Thomas Jefferson University, Philadelphia, PA, USA
[15]Present address: Saint Louis University School of Medicine, St. Louis, MO, USA
[16]Present address: New York University College of Dentistry, New York, NY, USA

[†]Full list of Genomic Education Partnership co-authors in Appendix.
***Corresponding author:** E-mail: chris.ellison@rutgers.edu.*
Associate editor: Harmit Malik

Abstract

Genome size varies widely, even among closely related species, yet much less is known about chromosome size variation. Here we use the fourth chromosome of *Drosophila*, also known as the "Muller F element" or "dot chromosome", as a model to investigate chromosome-specific size expansion. The F element of most *Drosophila* species is small (~1.3 Mb) and almost entirely heterochromatic, yet harbors approximately 80 protein-coding genes. Here, we study *D. kikkawai*, *D. takahashii*, *D. ananassae*, and *D. bipectinata*, whose F elements are 2- to 15-fold larger in size compared to *D. melanogaster*. Through manual gene curation and comparative genomic analysis, we find that their F elements have expanded primarily via accumulation of transposable elements (TEs) in introns and intergenic regions. Natural selection appears less efficient on these expanded F elements: they have smaller effective population sizes and their genes exhibit reduced usage of optimal codons, compared to *D. melanogaster*. We propose that F element size variation is driven by differences in F element recombination rates. The ultra-long (~20 Mb) F elements of *D. ananassae* and *D. bipectinata* display high rates of rearrangement and sequence evolution and exhibit independent TE-driven expansions. Our results suggest that F elements of most *Drosophila* species likely recombine enough to prevent size expansion, while F element recombination in *D. ananassae* and *D. bipectinata* is either absent or rare enough to allow TEs and other deleterious mutations to accumulate via Muller's ratchet; thus, these chromosomes evolve more like a Y chromosome than a typical *Drosophila* F element.

Keywords: *Drosophila*, genome size, heterochromatin

Introduction

Eukaryotic genomes can vary greatly in size, ranging from less than 5 Mb in some unicellular parasites (Corradi et al. 2010) to greater than 160 Gb in the plant *Tmesipteris oblanceolata* (Fernández et al. 2024), with ca. only 12-fold variation in protein-coding gene content. This variation in genome size does not correlate with organismal complexity, a phenomenon which was termed the "C-value paradox" (Thomas 1971). Instead, it is now clear that variation in genome size

Received: August 20, 2025. **Revised:** October 17, 2025. **Accepted:** October 30, 2025
© The Author(s) 2025. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

among species is due primarily to differences in the abundance of noncoding, repetitive DNA, including transposable elements (TEs) and DNA satellites (Elliott and Gregory 2015). This observation explains how genome size varies but does not explain why genome size varies widely, even among closely related species.

One potential explanation, known as the mutational hazard hypothesis (Lynch and Conery 2003), is based on the observation that effective population size (*Ne*) is negatively correlated with genome size. Natural selection is less efficient in species with lower effective population size, which limits the ability of purifying selection to remove mildly deleterious mutations, including repetitive DNA, from the genome. This hypothesis remains controversial, with studies across diverse species groups split between those supporting an association between *Ne* and genome size (eg Lynch and Conery 2003; Lefébure et al. 2017; Cui et al. 2019; Chak et al. 2021; Mérel et al. 2021; Mérel et al. 2025) and a similar number finding no such association (eg Whitney and Garland 2010; Whitney et al. 2011; Kapheim et al. 2015; Bast et al. 2016; Mackintosh et al. 2019; Roddy et al. 2021; Marino et al. 2024; Yang et al. 2024).

Relatedly, the mutational equilibrium hypothesis (Petrov 2002) posits that genome size results from a balance between DNA insertions (eg duplications and/or mobile genetic elements) and DNA deletions. Therefore, variation in genome size is primarily due to differences among species in the rates of new insertions and/or deletions. More recent work describes a similar "accordion model," where genome size expansions occur rapidly via large duplications and/or bursts of TE activity but are counterbalanced by large deletions (Kapusta et al. 2017). Both of the above hypotheses (mutational hazard and mutational equilibrium) suggest that genetic drift plays a dominant role in genome size expansion (for review, see Blommaert 2020).

Repetitive DNA is largely composed of mobile genetic elements, which can selfishly and rapidly replicate to high copy numbers if left unchecked by their host genome (Talla et al. 2017; Naville et al. 2019; Chen et al. 2020; Oggenfuss et al. 2021). Differences among species in the efficiency by which repetitive DNA is silenced and/or purged from the genome will therefore lead to genome size variation. In fact, the evolution of repetitive DNA silencing mechanisms may facilitate genome size expansion (Zhou et al. 2020): once a particular repeat class is effectively silenced by its host genome, its presence in the genome becomes less deleterious or even neutral (Lu and Clark 2010).

Little is understood about how genome size expansion occurs at the level of individual chromosomes. In general, chromosome lengths within a given genome tend to be similar regardless of the genome's absolute size (Li et al. 2011), implying that genome size expansion often involves an increase in size across all chromosomes. However, there are exceptions to this pattern: chromosomes do not change in size at all, at least initially, when genome expansion occurs via polyploidization (Wang et al. 2021); in other cases, individual chromosomes can show extreme size variation across species. For example, across the *Drosophila* genus, there is less than threefold variation in overall genome size (Bosco et al. 2007; Gregory and Johnston 2008; Hjelmen et al. 2019), yet one particular chromosome, the Muller F element (known in *D. melanogaster* as the fourth chromosome or "dot" chromosome), shows more than tenfold variation in size (Leung et al. 2017). Here, we conduct comparative and population genomic analyses across four *Drosophila* species whose Muller F elements have expanded in size in order to determine whether the same evolutionary forces–genetic drift and repetitive DNA proliferation–that influence genome size expansion also play a role in the size expansion of individual chromosomes.

The F element (also known as Muller F) in *D. melanogaster* is unique in that it is generally very small and almost entirely heterochromatic, with only the distal 1.3 Mb region of the long arm containing ~80 protein-coding genes (Riddle and Elgin 2018). These genes are expressed at levels similar to euchromatic genes on the other autosomes (Leung et al. 2017). In related species, these genes are well conserved on their respective F elements, but the F elements themselves vary significantly in size and structure (Bhutkar et al. 2008 ; Leung et al. 2015). For example, the F elements of the *montium* group species *Drosophila kikkawai* and the *melanogaster* group species *Drosophila takahashii* are 2.3 Mb and 3.2 Mb, respectively (Fig. 1). Larger expansions are found in the *ananassae* group species *Drosophila ananassae* and *Drosophila bipectinata*, whose F elements are 19.4 Mb and 20.5 Mb, respectively, and have adopted a metacentric chromosome structure (Fig. 1). In this study, we assess the expansion of the Muller F elements in these four *Drosophila* species using chromosome-level long-read genome assemblies generated by the Genomics Education Partnership (GEP) (Shaffer et al. 2010; Leung et al. 2023). We find that F element expansion occurs in both intronic and intergenic regions, driven by an accumulation of TEs and other repetitive DNA. Genes on the expanded F elements of all four species show reduced codon usage bias relative to other genes in the same genome and genes on the *D. melanogaster* F element, consistent with less efficient natural selection on these chromosomes.

D. ananassae and *D. bipectinata* both have similarly sized, ultra-long F elements (Leung et al. 2023). Although the expanded regions of the F element are composed primarily of repetitive DNA in both species, we find that different repeats have accumulated in each species. Thus, either F element expansion has occurred independently or the ancestral repeats responsible for the initial expansion have been deleted and replaced by younger, lineage-specific repeat insertions in each species (ie repeat turnover). We propose that chromosome-specific differences in recombination rate may explain why the F element expands in size in some *Drosophila* species and not others. Furthermore, we propose that a large reduction (or elimination) of recombination may explain the extreme properties of the ultra-long F elements in the *D. ananassae* species group. Thus, the F element, which represents the ancestral X chromosome of higher dipteran species (Vicoso and Bachtrog 2015), is more similar to a Y chromosome in the *D. ananassae* species complex. Together, our results suggest that the evolutionary forces that determine genome size variation also play a role in chromosome-specific size expansion.

Results

Gene Annotations and Ortholog Identification

Using improved, full-length genome assemblies of *D. takahashii*, *D. kikkawai*, *D. ananassae*, and *D. bipectinata* (Leung et al. 2023), we annotated all protein-coding genes on the heterochromatic Muller F element and manually identified their *D. melanogaster* orthologs (see Methods). We also annotated all protein-coding genes in a

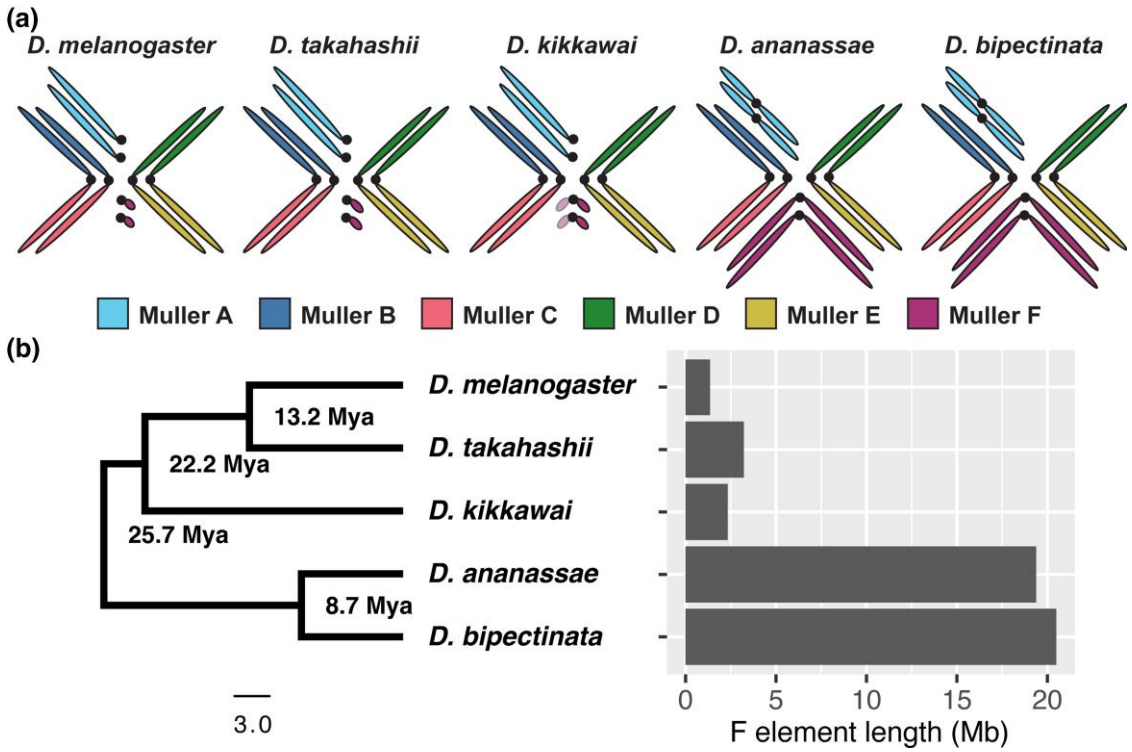


Fig. 1. *Drosophila* species with expanded F elements. a) Karyotypes of the four expanded F element species. Muller elements A to F correspond to *D. melanogaster* chromosomes X, 2L, 2R, 3L, 3R, and 4, respectively. The X chromosomes (Muller A) and F elements of *D. ananassae* and *D. bipectinata* are both metacentric. Both telocentric (dark red arm only) and metacentric (both dark and light red arms) F elements have been reported in *D. kikkawai* (Baimai and Chumchong 1980). Previous work suggests that the strain used here has the acrocentric F element (Leung et al. 2023). Panel adapted from Leung et al. (2023). b) Phylogenetic relationships among the species studied here, along with the F element scaffold length from their respective genome assemblies (Leung et al. 2023). Divergence times (in millions of years) are shown for each node of the tree (Suvorov et al. 2022).

Table 1 Summary of gene annotations by species

Species	Muller D			Muller F		
	Genes	Transcripts	Genes/Mb	Genes	Transcripts	Genes/Mb
D. melanogaster	119	246	91.3	80	298	59.3
D. takahashii	96	244	69.0	77	313	23.9
D. kikkawai	149	342	70.0	80	337	34.4
D. ananassae	123	284	68.1	78	295	4.0
D. bipectinata	177	372	91.2	79	291	3.9

euchromatic region (~2 Mb) near the base of Muller D, for use as a set of reference genes in each species (hereafter referred to as Muller D reference genes or Muller D reference region) (Table 1). The Muller D reference genes were selected based on their proximity to the pericentromeric heterochromatin (Leung et al. 2015). Although the majority of genes (ca. 95%) have been shown to remain on the same Muller element across species (Bhutkar et al. 2008), only 29 genes (16% to 30% of all reference genes, depending on species) are found in the reference regions of all four species, likely due to inversions and/or other chromosomal rearrangements.

We find that three genes on the F element in *D. melanogaster* are absent from the F elements of one or more species: in *D. kikkawai*, *CG31999* has translocated to Muller B; in *D. ananassae* and *D. bipectinata*, *Kif3C* has translocated to Muller C. We were unable to find a *RhoGAP102A* homolog in the genomes of *D. ananassae* and *D. bipectinata*, suggesting that this gene has been lost entirely from the genomes of these two species (Table 2). In addition, we identified several genes that have recently moved onto the Muller F element in at least one of the four expanded F species studied here: these genes represent either translocations or duplication events from other Muller elements. We term these "wanderer genes" and have inferred their ancestral chromosome location via synteny analysis (Table 2). Overall, the vast majority of F element genes (74 of ~80) are F-linked in all species studied here, including *D. melanogaster*. This result ensures that our comparisons of F element gene characteristics across species involve the same pool of genes.

F Element Expansion Occurs Within Both Introns and Intergenic Regions

To identify which F element features have expanded in size in these species, we compared sizes of various genic and intergenic features both within species (ie F element features versus features from the reference D element region) and between species, focusing specifically on the coding span (ie distance from start to stop codon, including introns) of each gene. In *D. melanogaster*, F element genes have larger coding spans

Table 2 Muller F genes duplicated, lost, or translocated across species

Gene	Locus (Muller element)					Inferred	Notes
	Dmel	*Dtak*	*Dkik*	*Dana*	*Dbip*	Ancestral	
PMCA-alpha	F	F	F,F	F	F	F	duplicated on Muller F in *Dkik*
CG31999	F	F	B	F	F	F	translocated to Muller B in *Dkik*
Kif3C	F	F	F	C	C	F	translocated to Muller C in *Dana* and *Dbip*
RhoGAP102A	F	F	F	—	—	F	deleted in *Dana* and *Dbip*
Gar1	C	C	F	F	F	F	translocated to Muller C in *Dmel* and *Dtak*
CG11076	F	A	A	AL	AL	A	translocated to Muller F in *Dmel*
JYalpha	F	D	D	—	—	Y	translocated to Muller F in *Dmel*, deleted in *Dana* and *Dbip*
Mpv17	F	A	A	AL	AL	A	translocated to Muller F in *Dmel*
Mink-alpha	E	E,F	E	E	E	E	duplicated on Muller F in *Dtak*
Gapdh2-alpha	A	A,C	A,C,F	E	E	C	duplicated on Muller F in *Dkik*
RpS16	C	B	C,F	C	C	C	duplicated on Muller F in *Dkik*
mael-alpha	D	D,D	D	D,F	D	D	duplicated on Muller F in *Dana*
CG5028-alpha	E	E	E	E,F	E	E	duplicated on Muller F in *Dana*
CG9922-alpha	E	E	E	E	E,F	E	duplicated on Muller F in *Dbip*
CG43088	A	—	D	—	AR,F	D	duplicated on Muller F in *Dbip*, deleted in *Dtak* and *Dana*
His2A:CG31618	B	B	B	B	C,F	B	translocated and duplicated in *Dbip*, copies on Muller C & F

1. "*-alpha*' designates a duplication event.
2. "AL' and "AR' correspond to the left and right arms of the Muller A element, respectively.

compared to genes in the reference region on the D element (Leung et al. 2015). This same pattern was shown to be exaggerated in *D. ananassae*, indicating that the *D. ananassae* F element genes have expanded in size compared to their *D. melanogaster* orthologs (Leung et al. 2017). Here, we find that, similar to *D. melanogaster*, F element genes within each of the four species have larger coding spans and total intron lengths than their respective D element genes. The F element genes of *D. takahashii*, *D. ananassae*, and *D. bipectinata* also have larger coding spans and total intron lengths compared to the F element genes of *D. melanogaster*. In contrast, the median coding span and intron length of genes on the *D. kikkawai* F element are both similar to those of *D. melanogaster* (Fig. 2a and b). Further investigation on a gene-by-gene basis shows that, across the four expanded F elements, 34 intron-containing genes are either the same size compared to their *D. melanogaster* orthologs or have decreased in size in at least one species (Figure S1a and b). The majority of these (18 genes) are non-expanded only in *D. kikkawai*, which explains why the *D. kikkawai* F element coding spans and intron lengths are not significantly larger than those of *D. melanogaster* (Fig. 2a and b). Eleven other genes are non-expanded in *D. kikkawai* and at least one other species, four of which (*Arf4*, *ATPsynbeta*, *RpS3A*, and *Zip102B*) show either no size expansion or size contraction in all four species, compared to their *D. melanogaster* orthologs (Figure S1c). The five remaining non-expanded genes are distributed across the other three species (Figure S1c).

Comparing *D. melanogaster* F element genes to those on each of the four expanded F elements, there is no difference in the total size of coding sequence (Fig. 2c) or the number of coding exons (Fig. 2d). Finally, all four expanded F elements have larger intergenic regions compared to the *D. melanogaster* F element (Fig. 2e). In fact, intergenic expansion contributes disproportionately to total F element size expansion in all four species, explaining between 52% and 69% of the total expansion (Figure S2). These results suggest that F element expansion is accompanied by an expansion of intergenic sequence as well as an increase in coding span, driven by an increase in intron length.

Notably, several unique aspects of F element genes are present in all five species (*D. melanogaster* plus the four expanded F species). Specifically, the F element genes within all five species consistently have more coding sequence (ie encode larger proteins) (Fig. 2c), more coding exons (Fig. 2d), and more introns (Fig. 2f), compared to the D element reference genes. These measures do not vary significantly among species. Thus, although the ancestral gene content of the F element is biased in terms of exon number and protein size, these properties are not affected by F element size expansion.

Repetitive Elements Drive F Element Expansions

Our analysis of F element genes suggests that repetitive DNA is a major contributor to F element expansion. To more fully characterize the F element repeat landscape in these species, we used *Earl Grey* (Baril et al. 2024) to generate de novo libraries of complex repeats (including TEs) and Tandem Repeats Finder (Benson 1999) to identify simple repeats in each species (see Methods). Roughly a quarter of the *D. melanogaster* F element in the iso-1 reference genome assembly is composed of TEs (Leung et al. 2017). In the other four species, we find that increased F element size generally correlates with increased TE density: *D. kikkawai* (1.7-fold expansion, 57% TE density), *D. takahashii* (2.4-fold expansion, 67% TE density), *D. ananassae* (14.4-fold expansion, 87% TE density), *D. bipectinata* (15.2-fold expansion, 85% TE density).

The repetitive portion of the F element is significantly larger in each expanded F species, compared to *D. melanogaster* (Fisher's Exact Test adjusted $P < 0.05$, Fig. 3a). This expansion of repetitive DNA is driven almost entirely by an increase in TE abundance, with most major TE classes/orders showing significant expansion in each species (Fig. 3a and Table S1). However, the degree of expansion of each TE class/order on the F element differs among species. For example, long terminal repeat (LTR) and long interspersed nuclear element (LINE) TEs show the largest expansion in *D. kikkawai*, *D. bipectinata* and *D. ananassae*, whereas DNA and helitron TEs show the largest expansion in *D. takahashii* (Table S1). In *D. kikkawai*, simple repeats show significant expansion in

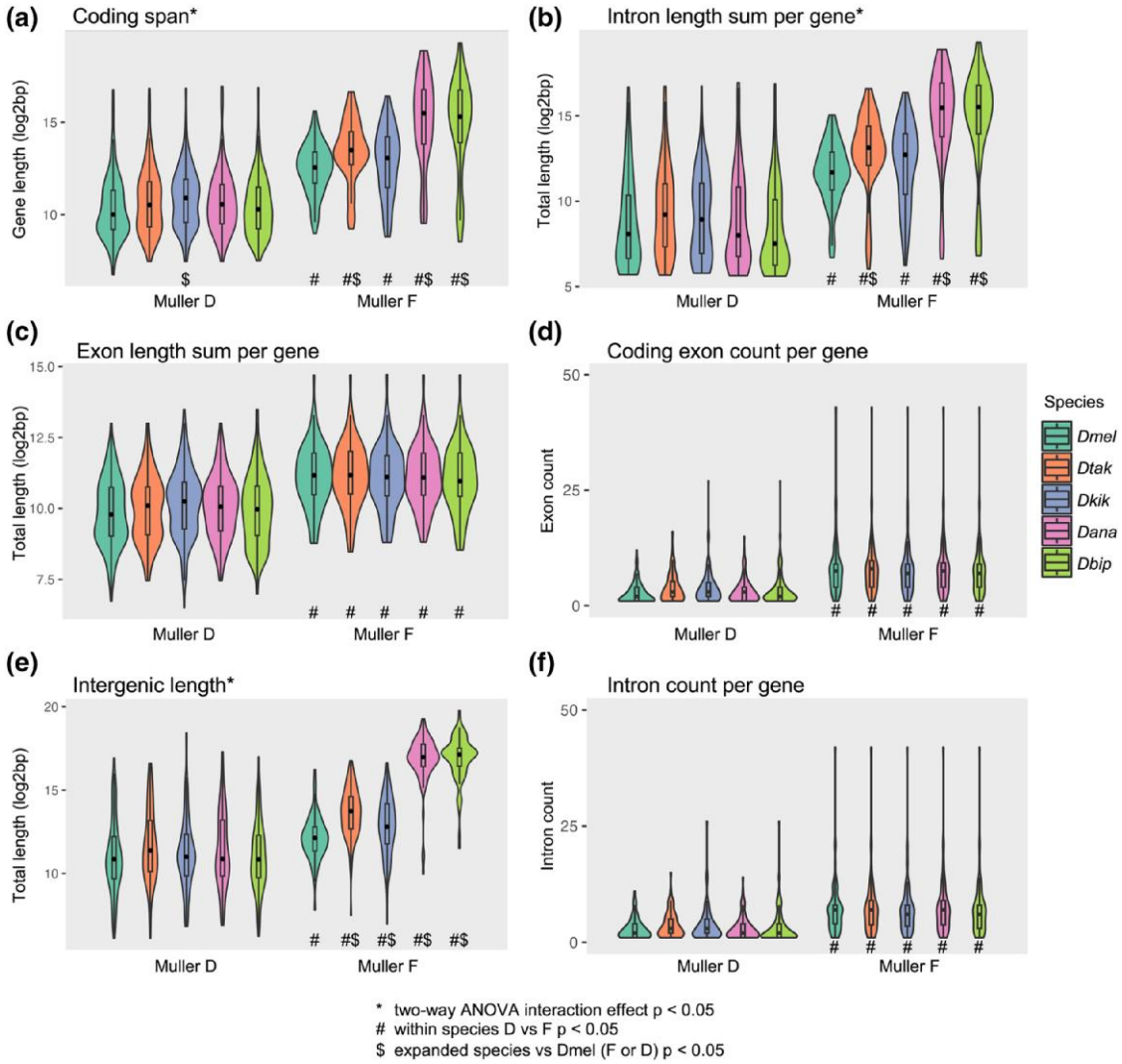


Fig. 2. F element expansion occurs within both introns and intergenic regions. Genes on the Muller F and the reference region of Muller D for each species were annotated by members of the GEP. a) Total coding span of Muller D and Muller F genes from each species. b) Sum of intron lengths per gene. c) Sum of coding exon lengths per gene. d) Total exon count per gene. e) Intergenic lengths. f) Total intron count per gene. For each violin, the dot demarcates the median, the box represents the interquartile range, and the whiskers represent 1.5× the interquartile range. For each panel, an asterisk after the plot header identifies statistically significant interactions between the independent variables chromosome and species (two-way non-parametric analysis of variance, see Methods). For post-hoc analyses (ART-C procedure with Holm correction, see Methods), a pound symbol (#) identifies statistically significant (adjusted $P < 0.05$) comparisons of Muller F to Muller D within a species, while the dollar symbol ($) identifies statistically significant (adjusted $P < 0.05$) comparisons of the same Muller element between each expanded F species and *D. melanogaster*. All tests and their associated statistical values can be found in Table S1. Values for gene coding span length, CDS length, intron length, CDS count, intron count, and intergenic length are available in Table S2.

addition to TEs, but this is not the case in the other three species (Fig. 3a and Table S1). Comparing the fold-change in size in base pairs between each species and *D. melanogaster* reveals that TEs show the largest expansion in three of the four species (Fig. 3b). In *D. kikkawai*, simple repeats show the largest expansion, despite the fact that they occupy less of the F element compared to TEs. These findings are consistent with previous studies that have documented TEs and other repeats as playing a major role in chromosome (and genome) size expansion (Kidwell 2002; Leung et al. 2017; Gebert et al. 2025).

F Element Genes Show Less Optimal Codon Usage

Previous work has demonstrated that, across multiple *Drosophila* species, F element genes consistently show less biased usage of synonymous codons compared to genes on the other autosomes (Vicario et al. 2007; Leung et al. 2015, 2017). To assess how F element expansion affects the usage of synonymous codons, we characterized codon usage bias for F and D element genes using a version of the codon adaptation index (CAI) metric (Sharp and Li 1987) that is normalized to account for biases in nucleotide composition (Puigbò et al. 2008) (see Methods). Within each species, F element genes display a strong reduction in optimal codon usage compared to their respective D element genes (Fig. 4a), which is consistent with previous observations in *D. ananassae* (Leung et al. 2017). Furthermore, genes from the expanded F elements of three species (*D. ananassae*, *D. bipectinata*, and *D. takahashii*) have significantly reduced usage of preferred codons compared to the F element genes of *D. melanogaster*, with the ultra-long F elements of *D. ananassae* and *D. bipectinata* showing the largest reduction in codon usage bias (Fig. 4a). Surprisingly, the *D. kikkawai* F element genes show a significant increase in codon usage bias

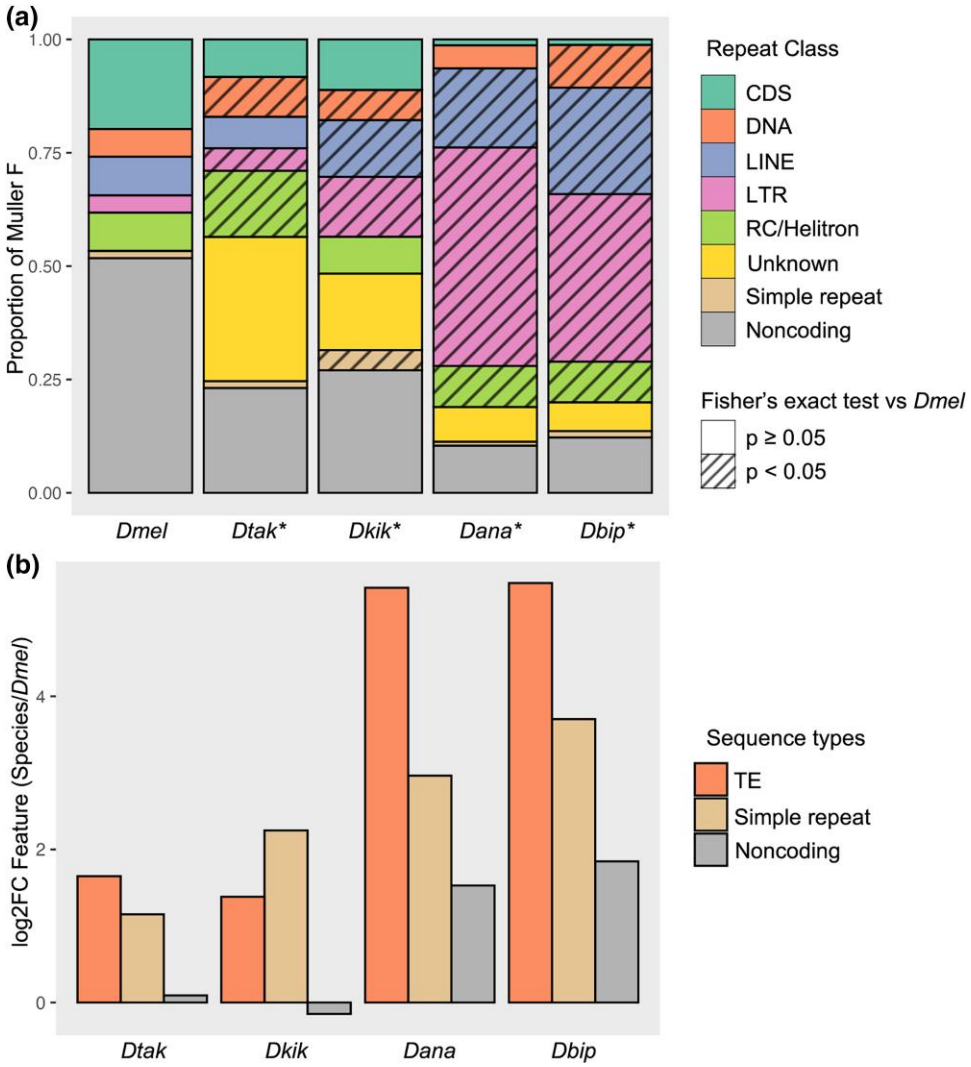


Fig. 3. Repetitive elements drive F element expansions. a) Repetitive and non-repetitive sequence composition of Muller F, by species. Each species with an asterisk has a significantly larger proportion of repetitive DNA on its F element compared to *D. melanogaster* (Fisher's exact test adjusted $P < 0.05$). Diagonal patterning identifies repeat classes in each species that comprise a significantly larger proportion of the F element compared to *D. melanogaster* (Fisher's exact test adjusted $P < 0.05$). Note: Manual inspection of the most abundant elements from the "Unknown" category in *D. takahashii* and *D. kikkawai* suggests that they are primarily DNA transposons. b) Log2-fold change in size (bp) relative to the respective *D. melanogaster* category. Abbreviations: CDS, coding sequence; DNA, DNA transposon; LINE, long interspersed nuclear element retrotransposon; LTR, long terminal repeat retrotransposon; RC/Helitron, rolling circle/Helitron transposon; Unknown, unknown/unclassified repeat, as reported by *Earl Grey* (Baril et al. 2024). The *simple repeat* category is composed of noncoding simple repeats (excluding transposon sequences). The *noncoding* category is composed of unmasked noncoding sequence (see Methods). The *TE* category is composed of DNA, LINE, LTR, and RC/Helitron transposons (excluding unknown/unclassified). All tests and their associated statistical values can be found in Table S1.

compared to *D. melanogaster*. This pattern is not solely due to the subset of *D. kikkawai* genes that have not expanded in size compared to *D. melanogaster*; in fact, these non-expanded genes have a lower median normalized CAI value compared to expanded genes, although the difference is not statistically significant (Figure S3). Overall, these results suggest that, with the exception of *D. kikkawai* (see Discussion), F element size expansion is associated with less efficient selection for preferred codons.

Recombination is reduced on the *D. melanogaster* F element, potentially due to its heterochromatic nature and/or centromere-mediated exclusion of crossing over (ie the "centromere effect" [Beadle 1932]) extending across its relatively short chromosome arm (Sandler and Szauter 1978; McMahan et al. 2013; Hartmann and Sekelsky 2017; Hatkevich et al. 2017; Riddle and Elgin 2018). It is unknown whether recombination occurs within the F elements of the four expanded F species. However, if the centromere effect is the main inhibitor of recombination on the *D. melanogaster* F element, the expanded F elements of these four species may be large enough to escape the centromere effect, at least at their distal ends (Riddle and Elgin 2018). Several previous studies have reported an association between recombination rate and codon usage bias in *Drosophila* (Kliman and Hey 1993; Marais et al. 2001; Campos et al. 2012; Charlesworth and Campos 2014). If recombination occurs on the F elements of these other species at a rate that increases with distance from the centromere, then genes located further from the centromere should show increased codon usage bias. Instead, there is no significant correlation between gene distance from the centromere and normalized CAI in any species (Fig. 4b). Rather, almost all F element genes show similarly low levels of codon usage bias regardless of their distance from the centromere

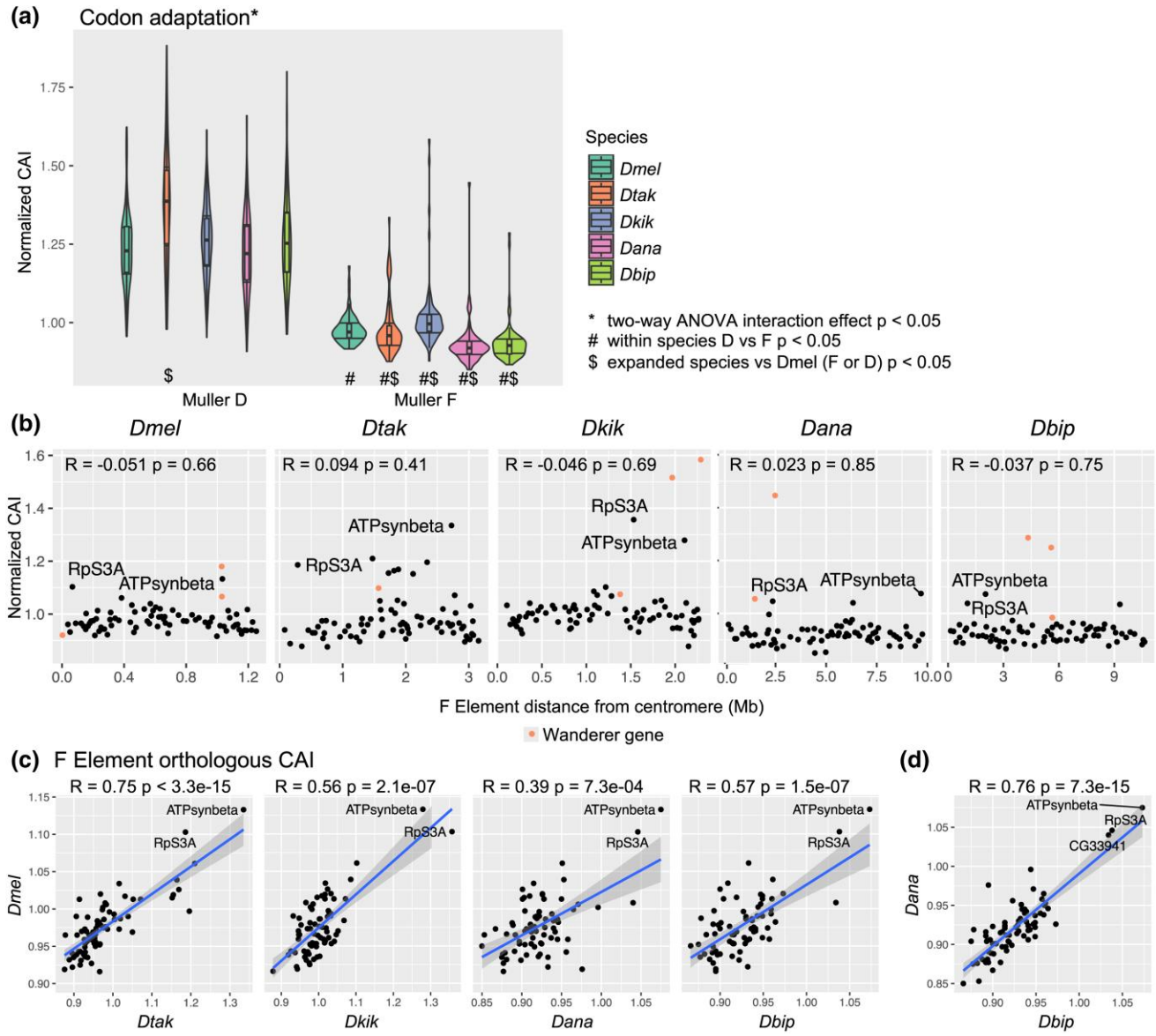


Fig. 4. F element genes show less biased codon usage. Normalized CAI were determined for all genes on the F element and D element reference region for each species (see Methods). a) Normalized CAI for Muller F genes and Muller D reference genes. For each violin, the dot demarcates the median, the box represents the interquartile range, and the whiskers represent 1.5× the interquartile range. b) Normalized CAI values versus distance from centromere. Orange points represent wanderer genes as annotated in Table 2. Correlation coefficients and *P*-values were calculated using Spearman's rank correlation coefficient, excluding wanderer genes. c) Scatterplots comparing normalized CAI values between F element orthologs in the expanded F species against *D. melanogaster*. Correlation coefficients and *P*-values were calculated using Spearman's rank correlation coefficient. d) Scatterplot comparing normalized CAI values between F element orthologs of *D. ananassae* against *D. bipectinata*. Note that a third gene, *CG33941*, shows strong codon usage bias in both *D. ananassae* and *D. bipectinata* but not in the other species. Wanderer genes were excluded from panels (c) and (d). For 4A, the asterisk after the plot header identifies statistically significant interactions between the independent variables chromosome and species (two-way non-parametric analysis of variance). For post-hoc analyses (ART-C procedure with Holm correction), a pound symbol (#) identifies statistically significant (adjusted $P < 0.05$) comparisons of Muller F to Muller D within a species, while the dollar symbol ($) identifies statistically significant (adjusted $P < 0.05$) comparisons of the same Muller element between each expanded F species and *D. melanogaster*. All tests and their associated statistical values can be found in Table S1.

(Fig. 4b). These results suggest that the reduced levels of recombination on the F element are more likely to be due to its packaging as a heterochromatic domain rather than any effect of the centromere.

We next compared normalized CAI values for genes from each of the expanded F species to their orthologs in *D. melanogaster*. We observe significant positive correlations in CAI values of F element orthologs between *D. melanogaster* and each respective species, with the closest relatives (ie *D. melanogaster* and *D. takahashii*; *D. ananassae* and *D. bipectinata*) showing stronger correlations (Fig. 4c and d). Thus, despite the differences in codon usage bias between expanded F species, the relative differences in CAI among F element genes within each species are largely conserved.

Highly expressed genes show stronger codon usage bias, likely due to selection for translational efficiency and/or accuracy (Moriyama and Powell 1997; Plotkin and Kudla 2011; Frumkin et al. 2018). Therefore, the cross-species correlation of CAI among F element genes likely arises from conserved expression levels of genes across species (Figure S4 and Table S3). For example, *ATPsynbeta* and *RpS3A* are among the 5% most highly expressed F element genes in all four

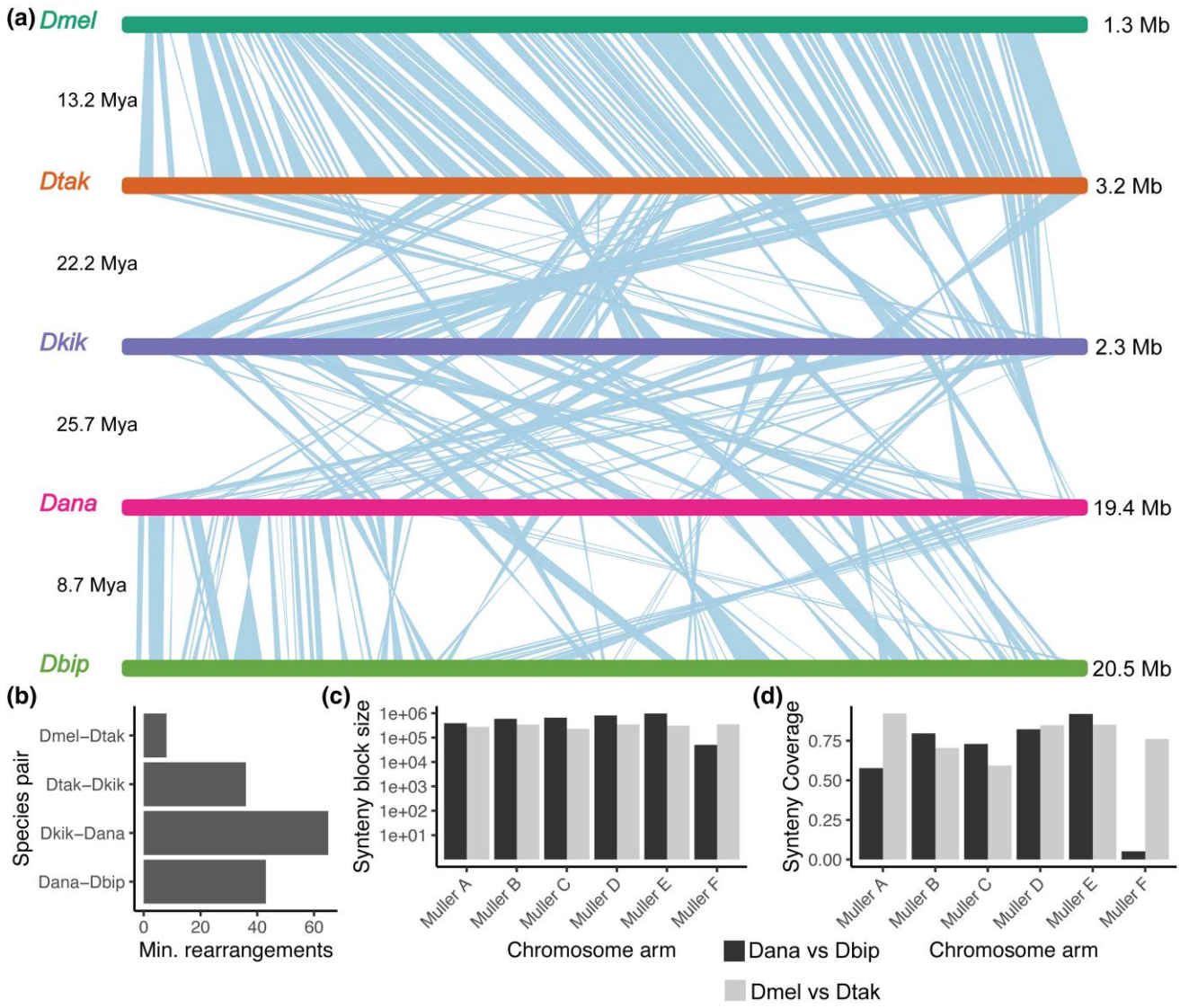


Fig. 5. F element synteny comparison. a) Multi-species visualization of F element synteny. Each link (light blue) connects a pair of orthologous genes between species. Divergence times between species pairs are shown on the left. Note: the F elements of each species are not drawn to scale–their lengths are shown on the right-hand side of the plot. b) The minimum number of chromosomal rearrangements between the species pairs shown in (a), as calculated by *GRIMM* (Tesler 2002). c) Synteny blocks were identified from whole genome alignments between *D. melanogaster* and *D. takahashii,* and between *D. bipectinata* and *D. ananassae.* Mean synteny block size for each Muller element is shown for pairwise comparisons between *ananassae/bipectinata* versus *melanogaster/takahashii.* d) Synteny coverage (ie aligned fraction of chromosome) is shown for the same comparison as in (c).

expanded F species and also show strong codon usage bias across all species, including *D. melanogaster* (Fig. 4b). This indicates that selection may still be responsible for differences in codon usage among F element genes in these species, but the overall efficiency of selection to drive higher levels of codon usage bias is reduced in species with expanded F elements, at least for the three species whose F elements show the largest expansion in size.

D. ananassae and *D. bipectinata* F Elements Show a High Rate of Structural Change

In the genome assemblies of *D. melanogaster* and the four expanded F species, the F element is represented by a single scaffold, allowing us to compare F element gene order across all five species (Fig. 5a). While the F elements of *D. melanogaster* and *D. takahashii* appear completely collinear, the F elements of the other three species are highly rearranged, both compared to each other and to *D. melanogaster* (Fig. 5a and b). The lack of synteny between *D. ananassae* and *D. bipectinata* is especially striking, given that these two species share a more recent common ancestor (divergence time of 8.7 Mya) compared to *D. melanogaster* and *D. takahashii* (divergence time of 13.2 Mya) (Suvorov et al. 2022). These results suggest that the large expansion of TEs and other repeats on the *D. ananassae* and *D. bipectinata* F elements may be causing a higher rate of chromosomal rearrangements on this chromosome.

To further investigate this discrepancy, we generated a whole genome alignment between these four species, which we used to calculate the size, in base pairs, of collinear (ie syntenic) alignment blocks between species pairs. On all other Muller elements, the mean synteny block size from the *D. ananassae* and *D. bipectinata* pairwise alignment is larger than that of *D. melanogaster* and *D. takahashii*, as expected given their more recent divergence time (Fig. 5c). However, the mean *ananassae/bipectinata* F element synteny block size is almost an order of magnitude smaller than that of *melanogaster/*

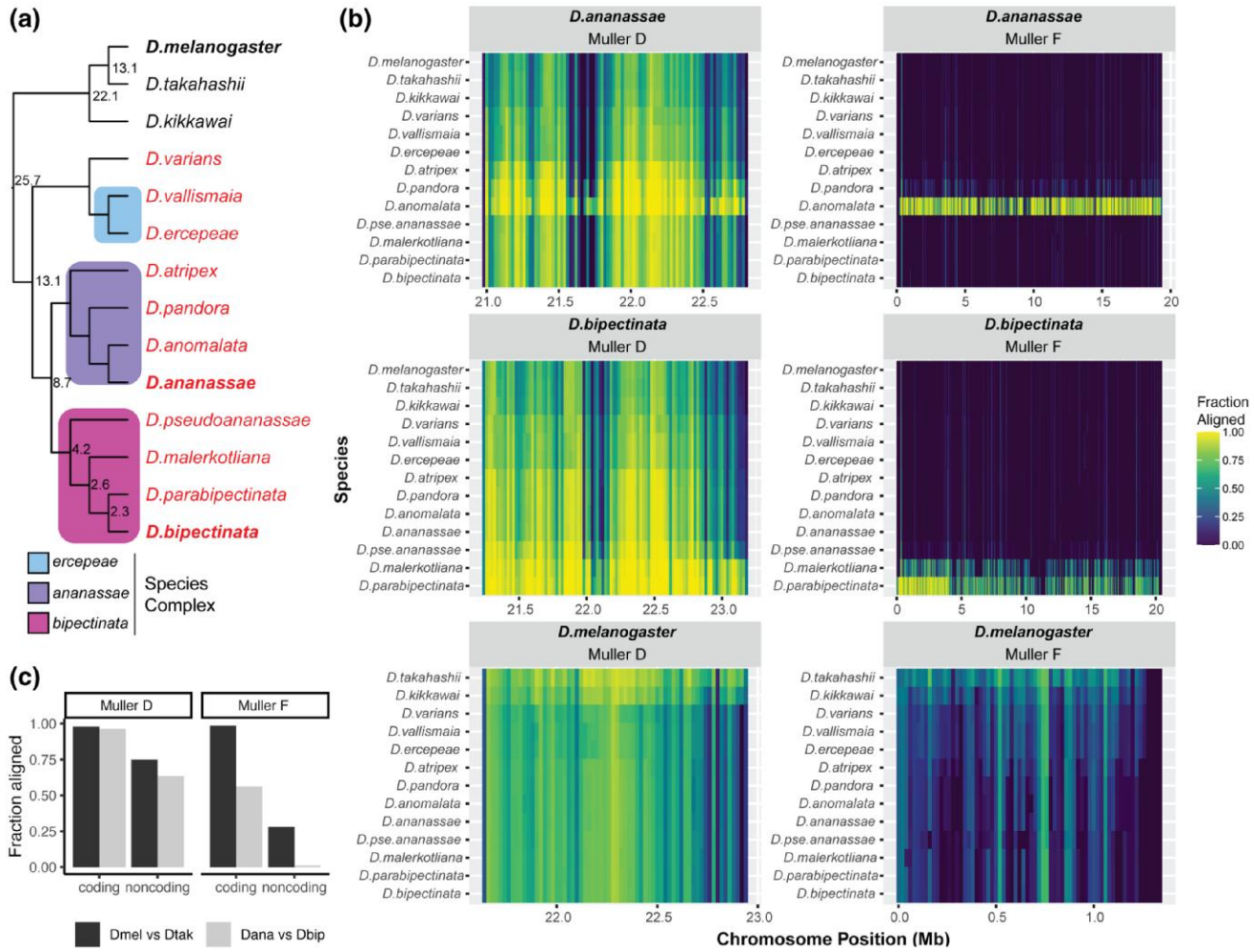


Fig. 6. F element sequence conservation. a) Whole genome alignments were constructed using genome assemblies from the 14 species shown in the phylogenetic tree. Red text indicates the *ananassae* species group, the members of which all have a highly expanded F element. Bold text indicates the focal species used in the conservation analyses shown in (b). Node labels indicate divergence times (in millions of years, where available) from Suvorov et al. (2022). Note that branch lengths are not drawn to scale. b) For each bold focal species shown in (a), pairwise comparisons were made to each of the other species in the tree. For each comparison, the fraction of aligned bases between the two species was calculated for 20 kb windows across either the Muller D reference region (d) or the entire F element (f). Note: *D. pseudoananassae* is abbreviated *D.pse.ananassae*. c) Pairwise alignment rates (*D. melanogaster* versus *D. takahashii* and *D. ananassae* versus *D. bipectinata*) for coding and noncoding regions of the Muller D reference region and the Muller F element.

takahashii (Fig. 5c). Synteny coverage, which represents the aligned fraction of each chromosome for a given species pair, is similar between both species pairs for the autosomal Muller elements B, C, and D (Fig. 5d). The *ananassae/bipectinata* X chromosome (Muller A) is metacentric in both species and contains a large region of pericentromeric heterochromatin (Leung et al. 2023), which likely explains the reduced synteny coverage in this species pair compared to *melanogaster/takahashii*. Most striking, however, is the F element, whose synteny coverage is dramatically reduced in the *ananassae/bipectinata* pairwise alignment compared to all other Muller elements across both the *ananassae/bipectinata* and *melanogaster/takahashii* pairwise alignments (Fig. 5d). These results suggest that, outside of coding regions, the ultra-long F elements of *D. ananassae* and *D. bipectinata* share very little sequence in common.

D. ananassae and *D. bipectinata* F Elements Show High Rates of Sequence Evolution

To further investigate the sequence evolution of the *D. ananassae* and *D. bipectinata* F elements, we sought to analyze additional relatives of these two species. *D. ananassae* and *D. bipectinata* are members of the *ananassae* species group, which contains 21 species divided among three species complexes: *ananassae*, *ercepeae*, and *bipectinata*, plus *D. varians*, which is related to *ercepeae* group species but excluded from the complex (Matsuda et al. 2009; Signor et al. 2013). We used publicly available long-read genome assemblies (Kim et al. 2024) for nine additional species from within the *ananassae* group: three additional species each from the *ananassae* and *bipectinata* complexes, two species from the *ercepeae* complex, and *D. varians* (Fig. 6a). Using gene content to identify F-linked contigs, we estimate that the size of the F element across these nine species ranges from 8 Mb to 22 Mb. These sizes are likely underestimated because the F elements in each assembly are composed of multiple contigs, whereas our size estimates include only contigs containing at least one F element gene. Thus, any F-linked contigs containing only noncoding sequence were not counted towards the total size. Nevertheless, these estimates suggest that all species within the *ananassae* species group have highly expanded F elements compared to *D. melanogaster*,

which is consistent with previous cytological evidence (Matsuda et al. 2009).

To compare F element sequence conservation across this group in more detail, we created a whole genome alignment from the four species studied here, plus the nine additional *ananassae* group species and *D. melanogaster* (see Methods). We selected three focal species: *D. ananassae*, *D. bipectinata*, and *D. melanogaster*, and compared each in turn to the 13 other species in the whole genome alignment. We calculated the percentage of aligned bases in 20 kb windows across the entire F element of each focal species, as well as across their Muller D reference region (Fig. 6b).

For the Muller D element, each focal species shows the highest alignment rate with its closest relative, as expected, and many blocks of aligned sequence are conserved across all 14 species (Fig. 6b). The *D. melanogaster* F element shows a similar pattern, although there are fewer conserved alignment blocks (Fig. 6b). Strikingly, the *D. ananassae* and *D. bipectinata* F elements appear to align only to the F elements of their two closest relatives (*D. ananassae*: *D. anomalata* and *D. pandora*; *D. bipectinata*: *D. parabipectinata* and *D. malerkotliana*) (Fig. 6a and b). This result is consistent with our previous observation that the *D. ananassae* and *D. bipectinata* expanded F elements share little sequence in common (Fig. 5c and d). The sequences that do align are highly enriched for coding sequences of genes. For example, ~56% of *D. ananassae* F element coding sequence is aligned to the *D. bipectinata* F element, whereas only ~1% of F element non-coding sequence aligns (Fig. 6c). There are two possible explanations for these results: (i) the large F elements carried by *D. ananassae* and *D. bipectinata* expanded in size independently, or (ii) the large F elements are evolving so rapidly that their non-coding sequence has completely turned over since their divergence from a common ancestor (see Discussion).

F Elements of the *Ananassae* Species Group Have Distinct Repeat Content

To further investigate the evolution of the noncoding portion of the F element, we assessed conservation of repeat content between all *ananassae* group species plus *D. kikkawai*, *D. takahashii*, and *D. melanogaster*. We used F-specific repeat libraries from each species to mask the assembled F element sequence in each of the other species. Two species whose F elements are populated by the same repeat families will show similar fractions of masked sequence regardless of which of the two repeat libraries are used. These reciprocal repeat masking analyses show that the F elements within each of the *ananassae* group species complexes (ie *ananassae*, *bipectinata*, and *ercepeae*) have similar repeat content. *D. varians* appears to have a subset of distinct repeats that are not found in its closest relatives, *D. vallismaia* and *D. ercepeae*, supporting its exclusion from the *ercepeae* species complex (Fig. 7a). Similarly, members of the *ananassae* species complex share a subset of repeats that are missing from members of the *bipectinata* species complex, and *vice versa*.

To more directly visualize the turnover of F-linked repeats across the *ananassae* group species, we focused on the F element gene *zfh2* and independently annotated all *D. ananassae* and *D. bipectinata* repeats in the introns of this gene across 11 *ananassae* group species. The repeat content of orthologous introns appears to be entirely distinct between species that diverged more than 2 to 3 Mya (Figs. 7b and 6a). We repeated this analysis on six additional F element genes (*Tdg*, *Pur-alpha*, *ND-49*, *Ephrin*, *CG1909*, and *Rad23*) and obtained similar results, suggesting that distinct intronic repeat content is not unique to *zfh2* but is likely a common feature across all or most F element genes in the *ananassae* group (Figure S5). Thus, F element expansion (or repeat turnover) has occurred independently in four different lineages within the *ananassae* group: the *ercepeae*, *bipectinata*, and *ananassae* species complexes, as well as in *D. varians*.

Expanded F Elements Have a Reduced Effective Population Size

Genome size expansion has previously been suggested to be associated with population bottlenecks or other reductions in effective population size (eg Lefébure et al. 2017). We used publicly available population resequencing data for *D. melanogaster*, *D. kikkawai*, *D. ananassae*, and *D. bipectinata* (there are no publicly available population data for *D. takahashii*) to estimate the effective population size (*Ne*) of both the F element and the autosomes (see Methods). We find that the F elements of all three expanded F species show much smaller *Ne* compared to the F element of *D. melanogaster* (Fig. 8a). The *D. kikkawai* and *D. bipectinata* autosomal *Ne* is also smaller than that of *D. melanogaster*, whereas the autosomal *Ne* of *D. ananassae* is larger than that of *D. melanogaster*, consistent with previous work (Grath et al. 2009).

Given the positive correlation between recombination and *Ne* in *Drosophila* (Gossmann et al. 2011), we next sought to determine whether the F elements of these species differ in terms of signatures of recombination. We calculated the mean physical distance at which linkage disequilibrium (LD) decays to ¾ of its maximum value across the F element and across the autosomes of each species (Fig. 8b). Recombination reduces LD between adjacent alleles; therefore, as the rate of recombination increases, LD will decay to background levels across a shorter physical distance. The LD decay results mirror our estimates of *Ne*: for the autosomes, *D. ananassae* has both the largest *Ne* and the fastest rate of LD decay, while *D. kikkawai* has the smallest *Ne* and the slowest rate of decay in LD (Fig. 8a and b). For the F element, *D. melanogaster* has both the largest *Ne* and the fastest rate of LD decay, while *D. ananassae* has the smallest *Ne* and the slowest rate of decay in LD (Fig. 8a and b). The ratio of autosomal versus F element LD decay for each species recapitulates the degree of F element size expansion across species: *D. kikkawai* shows a slightly larger ratio compared to *D. melanogaster*, while the ratios for *D. ananassae* and *D. bipectinata* are much larger (Fig. 8c). Population genetic analyses have previously suggested that recombination does occur on the F element in *Drosophila*, even in *D. melanogaster* (Wang et al. 2002; Arguello et al. 2010). Our results support this previous work and also raise the possibility that a sharp reduction in recombination, specifically on the F element, may explain its large expansion in size in *D. ananassae* and *D. bipectinata*.

Discussion

While genome size evolution has been studied fairly continuously since the C-value paradox was first described in 1971 (Thomas 1971), much less is known about where in the genome this expansion occurs. For example, do all chromosomes increase in size proportionally? Or, given that genome size expansion is largely driven by increases in repetitive DNA, do expanding genomes instead exhibit a large increase in the size of

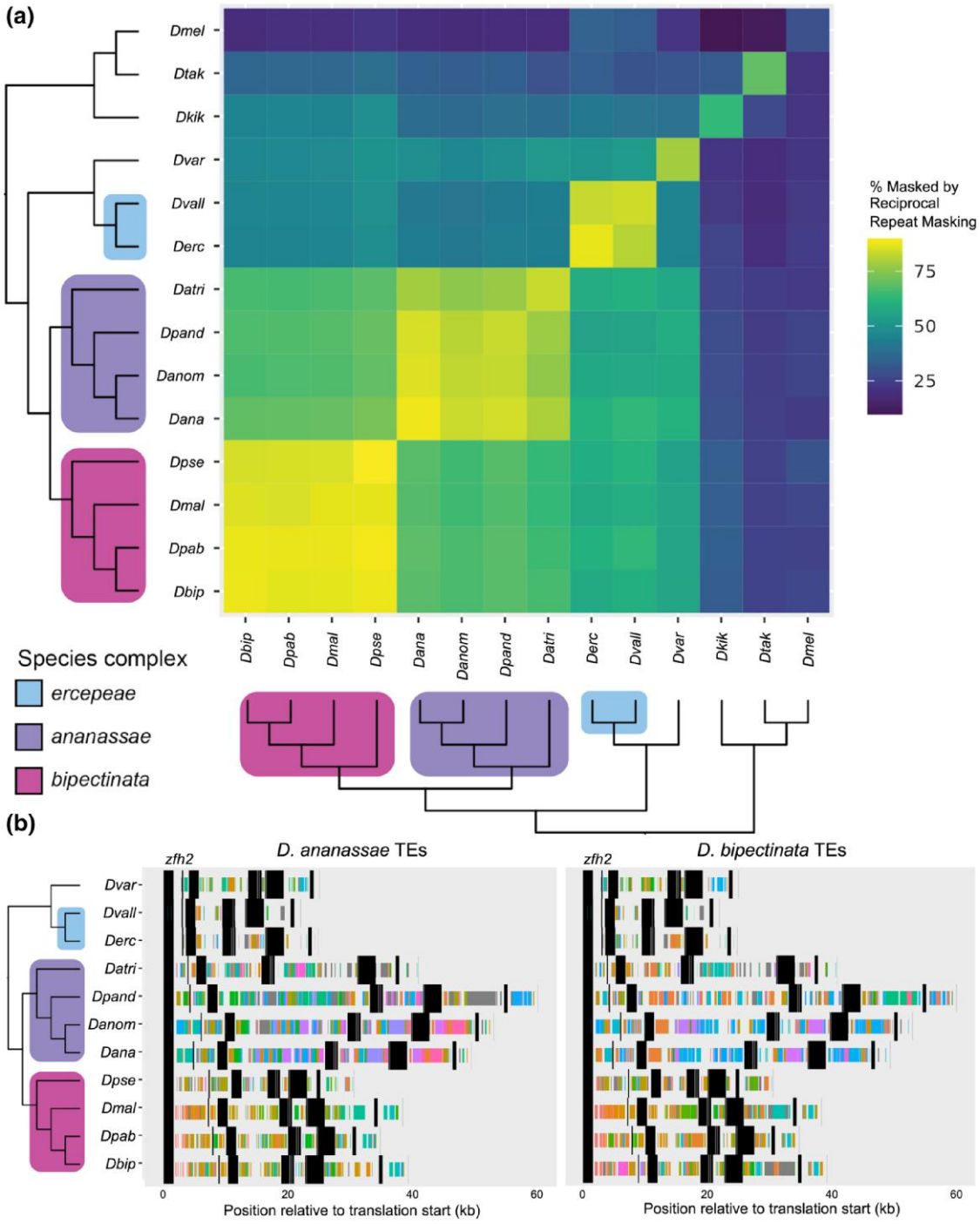


Fig. 7. Distinct repeat content of *ananassae* group F elements. a) Reciprocal repeat masking results for eleven *ananassae* group species plus *D. melanogaster, D. takahashii,* and *D. kikkawai* (see Methods). The F elements of each *ananassae* group species complex (ie *ananassae, bipectinata,* and *ercepeae*) carry a subset of repeats that are not found in the other complexes. The heatmap cells are colored based on the percentage of the F element that is masked. Differences in color along the diagonal are due to differences in F element repeat density among species. b) Visualization of repeat content in the introns of the *zfh2* gene across members of the *ananassae* species group. Black rectangles represent *zfh2* coding exons. Colored rectangles represent matches to individual TE families identified in either *D. ananassae* (left panel) or *D. bipectinata* (right panel) TEs. TE family names and coordinates within *zfh2* for each species are provided in Table S4.

the heterochromatic portion of their genome but little to no increase in the size of euchromatic domains? Our findings show that even species with similar genome sizes can show dramatic differences in chromosome size, suggesting that, in some cases, chromosome-specific expansions may be counterbalanced by shrinkage of other chromosomes, similar to the accordion model of genome size expansion (Kapusta et al. 2017). For example, the heterochromatin of the F element, as well as the pericentromeres of the X chromosome (Muller A), show significant size expansion in *D. ananassae* and *D. bipectinata*, yet their other Muller elements have decreased in size compared to *D. kikkawai* and *D. takahashii* (Figure S6) (Leung et al. 2023).

How is it possible for some chromosomes to expand while others shrink in the same genome? Variation in recombination rate is a logical explanation. In the absence of recombination,

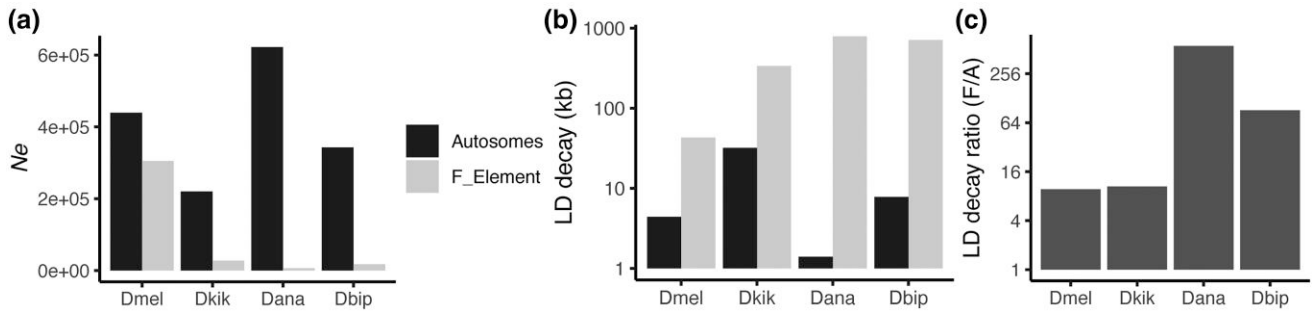


Fig. 8. Effective population size and linkage disequilibrium. a) Estimates of effective population size (*Ne*) for both the autosomes and F elements. b) The mean physical distance at which linkage disequilibrium (LD) decays to ¾ of its maximum value. c) The F element versus autosomal (F/A) ratios of LD decay.

deleterious mutations will irreversibly accumulate in a population (Muller 1964). This phenomenon is commonly referred to as "Muller's ratchet" (Muller 1964; Felsenstein 1974). However, simulations show that even a small amount of recombination is enough to prevent TE accumulation via Muller's ratchet (Dolgin and Charlesworth 2008). It is therefore possible that the F elements of most *Drosophila* species recombine just enough to escape Muller's ratchet. Indeed, population genetic analysis of F-linked genes in *D. melanogaster* and *D. simulans* support this possibility (Wang et al. 2002; Arguello et al. 2010), despite the fact that recombination has never been detected in laboratory crosses (Sandler and Szauter 1978; McMahan et al. 2013; Hartmann and Sekelsky 2017). In the *ananassae* species complex, however, severe reduction or complete loss of recombination on the ancestral F element could provide the mechanism to engage Muller's ratchet, leading to the accumulation of TEs and driving F element expansion. Similarly, a smaller reduction in recombination on the *D. kikkawai* F element could explain its smaller increase in size. Importantly, our estimates of codon usage bias, effective population size, and linkage disequilibrium are consistent with substantially lower recombination rates on the *D. ananassae* and *D. bipectinata* F elements compared to both *D. melanogaster* and *D. kikkawai*.

Our results are likely relevant to other chromosomes as well. Interchromosomal variation in recombination rate among the pericentromeres of other Muller elements could explain why some pericentromeres expand in size (such as those of Muller A in *D. ananassae/bipectinata* and Muller B and C in *D. takahashii*) while others shrink (Figure S6) (Leung et al. 2023). Notably, in *D. melanogaster*, recombination in the pericentromeres does occur, though at a much lower rate than in the euchromatic chromosome arms (Comeron et al. 2012; Kapperud 2023; Koury 2023), and the rate and positions of crossover events are known to vary widely, even among closely related species (Brand et al. 2018; Hawley et al. 2025).

Chromosome size expansion driven by local changes in recombination rate will have important fitness implications for the host genome: the efficiency of natural selection on genes in these regions of the genome will decrease, as we show for codon usage bias (Fig. 4), and chromosomal rearrangements will increase, due to expansions of repetitive DNA (Balachandran et al. 2022; Badet et al. 2024), as seen on the *D. ananassae* and *D. bipectinata* F elements (Fig. 5). However, there is evidence that purifying selection still operates on these chromosomes. For example, there are four intron-containing F element genes whose size has remained unchanged or even decreased in all four expanded F species, compared to *D. melanogaster* (Figure S1). Although there are no obvious shared functions or expression patterns among these four genes, their lack of expansion suggests that purifying selection is acting to prevent accumulation of TE insertions and other repeat expansions in the introns of these genes. Interestingly, in *D. kikkawai*, whose F element has expanded the least of the species studied here, almost half (45%, 36 genes; Figure S1) of all F element genes are either the same size or smaller relative to their *D. melanogaster* orthologs. This observation raises the possibility that the process of chromosome size expansion does not occur randomly but instead in a stepwise manner across groups of genes where the fitness cost of size expansion is similar. Transcription of genes with large introns is energetically costly, time-consuming, and prone to splicing errors (Fox-Walsh et al. 2005; Fingerhut et al. 2019; Fontan et al. 2025). Thus, we would expect the first group of genes to expand in size to be those where intronic TE insertions are either neutral or only slightly deleterious. In support of this prediction, the F element genes that have expanded in size in *D. kikkawai* are expressed at significantly lower levels compared to non-expanded F element genes (Figure S7).

Three of our findings support a scenario where the *D. kikkawai* F element is in the early stages of size expansion: (i) its F element is only modestly (ie 1.7-fold) larger than the *D. melanogaster* F element; (ii) almost half of the *D. kikkawai* F element genes are either the same size or smaller than their *D. melanogaster* orthologs; (iii) *D. kikkawai* F element genes show a significant increase in optimal codon usage compared to *D. melanogaster* (Fig. 4). This latter observation is surprising given that the *D. kikkawai* F element has both a smaller effective population size and a slower rate of LD decay compared to *D. melanogaster* (Fig. 8). However, both our *Ne* and LD estimates arise from short-term *D. kikkawai* population dynamics. For example, our *Ne* estimate covers the past ~3,000 years (see Methods), while codon usage biases evolve over millions of years and therefore reflect long-term *Ne* (Jensen and Bachtrog 2011). Thus, a recent decrease in selection efficiency on the *D. kikkawai* F element, potentially caused by a decrease in recombination rate, would explain both our *Ne*/LD and codon usage estimates while also explaining why this chromosome appears to be in the early stages of size expansion.

The most recent common ancestor of all *Drosophila* species likely carried a small dot-like F element (Leung et al. 2015). Compared to extant species with dot-like F elements

(ie *D. erecta*, *D. mojavensis*, and *D. grimshawi*) (Leung et al. 2015) as well as species with slightly expanded F elements (ie *D. takahashii* and *D. kikkawai*), the ultra-long F elements of *D. ananassae* and *D. bipectinata* stand out, not only due to their size, but also their extremely low usage of optimal codons and their extremely high repeat densities and rates of chromosomal rearrangement. These properties are unusual compared to other F elements, but quite common among Y chromosomes. For example, in multiple *Drosophila* species, Y-linked genes tend to show lower usage of optimal codons compared to F element genes (Singh et al. 2014). Notably, *D. ananassae* exhibits the opposite pattern (*D. bipectinata* was not studied) (Singh et al. 2014). Both within and among closely related *Drosophila* species, Y chromosomes have also been shown to accumulate extremely high densities of repetitive elements while undergoing rapid changes in repeat content and gene order (Mahajan et al. 2018; Chang et al. 2022; Nguyen et al. 2022). Thus, in certain ways, the F elements of *D. ananassae* and *D. bipectinata* are more similar to a Y chromosome than they are to the F elements of other *Drosophila* species.

While rare among *Drosophila* species, ultra-long F elements are not limited to the *ananassae* species group. Ultra-long F elements have been identified based on cytogenetic evidence in several Hawaiian *Drosophila* species (Clayton 1988; Craddock et al. 2016) as well as the *repleta* group species *D. mercatorum* (Sperling et al. 2025). Recent work in *D. mercatorum* proposes that F element expansion is associated with the ability of this species to undergo parthenogenesis (Sperling et al. 2023, 2025). Intriguingly, *D. ananassae* is also known to undergo facultative parthenogenesis (Futch 1972, 1973), raising the possibility that chromosome size expansion could influence life history traits in *Drosophila*.

Our findings have important implications for genome evolution beyond the F element. Chromosome-specific size expansion may be relatively common, given that we find multiple examples of this phenomenon among the four species we study here, not including the F element (eg Muller A in *D. ananassae*/*D.bipectinata* and Muller B/C in *D. takahashii*, Figure S7). If so, this phenomenon may be an important contributor to organismal fitness, assuming size expansion of other chromosomes leads to less efficient selection and accumulation of repetitive elements and chromosomal rearrangements, as we see in the F elements of *D. ananassae* and *D. bipectinata*. Chromosome-scale genome assemblies (Leung et al. 2023) enabled our insights into chromosome-specific size expansions that would have otherwise been obscured by size reductions in other chromosomes. This masking effect likely contributes to why chromosome size variation remains understudied. As more such assemblies become available (eg Gebert et al. 2025), we will be better equipped to determine whether variation in chromosome size is even more prevalent than variation in total genome size.

Methods

Version numbers and command-line options for all software packages are available on GitHub: https://github.com/Ellison-Lab/4genomes_Felement.

Genome Assembly and Annotation

Strain information and approaches used to construct the genome assemblies for *D. takahashii*, *D. kikkawai*, *D. ananassae*, and *D. bipectinata* have been previously described (Leung et al. 2023). The annotations for all genes on the F element scaffold (and Muller D reference region) of each species were hand-curated by undergraduate student participants in the Genomics Education Partnership (GEP) (Shaffer et al. 2010) with guidance from their faculty mentors at their home institutions. GEP student annotators identified *D. melanogaster* orthologs for each gene via protein *BLAST* searches conducted on the FlyBase website (Öztürk-Çolak et al. 2024).

Annotators then used a mirror of the UCSC Genome Browser to examine different lines of experimental (eg RNA-Seq) and computational evidence (eg sequence similarity to *D. melanogaster* protein sequences, gene predictions, splice junction predictions) to construct gene models across the four species. The paired-end RNA-Seq data from adult females, adult males, and mixed embryos used in the gene annotation process were produced by the modENCODE project (Graveley et al. 2011); individual NCBI BioProject accessions are listed in Table S5. The evidence tracks derived from these RNA-Seq datasets included RNA-Seq read coverage, splice junction predictions from *RegTools* (Cotto et al. 2023), and assembled transcripts from *StringTie2* (Trapnell et al. 2010; Kovaka et al. 2019) and *Oases* (Schulz et al. 2012).

Each gene annotation project was completed by at least two GEP student annotators working independently. Reconcilers are undergraduate students, or recent graduates, with extensive training in gene annotation. Working with supervision from GEP faculty and staff, they reconciled these gene annotations using *Apollo 2* (Dunn et al. 2019) to create the gene models analyzed in this study. In some cases, such as when genes were missing two independent student annotations, reconcilers performed initial gene annotation in addition to reconciliation. The Annotation Instruction Sheet (available on the GEP web site at https://thegep.org/projects/felement/) describes the protocols for handling cases where the annotations were ambiguous because of insufficient evidence. Additional curricula (eg walkthroughs, presentations) that describe the GEP annotation protocol are available under the "Curriculum" section of the GEP web site (https://thegep.org/curriculum/).

Orientation of Muller F Scaffolds

For this study, we used the *D. kikkawai*, *D. takahashii*, *D. ananassae*, and *D. bipectinata* genome assemblies described in Leung et al (2023), each of which has Muller F assembled as a single scaffold. We then performed additional analyses to confirm that the Muller F scaffolds are oriented correctly within each assembly. The Muller F element in the *D. melanogaster* reference genome assembly is oriented such that the centromere is at the beginning of the scaffold and the telomere is at the end, including a head-to-tail array of *HTT* telomeric retrotransposons. The *D. takahashii* Muller F is almost perfectly collinear with the *D. melanogaster* F element, and we therefore oriented this scaffold based on synteny. One end of the *D. kikkawai* F element contains a ~95 kb array of tandem repeats composed of two large DNA satellites (640 bp and 1.9 kb, respectively), which potentially represents centromeric or pericentromeric sequence. We were unable to identify telomeric retrotransposons at the other end of the chromosome scaffold; however, we noted the presence of the genes *Mitf*, *Arf4*, *CG33521*, *PIP4K*, *pho*, and *Kif3C* near this end, all of which are also near the telomere in the *D. melanogaster* Muller F scaffold. We therefore oriented

the scaffold such that the satellite array is at the beginning and the genes with telomeric orthologs in *D. melanogaster* are at the end.

The F elements of *D. ananassae* and *D. bipectinata* are metacentric. The two chromosome arms were previously oriented based on Hi-C contact maps (Leung et al. 2023). We used a combination of synteny, putative centromeric tandem repeats, and telomeric retrotransposons to determine whether the arms were oriented correctly. At the beginning of the *D. ananassae* Muller F scaffold, we identified orthologs of *Arf4*, *Mitf*, *PIP4K* and *CG33521*. We also found the ortholog of *D. melanogaster yellow-h*, which is near the telomere in *D. kikkawai*, and the orthologs of *D. melanogaster Dyrk3* and *Cadps*, which are near the telomere of the *D. melanogaster* F element. Interestingly, the other end of the *D. ananassae* F element scaffold contains the genes *Zyx*, *ATPsynbeta*, *PlexA*, *toy*, *CaMKII*, and *apolpp*, which are located in the subtelomeric regions of either the *D. melanogaster* or *D. kikkawai* F elements, or both. Together, these results suggest that both arms of the *D. ananassae* F element were indeed oriented correctly during Hi-C scaffolding and raise the possibility that there is selection to maintain the telomeric location of a subset of genes on the F element. Most *Drosophila* species have telomere-specialized *Jockey* retrotransposons at their chromosome ends (Villasante et al. 2007; Saint-Leandre et al. 2019; Son et al. 2025). We identified a head-to-tail array of *Jockey* clade retrotransposons at the beginning of the *D. ananassae* F scaffold, which likely represent telomeric retrotransposons, as well as a ~100 kb array of simple tandem repeats in the middle of the scaffold, which may mark the location of the centromere.

We then compared synteny between *D. ananassae* and *D. bipectinata* F element scaffolds, which suggested that the two chromosome arms of *D. bipectinata* were previously oriented incorrectly. In the middle of the *D. bipectinata* F element scaffold is a large array of tandem repeats that may mark the location of the centromere. There are gaps on either side of the array where the chromosome arms were scaffolded together. Orthologs of the telomeric *D. melanogaster* genes *Cadps*, *Dyrk3*, and *PIP4K* were near this putative centromere in the previous version of the assembly (Leung et al. 2023); however, in the new version of the assembly used here, we have inverted this arm, placing these genes at the telomere. This arrangement places *Rad23* and *pho* near one telomeric end, and *bt*, *Cadps*, *Dyrk3*, and *PIP4K* near the opposite end, adjacent to the telomeric TE array.

Reciprocal Repeat Masking

Reciprocal repeat masking was performed using *RepeatMasker* along with the full *Earl Grey* repeat library from each species shown in Fig. 7 to mask the F element sequences of all other species shown in Fig. 7. The percentage of masked sequence was calculated by dividing the number of masked base pairs by the total number of F element base pairs.

Gene Feature Analysis

For the various calculations regarding coding span, coding exons, introns, and intergenic regions, we identified in each species the transcript ortholog corresponding to the principal splice isoform as defined in the APPRIS database (Rodriguez et al. 2022) using the BDGP6 assembly (Hoskins et al. 2015) and Ensembl release 110 annotation (Harrison et al. 2024). For any gene with more than one splice isoform marked principal, the longest splice isoform was selected as the representative transcript for that gene. Two-way non-parametric analysis of variance tests were performed using the *ARTool* package (Wobbrock et al. 2011) with default options, with chromosome and species as independent variables. Post hoc pairwise comparisons were performed using the *ART-C* (Aligned Rank Transform for Contrasts) procedure (Elkin et al. 2021) implemented in the *art.con()* function, with *P*-values adjusted for multiple comparisons using the Holm method. For Fig. 3, noncoding sequences were identified as regions that do not overlap annotated coding sequences or repeats identified by RepeatMasker and Tandem Repeats Finder, respectively. Note that the "intergenic" category includes the untranslated regions of F and D element genes because these regions were not annotated during the manual curation process. Genes were considered to be non-expanded if their coding span was less than 1.1-fold the length of their ortholog in *D. melanogaster*.

Repeat Identification and Annotation

We created de novo repeat libraries using *Earl Grey* (Baril et al. 2024) and used *RepeatMasker* (www.repeatmasker.org) to identify repeat locations within each genome assembly. We used *Tandem Repeats Finder* (Benson 1999) to identify simple repeats in each assembly. For each expanded F species, the F-specific repeat library was constructed by extracting the subset of repeat consensus sequences from the species-specific *Earl Grey* repeat library that have one or more matches on the F element. To annotate F element genes in additional *ananassae* group species, we used *LiftOn* (Chao et al. 2025) with options "-copies -sc 0.95" and our hand-curated *D. ananassae* F element gene annotations as a reference.

Codon Usage Analysis

To calculate the CAI for each species, we used the top 10% of genes in each species with the lowest effective number of codons (ENC) as the reference gene set, similar to Vicario et al. (2007). To identify these genes, we used the NCBI RefSeq annotations for each species and the longest isoform per gene. We used the *codon-bias* package (Diament 2025) to calculate ENC and to export the reference gene set for each species. We used the options within the *codon-bias* package that allow for calculating ENC in a way that corrects for both background nucleotide composition (Novembre 2002) and sequence length (Sun et al. 2013). Biased nucleotide and amino acid composition can distort the CAI measure. We therefore used the approach described in Puigbò et al. (2008), implemented in the *codonuse* package (https://github.com/s-andrews/codonuse), to calculate the expected CAI (eCAI) for each gene, which is the upper limit of CAI when codon usage bias results only from mutational bias (Puigbò et al. 2008). The *codonuse* package calculates the normalized CAI for each gene by dividing its true CAI by its eCAI. Thus, selection for preferred codons is likely to be contributing to codon usage bias in genes with normalized CAI values larger than one, whereas the codon usage bias of genes with normalized CAI values less than one can be explained mainly by mutational bias (Puigbò et al. 2008).

F Element Synteny and Conservation

To infer the ancestral chromosome location of each gene in Table 2, we used a combination of prior literature (Carvalho

and Clark 2013), synteny, and parsimony. In cases where the ancestral chromosome was ambiguous based on the five species in Table 2, we identified the chromosome locations of homologs in *D. miranda* and *D. virilis*.

We used *NGenomeSyn* (He et al. 2023) to create the F element synteny maps shown in Fig. 5a and *GRIMM* (Tesler 2002) to infer the minimum number of rearrangement events between the F elements of species pairs in Fig. 5b.

We used *progressive cactus* (Armstrong et al. 2020) to create whole genome alignments and the *HAL* utilities *halSynteny* (Krasheninnikova et al. 2020) to identify synteny blocks and calculate synteny coverage and *halAlignmentDepth* to calculate the fraction of aligned bases between species (Hickey et al. 2013).

Effective Population Size and Linkage Disequilibrium

We obtained Illumina population resequencing data for *D. melanogaster* (Sprengelmeyer et al. 2020), *D. kikkawai* (Li et al. 2022), *D. ananassae* (Choi et al. 2015), and *D. bipectinata* (Li et al. 2022). We used *bowtie2* (Langmead and Salzberg 2012) to align Illumina reads to their corresponding reference genome, sorted read alignments and marked duplicates with *sambamba* (Tarasov et al. 2015), and used *freebayes* (Garrison and Marth 2012) to identify nucleotide variants. We filtered *freebayes* variants to include only biallelic sites using *vcffilter* from *vcflib* (Garrison et al. 2022).

We used *SMC++* (Terhorst et al. 2017) and the filtered *freebayes* variants to infer population size histories for each species. We then calculated the harmonic mean of effective population size estimates from generations 3,600 to 36,000 (corresponding to 300 - 3,000 years ago, assuming 12 generations per year), as in Bergeron et al. (2023).

To examine the decay of linkage disequilibrium, we used *vcftools* (Danecek et al. 2011) to calculate LD as the squared correlation coefficient between genotypes for all pairs of nucleotide variants within 1 Mb of each other on the F element and within 500 kb of each other on the other autosomes. We then calculated the mean LD for 1,000 bp bins (F element) and 100 bp bins (autosomes) and calculated the bin size at which LD decreases to ¾ of its maximum value. We note that, although it is more common to calculate LD decay as the distance at which LD decays to half of its maximum value (eg Kim et al. 2007), we use ¾ here because F element LD plateaus in some species before reaching that point.

Supplementary material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

The authors acknowledge the Office of Advanced Research Computing (OARC) at Rutgers, The State University of New Jersey, for providing access to the Amarel cluster and associated research computing resources that have contributed to the results reported here. The authors acknowledge New Jersey City University, Department of Biology, as the host institution which provided support to the reconcilers. C.E.E. and T.J.S. thank Dr. Grace Lee for insightful discussion of these results. The authors also acknowledge the contributions of the late Dr. James Youngblom, whose students completed gene annotation reports for this project, and thank Dr. Janey Youngblom, who facilitated the review and approval of this manuscript by these students after his passing. The authors also thank the additional students who contributed gene annotation effort to this project, but for various reasons did not participate in manuscript critique and revision.

Funding

This work was supported by the National Science Foundation grant 2114661 to C.J.A. and the National Institutes of Health grants R01GM130698 to C.E.E., R35GM152168 to C.E.E., and K12GM093854 to T.J.S. The Genomics Education Partnership is supported by National Science Foundation grant 1915544 to L.K.R. and National Institutes of Health grant R25GM130517 to L.K.R.

Data Availability

Version numbers and command-line options for all software packages are available on GitHub: https://github.com/Ellison-Lab/4genomes_Felement. The curated gene annotations are available as evidence tracks ('GEP/*[species]*HiC1' for *D. takahashii*, *D. kikkawai*, and *D. ananassae*; "GEP/DbipHiC2' for *D. bipectinata*) on the GEP UCSC Genome Browser mirror (https://gander.wustl.edu). Associated Muller F and D gene annotations (GFF format), repeat annotations (RepeatMasker format), repeat libraries (FASTA format), and gene order and synteny annotations are available in the Figshare repository https://doi.org/10.6084/m9.figshare.30251578. The *D. melanogaster* release 6 assembly was produced by the Berkeley *Drosophila* Genome Project (BDGP; [Hoskins et al. 2015]) and the release 6.55 gene annotations for *D. melanogaster* used here are available on GitHub (https://github.com/Ellison-Lab/4genomes_Felement). Values for gene coding span length, CDS length, intron length, CDS count, intron count, intergenic length, gene size fold change/*D. melanogaster*, and normalized CAI are available in Table S2.

References

Arguello JR *et al.* Recombination yet inefficient selection along the *Drosophila melanogaster* subgroup's fourth chromosome. *Mol Biol Evol.* 2010:27:848–861. https://doi.org/10.1093/molbev/msp291.

Armstrong J *et al.* Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature.* 2020:587:246–251. https://doi.org/10.1038/s41586-020-2871-y.

Badet T, Tralamazza SM, Feurtey A, Croll D. Recent reactivation of a pathogenicity-associated transposable element is associated with major chromosomal rearrangements in a fungal wheat pathogen. *Nucleic Acids Res.* 2024:52:1226–1242. https://doi.org/10.1093/nar/gkad1214.

Baimai V, Chumchong C. Karyotype variation and geographic distribution of the three sibling species of the *Drosophila kikkawai* complex. *Genetica.* 1980:54:113–120. https://doi.org/10.1007/BF00055979.

Balachandran P *et al.* Transposable element-mediated rearrangements are prevalent in human genomes. *Nat Commun.* 2022:13:7115. https://doi.org/10.1038/s41467-022-34810-8.

Baril T, Galbraith J, Hayward A. Earl grey: a fully automated user-friendly transposable element annotation and analysis pipeline. *Mol Biol Evol.* 2024:41:msae068. https://doi.org/10.1093/molbev/msae068.

Bast J *et al.* No accumulation of transposable elements in asexual arthropods. *Mol Biol Evol.* 2016:33:697–706. https://doi.org/10.1093/molbev/msv261.

Beadle GW. A possible influence of the spindle fibre on crossing-over in *Drosophila*. *Proc Natl Acad Sci U S A*. 1932:18:160–165. https://doi.org/10.1073/pnas.18.2.160.

Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999:27:573–580. https://doi.org/10.1093/nar/27.2.573.

Bergeron LA *et al*. Evolution of the germline mutation rate across vertebrates. *Nature*. 2023:615:285–291. https://doi.org/10.1038/s41586-023-05752-y.

Bhutkar A *et al*. Chromosomal rearrangement inferred from comparisons of 12 *Drosophila* genomes. *Genetics*. 2008:179:1657–1680. https://doi.org/10.1534/genetics.107.086108.

Blommaert J. Genome size evolution: towards new model systems for old questions. *Proc Biol Sci*. 2020:287:20201441. https://doi.org/10.1098/rspb.2020.1441.

Bosco G, Campbell P, Leiva-Neto JT, Markow TA. Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. *Genetics*. 2007:177:1277–1290. https://doi.org/10.1534/genetics.107.075069.

Brand CL, Cattani MV, Kingan SB, Landeen EL, Presgraves DC. Molecular evolution at a meiosis gene mediates species differences in the rate and patterning of recombination. *Curr Biol*. 2018:28:1289–1295.e4. https://doi.org/10.1016/j.cub.2018.02.056.

Campos JL, Charlesworth B, Haddrill PR. Molecular evolution in nonrecombining regions of the *Drosophila melanogaster* genome. *Genome Biol Evol*. 2012:4:278–288. https://doi.org/10.1093/gbe/evs010.

Carvalho AB, Clark AG. Efficient identification of Y chromosome sequences in the human and *Drosophila* genomes. *Genome Res*. 2013:23:1894–1907. https://doi.org/10.1101/gr.156034.113.

Chak STC, Harris SE, Hultgren KM, Jeffery NW, Rubenstein DR. Eusociality in snapping shrimps is associated with larger genomes and an accumulation of transposable elements. *Proc Natl Acad Sci U S A*. 2021:118:e2025051118. https://doi.org/10.1073/pnas.2025051118.

Chang C-H, Gregory LE, Gordon KE, Meiklejohn CD, Larracuente AM. Unique structure and positive selection promote the rapid divergence of *Drosophila* Y chromosomes. *Elife*. 2022:11:e75795. https://doi.org/10.7554/eLife.75795.

Chao K-H *et al*. Combining DNA and protein alignments to improve genome annotation with LiftOn. *Genome Res*. 2025:35:311–325. https://doi.org/10.1101/gr.279620.124.

Charlesworth B, Campos JL. The relations between recombination rate and patterns of molecular variation and evolution in *Drosophila*. *Annu Rev Genet*. 2014:48:383–403. https://doi.org/10.1146/annurev-genet-120213-092525.

Chen J *et al*. Genomic diversity generated by a transposable element burst in a rice recombinant inbred population. *Proc Natl Acad Sci U S A*. 2020:117:26288–26297. https://doi.org/10.1073/pnas.2015736117.

Choi JY, Bubnell JE, Aquadro CF. Population genomics of infectious and integrated *Wolbachia pipientis* genomes in *Drosophila ananassae*. *Genome Biol Evol*. 2015:7:2362–2382. https://doi.org/10.1093/gbe/evv158.

Clayton FE. The role of heterochromatin in karyotype variation among Hawaiian picture-winged *Drosophila*. *Pac Sci*. 1988:42:28–47. https://core.ac.uk/download/pdf/5093761.pdf.

Comeron JM, Ratnappan R, Bailin S. The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genet*. 2012:8:e1002905. https://doi.org/10.1371/journal.pgen.1002905.

Corradi N, Pombert J-F, Farinelli L, Didier ES, Keeling PJ. The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. *Nat Commun*. 2010:1:77. https://doi.org/10.1038/ncomms1082.

Cotto KC *et al*. Integrated analysis of genomic and transcriptomic data for the discovery of splice-associated variants in cancer. *Nat Commun*. 2023:14:1589. https://doi.org/10.1038/s41467-023-37266-6.

Craddock EM, Gall JG, Jonas M. Hawaiian *Drosophila* genomes: size variation and evolutionary expansions. *Genetica*. 2016:144:107–124. https://doi.org/10.1007/s10709-016-9882-5.

Cui R *et al*. Relaxed selection limits lifespan by increasing mutation load. *Cell*. 2019:178:385–399.e20. https://doi.org/10.1016/j.cell.2019.06.004.

Danecek P *et al*. The variant call format and VCFtools. *Bioinformatics*. 2011:27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Diament A. codon-bias (python package). Zenodo 2025. https://zenodo.org/records/15017252

Dolgin ES, Charlesworth B. The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics*. 2008:178:2169–2177. https://doi.org/10.1534/genetics.107.082743.

Dunn NA *et al*. Apollo: democratizing genome annotation. *PLoS Comput Biol*. 2019:15:e1006790. https://doi.org/10.1371/journal.pcbi.1006790.

Elkin LA, Kay M, Higgins JJ, Wobbrock JO. 2021. An aligned rank transform procedure for multifactor contrast tests. In: The 34th Annual ACM Symposium on User Interface Software and Technology. Virtual Event USA Vol. 5. ACM. p. 754–768.

Elliott TA, Gregory TR. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philos Trans R Soc Lond B Biol Sci*. 2015:370:20140331. https://doi.org/10.1098/rstb.2014.0331.

Felsenstein J. The evolutionary advantage of recombination. *Genetics*. 1974:78:737–756. https://doi.org/10.1093/genetics/78.2.737.

Fernández P *et al*. A 160 Gbp fork fern genome shatters size record for eukaryotes. *iScience*. 2024:27:109889. https://doi.org/10.1016/j.isci.2024.109889.

Fingerhut JM, Moran JV, Yamashita YM. Satellite DNA-containing gigantic introns in a unique gene expression program during *Drosophila* spermatogenesis. *PLoS Genet*. 2019:15:e1008028. https://doi.org/10.1371/journal.pgen.1008028.

Fontan A, Lannes R, Fingerhut JM, Flynn JM, Yamashita YM. 2025. Defective splicing of Y-chromosome-linked gigantic genes contributes to hybrid male sterility in *Drosophila* [preprint]. bioRxiv 655455. https://doi.org/10.1101/2025.05.22.655455.

Fox-Walsh KL *et al*. The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc Natl Acad Sci U S A*. 2005:102:16176–-16181. https://doi.org/10.1073/pnas.0508489102.

Frumkin I *et al*. Codon usage of highly expressed genes affects proteome-wide translation efficiency. *Proc Natl Acad Sci U S A*. 2018:115:E4940–E4949. https://doi.org/10.1073/pnas.1719375115.

Futch DG. A preliminary note on parthenogenesis in *D. ananassae*. *Drosoph. Inf. Serv*. 1972:48:78.

Futch DG. Parthenogenesis in Samoan *Drosophila ananassae* and *Drosophila pallidosa*. *Genetics*. 1973:74:s86–s87.

Garrison E, Kronenberg ZN, Dawson ET, Pedersen BS, Prins P. A spectrum of free software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2, hts-nim and slivar. *PLoS Comput Biol*. 2022:18:e1009123. https://doi.org/10.1371/journal.pcbi.1009123.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing [preprint], arXiv, arXiv:1207.3907. http://arxiv.org/abs/1207.3907.

Gebert D *et al*. Analysis of 30 chromosome-level *Drosophila* genome assemblies reveals dynamic evolution of centromeric satellite repeats. *Genome Biol*. 2025:26:63. https://doi.org/10.1186/s13059-025-03527-4.

Gossmann TI, Woolfit M, Eyre-Walker A. Quantifying the variation in the effective population size within a genome. *Genetics*. 2011:189:1389–1402. https://doi.org/10.1534/genetics.111.132654.

Grath S, Baines JF, Parsch J. Molecular evolution of sex-biased genes in the *Drosophila ananassae* subgroup. *BMC Evol Biol*. 2009:9:291. https://doi.org/10.1186/1471-2148-9-291.

Graveley BR *et al*. The developmental transcriptome of *Drosophila melanogaster*. *Nature*. 2011:471:473–479. https://doi.org/10.1038/nature09715.

Gregory TR, Johnston JS. Genome size diversity in the family Drosophilidae. *Heredity (Edinb)*. 2008:101:228–238. https://doi.org/10.1038/hdy.2008.49.

Harrison PW *et al.* Ensembl 2024. *Nucleic Acids Res*. 2024:52:D891–D899. https://doi.org/10.1093/nar/gkad1049.

Hartmann MA, Sekelsky J. The absence of crossovers on chromosome 4 in *Drosophila melanogaster*: imperfection or interesting exception? *Fly (Austin)*. 2017:11:253–259. https://doi.org/10.1080/19336934.2017.1321181.

Hatkevich T *et al.* Bloom syndrome helicase promotes meiotic crossover patterning and homolog disjunction. *Curr Biol*. 2017:27:96–102. https://doi.org/10.1016/j.cub.2016.10.055.

Hawley RS *et al.* Patterns of crossover distribution in *Drosophila mauritiana* necessitate a re-thinking of the centromere effect on crossing over. *Genetics*. 2025:230:iyaf039. https://doi.org/10.1093/genetics/iyaf039.

He W *et al.* NGenomeSyn: an easy-to-use and flexible tool for publication-ready visualization of syntenic relationships across multiple genomes. *Bioinformatics*. 2023:39:btad121. https://doi.org/10.1093/bioinformatics/btad121.

Hickey G, Paten B, Earl D, Zerbino D, Haussler D. HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics*. 2013:29:1341–1342. https://doi.org/10.1093/bioinformatics/btt128.

Hjelmen CE, Blackmon H, Holmes VR, Burrus CG, Johnston JS. Genome size evolution differs between *Drosophila* subgenera with striking differences in male and female genome size in Sophophora. *G3 (Bethesda)*. 2019:9:3167–3179. https://doi.org/10.1534/g3.119.400560.

Hoskins RA *et al.* The release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Res*. 2015:25:445–458. https://doi.org/10.1101/gr.185579.114.

Jensen JD, Bachtrog D. Characterizing the influence of effective population size on the rate of adaptation: Gillespie's Darwin domain. *Genome Biol Evol*. 2011:3:687–701. https://doi.org/10.1093/gbe/evr063.

Kapheim KM *et al.* Social evolution. Genomic signatures of evolutionary transitions from solitary to group living. *Science*. 2015:348:1139–1143. https://doi.org/10.1126/science.aaa4788.

Kapperud Ø. Recombination and sterility in inversion homo- and heterokaryotypes under a general counting model of chiasma interference. *Genetics*. 2023:225:iyad120. https://doi.org/10.1093/genetics/iyad120.

Kapusta A, Suh A, Feschotte C. Dynamics of genome size evolution in birds and mammals. *Proc Natl Acad Sci U S A*. 2017:114:E1460–E1469. https://doi.org/10.1073/pnas.1616702114.

Kidwell MG. Transposable elements and the evolution of genome size in eukaryotes. *Genetica*. 2002:115:49–63. https://doi.org/10.1023/A:1016072014259.

Kim BY *et al.* Single-fly genome assemblies fill major phylogenomic gaps across the Drosophilidae tree of life. *PLoS Biol*. 2024:22:e3002697. https://doi.org/10.1371/journal.pbio.3002697.

Kim S *et al.* Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat Genet*. 2007:39:1151–1155. https://doi.org/10.1038/ng2115.

Kliman RM, Hey J. Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Mol Biol Evol*. 1993:10:1239–1258. https://doi.org/10.1093/oxfordjournals.molbev.a040074.

Koury SA. Predicting recombination suppression outside chromosomal inversions in *Drosophila melanogaster* using crossover interference theory. *Heredity (Edinb)*. 2023:130:196–208. https://doi.org/10.1038/s41437-023-00593-x.

Kovaka S *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol*. 2019:20:278. https://doi.org/10.1186/s13059-019-1910-1.

Krasheninnikova K *et al.* halSynteny: a fast, easy-to-use conserved synteny block construction method for multiple whole-genome alignments. *Gigascience*. 2020:9:giaa047. https://doi.org/10.1093/gigascience/giaa047.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012:9:357–359. https://doi.org/10.1038/nmeth.1923.

Lefébure T *et al.* Less effective selection leads to larger genomes. *Genome Res*. 2017:27:1016–1028. https://doi.org/10.1101/gr.212589.116.

Leung W *et al. Drosophila* Muller F elements maintain a distinct set of genomic properties over 40 million years of evolution. *G3 (Bethesda)*. 2015:5:719–740. https://doi.org/10.1534/g3.114.015966.

Leung W *et al.* Retrotransposons are the major contributors to the expansion of the *Drosophila ananassae* Muller F element. *G3 (Bethesda)*. 2017:7:2439–2460. https://doi.org/10.1534/g3.117.040907.

Leung W *et al.* Long-read genome assemblies for the study of chromosome expansion: *Drosophila kikkawai*, *Drosophila takahashii*, *Drosophila bipectinata*, and *Drosophila ananassae*. *G3 (Bethesda)*. 2023:13:jkad191. https://doi.org/10.1093/g3journal/jkad191/7249150.

Li F *et al.* Phylogenomic analyses of the genus Drosophila reveals genomic signals of climate adaptation. *Mol Ecol Resour*. 2022:22:1559–1581. https://doi.org/10.1111/1755-0998.13561.

Li X *et al.* Chromosome size in diploid eukaryotic species centers on the average length with a conserved boundary. *Mol Biol Evol*. 2011:28:1901–1911. https://doi.org/10.1093/molbev/msr011.

Lu J, Clark AG. Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in *Drosophila*. *Genome Res*. 2010:20:212–227. https://doi.org/10.1101/gr.095406.109.

Lynch M, Conery JS. The origins of genome complexity. *Science*. 2003:302:1401–1404. https://doi.org/10.1126/science.1089370.

Mackintosh A *et al.* The determinants of genetic diversity in butterflies. *Nat Commun*. 2019:10:3466. https://doi.org/10.1038/s41467-019-11308-4.

Mahajan S, Wei KH-C, Nalley MJ, Gibilisco L, Bachtrog D. De novo assembly of a young *Drosophila* Y chromosome using single-molecule sequencing and chromatin conformation capture. *PLoS Biol*. 2018:16:e2006348. https://doi.org/10.1371/journal.pbio.2006348.

Marais G, Mouchiroud D, Duret L. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. *Proc Natl Acad Sci U S A*. 2001:98:5688–5692. https://doi.org/10.1073/pnas.091427698.

Marino A, Debaecker G, Fiston-Lavier A-S, Haudry A, Nabholz B. Effective population size does not explain long-term variation in genome size and transposable element content in animals. *eLife*. 2024:13:RP100574.

Matsuda M, Ng C-S, Doi M, Kopp A, Tobari YN. Evolution in the *Drosophila ananassae* species subgroup. *Fly (Austin)*. 2009:3:157–169. https://doi.org/10.4161/fly.8395.

McMahan S, Kohl KP, Sekelsky J. Variation in meiotic recombination frequencies between allelic transgenes inserted at different sites in the *Drosophila melanogaster* genome. *G3 (Bethesda)*. 2013:3:1419–1427. https://doi.org/10.1534/g3.113.006411.

Mérel V *et al.* The worldwide invasion of *Drosophila suzukii* is accompanied by a large increase of transposable element load and a small number of putatively adaptive insertions. *Mol Biol Evol*. 2021:38:4252–4267. https://doi.org/10.1093/molbev/msab155.

Mérel V, Tricou T, Burlet N, Haudry A. Relaxed purifying selection is associated with an accumulation of transposable elements in flies. *Mol Biol Evol*. 2025:42:1–15. https://doi.org/10.1093/molbev/msaf111.

Moriyama EN, Powell JR. Codon usage bias and tRNA abundance in *Drosophila*. *J Mol Evol*. 1997:45:514–523. https://doi.org/10.1007/PL00006256.

Muller HJ. The relation of recombination to mutational advance. *Mutat Res*. 1964:106:2–9. https://doi.org/10.1016/0027-5107(64)90047-8.

Naville M *et al.* Massive changes of genome size driven by expansions of non-autonomous transposable elements. *Curr Biol*. 2019:29:1161–-1168.e6. https://doi.org/10.1016/j.cub.2019.01.080.

Nguyen AH, Wang W, Chong E, Chatla K, Bachtrog D. Transposable element accumulation drives size differences among polymorphic Y chromosomes in *Drosophila*. *Genome Res*. 2022:32:1074–1088. https://doi.org/10.1101/gr.275996.121.

Novembre JA. Accounting for background nucleotide composition when measuring codon usage bias. *Mol Biol Evol*. 2002:19:1390–1394. https://doi.org/10.1093/oxfordjournals.molbev.a004201.

Oggenfuss U *et al*. A population-level invasion by transposable elements triggers genome expansion in a fungal pathogen. *Elife*. 2021:10:e69249. https://doi.org/10.7554/eLife.69249.

Öztürk-Çolak A *et al*. FlyBase: updates to the *Drosophila* genes and genomes database. *Genetics*. 2024:227:iyad211. https://doi.org/10.1093/genetics/iyad211.

Petrov DA. Mutational equilibrium model of genome size evolution. *Theor Popul Biol*. 2002:61:531–544. https://doi.org/10.1006/tpbi.2002.1605.

Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet*. 2011:12:32–42. https://doi.org/10.1038/nrg2899.

Puigbò P, Bravo IG, Garcia-Vallvé S. E-CAI: a novel server to estimate an expected value of Codon adaptation Index (eCAI). *BMC Bioinformatics*. 2008:9:65. https://doi.org/10.1186/1471-2105-9-65.

Riddle NC, Elgin SCR. The *Drosophila* dot chromosome: where genes flourish amidst repeats. *Genetics*. 2018:210:757–772. https://doi.org/10.1534/genetics.118.301146.

Roddy AB, Alvarez-Ponce D, Roy SW. Mammals with small populations do not exhibit larger genomes. *Mol Biol Evol*. 2021:38:3737–3741. https://doi.org/10.1093/molbev/msab142.

Rodriguez JM *et al*. APPRIS: selecting functionally important isoforms. *Nucleic Acids Res*. 2022:50:D54–D59. https://doi.org/10.1093/nar/gkab1058.

Saint-Leandre B, Nguyen SC, Levine MT. Diversification and collapse of a telomere elongation mechanism. *Genome Res*. 2019:29:920–931. https://doi.org/10.1101/gr.245001.118.

Sandler L, Szauter P. The effect of recombination-defective meiotic mutants on fourth-chromosome crossing over in *Drosophila melanogaster*. *Genetics*. 1978:90:699–712. https://doi.org/10.1093/genetics/90.4.699.

Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*. 2012:28:1086–1092. https://doi.org/10.1093/bioinformatics/bts094.

Shaffer CD *et al*. The genomics education partnership: successful integration of research into laboratory classes at a diverse group of undergraduate institutions. *CBE Life Sci Educ*. 2010:9:55–69. https://doi.org/10.1187/09-11-0087.

Sharp PM, Li WH. The codon adaptation Index–a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*. 1987:15:1281–1295. https://doi.org/10.1093/nar/15.3.1281.

Signor S, Seher T, Kopp A. Genomic resources for multiple species in the *Drosophila ananassae* species group. *Fly (Austin)*. 2013:7:47–57. https://doi.org/10.4161/fly.22353.

Singh ND, Koerich LB, Carvalho AB, Clark AG. Positive and purifying selection on the Drosophila Y chromosome. *Mol Biol Evol*. 2014:31:2612–2623. https://doi.org/10.1093/molbev/msu203.

Son JH *et al*. 2025. Convergence and conflict among telomere specialized transposons across 60 million years of Drosophilid evolution. [preprint]. bioRxiv 658640. https://doi.org/10.1101/2025.06.09.658640.

Sperling AL, Fabian DK, Garrison E, Glover DM. A genetic basis for facultative parthenogenesis in *Drosophila*. *Curr Biol*. 2023:33:3545–3560.e13. https://doi.org/10.1016/j.cub.2023.07.006.

Sperling AL, Fabian DK, Glover DM. 2025. Expansion of chromosome F heterochromatin in parthenogenetic *Drosophila mercatorum*. [preprint]. bioRxiv 660940. https://doi.org/10.1101/2025.06.22.660940.

Sprengelmeyer QD *et al*. Recurrent collection of *Drosophila melanogaster* from wild African environments and genomic insights into species history. *Mol Biol Evol*. 2020:37:627–638. https://doi.org/10.1093/molbev/msz271.

Sun X, Yang Q, Xia X. An improved implementation of effective number of codons (N_c). *Mol Biol Evol*. 2013:30:191–196. https://doi.org/10.1093/molbev/mss201.

Suvorov A *et al*. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. *Curr Biol*. 2022:32:111–123.e5. https://doi.org/10.1016/j.cub.2021.10.052.

Talla V *et al*. Rapid increase in genome size as a consequence of transposable element hyperactivity in wood-white (Leptidea) butterflies. *Genome Biol Evol*. 2017:9:2491–2505. https://doi.org/10.1093/gbe/evx163.

Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics*. 2015:31:2032–2034. https://doi.org/10.1093/bioinformatics/btv098.

Terhorst J, Kamm JA, Song YS. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat Genet*. 2017:49:303–309. https://doi.org/10.1038/ng.3748.

Tesler G. Efficient algorithms for multichromosomal genome rearrangements. *J Comput Syst Sci*. 2002:65:587–609. https://doi.org/10.1016/S0022-0000(02)00011-9.

Thomas CA Jr. The genetic organization of chromosomes. *Annu Rev Genet*. 1971:5:237–256. https://doi.org/10.1146/annurev.ge.05.120171.001321.

Trapnell C *et al*. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010:28:511–515. https://doi.org/10.1038/nbt.1621.

Vicario S, Moriyama EN, Powell JR. Codon usage in twelve species of *Drosophila*. *BMC Evol Biol*. 2007:7:226. https://doi.org/10.1186/1471-2148-7-226.

Vicoso B, Bachtrog D. Numerous transitions of sex chromosomes in Diptera. *PLoS Biol*. 2015:13:e1002078. https://doi.org/10.1371/journal.pbio.1002078.

Villasante A *et al*. *Drosophila* telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase. *Genome Res*. 2007:17:1909–1918. https://doi.org/10.1101/gr.6365107.

Wang W, Thornton K, Berry A, Long M. Nucleotide variation along the *Drosophila melanogaster* fourth chromosome. *Science*. 2002:295:134–137. https://doi.org/10.1126/science.1064521.

Wang X, Morton JA, Pellicer J, Leitch IJ, Leitch AR. Genome downsizing after polyploidy: mechanisms, rates and selection pressures. *Plant J*. 2021:107:1003–1015. https://doi.org/10.1111/tpj.15363.

Whitney KD, Boussau B, Baack EJ, Garland T Jr. Drift and genome complexity revisited. *PLoS Genet*. 2011:7:e1002092. https://doi.org/10.1371/journal.pgen.1002092.

Whitney KD, Garland T Jr. Did genetic drift drive increases in genome complexity? *PLoS Genet*. 2010:6:e1001080. https://doi.org/10.1371/journal.pgen.1001080.

Wobbrock JO, Findlater L, Gergle D, Higgins JJ. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM. https://doi.org/10.1145/1978942.1978963.

Yang H *et al*. 2024. Consistent accumulation of transposable elements in species of the Hawaiian *Tetragnatha* spiny-leg adaptive radiation across the archipelago chronosequence. bioRxiv 574070. https://doi.org/10.1101/2024.01.03.574070.

Zhou W, Liang G, Molloy PL, Jones PA. DNA methylation enables transposable element-driven genome expansion. *Proc Natl Acad Sci U S A*. 2020:117:19359–19366. https://doi.org/10.1073/pnas.1921719117.