Reorganization of 3D genome structure in the *Drosophila melanogaster* species group

Nicole S. Torosin[1], Aparna Anand[1], Tirupathi Rao Golla[1], Weihuan Cao[1], Christopher E. Ellison[1]

April 2020

[1]Department of Genetics, Rutgers University, Piscataway, NJ 08854

Abstract

Topologically associating domains, or TADs, are functional units that organize chromosomes into 3D structures of interacting chromatin. TADs play an important role in regulating gene expression by constraining enhancer-promoter contacts; there is evidence that deletion of TAD boundaries leads to aberrant expression of neighboring genes. While the mechanisms of TAD formation have been well-studied, current knowledge on the extent of TAD conservation across species is inconclusive. Due to the integral role TADs play in gene regulation, their structure and organization is expected to be conserved during evolution. However, more recent research suggests that TAD structures diverge relatively rapidly. We use Hi-C chromosome conformation capture to measure evolutionary conservation of whole TADs and TAD boundary elements between *D. melanogaster* and *D. triauraria*, two early-branching species from the melanogaster species group which diverged ∼28 million years ago. We found that 75% of TAD boundaries are orthologous while only 25% of TAD domains are conserved and these are enriched for Polycomb-repressed chromatin. Our results show that TADs have been reorganized since the common ancestor of *D. melanogaster* and *D. triauraria*, yet the sequence elements that specify TAD boundaries remain highly conserved. We propose that evolutionary divergence in 3D genome organization results from shuffling of conserved boundary elements across chromosomes, breaking old TADs and creating new TAD architectures. This result supports the existence of distinct TAD subtypes: some may be evolutionarily flexible while others remain highly conserved due to their importance in restricting gene-regulatory element interactions.

1 Introduction

The recent development of Hi-C sequencing techniques has allowed inference of three-dimensional chromosome conformation through identification of inter- and intra-chromosomal interactions at high-resolution, across the entire genome. Visualization of gene contacts and frequencies led to the discovery of organizational features called topologically associating domains, or TADs, which bring genes in close proximity with their regulatory elements [30]. TADs are regions of highly interacting chromatin that contain genes with similar expression patterns and epigenetic states [8, 45]. Domains are demarcated by boundaries which are regions of decompacted chromatin bound by insulator proteins [46]. In vertebrates, CCCTC-binding factor (CTCF) along with the structural maintenance of chromosomes (SMC) cohesin complex, plays a major role in specifying TAD boundaries [40, 8, 50, 37], whereas in *Drosophila*, CTCF and SMC binding show little enrichment at TAD boundaries. Instead, other insulator proteins, including BEAF-32, Chromator, CP190, and M1BP are more frequently found at TAD boundaries [45, 39, 24, 23, 49] and depletion of M1BP has been shown to disrupt 3D genome organization in the *Drosophila* Kc167 cell line [39].

Most research thus far investigating 3D genome structure has operated under the prevailing theory that TADs regulate gene expression by limiting potential gene-enhancer interactions to those within a given domain. This theory is supported by a variety of studies. For example, genome-wide enhancer-promoter contacts in mouse neurons occur mainly within TADs [4] and reporter gene-enhancer interactions have been shown to be correlated with TAD structure [47]. Furthermore, disruption of TAD boundaries has been associated with aberrant enhancer/promoter contacts, gene misregulation, developmental abnormalities and cancer [32, 16, 15, 21, 33, 52].

The functional role of TADs with respect to the regulation of gene expression has important implications for 3D genome evolution. If TAD structure is critical for proper gene regulation, then the evolution of 3D genome organization should be highly constrained and related species should show strong conservation of TAD structures. Consistent with this prediction, a variety of studies in vertebrates have reported strong conservation of 3D genome organization using comparative Hi-C approaches [8, 50, 26, 28] . A recent study in *Drosophila* reported that 3D genome architecture is conserved over 40 million years of evolution in spite of extensive chromosomal rearrangements [41]. These studies support a model where chromosomal rearrangements that preserve TADs (i.e. their breakpoints are located within TAD boundaries) are much more likely to be retained over evolutionary time compared to rearrangements that disrupt TADs (Figure 1): Under this model, TADs are shuffled as whole units over evolutionary time due to selection to maintain the 3D interaction properties of genes and regulatory sequences within them.

However, more recent research suggests that TADs may be frequently reorganized over evolutionary time. Notably, one recent study found that only 43% of TADs are shared between humans and chimpanzees [13]. Furthermore, there are a number of studies showing that gene expression profiles remain unperturbed upon TAD reorganization. The extensive changes in chromosome topology caused by the rearrangements found on *Drosophila* balancer chromosomes is not associated with differences in gene expression [18]. Also in *Drosophila*, studies involving deletion mutations [29] and experimentally-induced inversions [34] found that these mutations, which should disrupt TAD organization, had little effect on gene expression. Similar observations have been made in mice: fusion of TADs does not have major effects on gene expression [6]. These studies suggest that TADs may diverge relatively rapidly over evolutionary time, with little effect on gene expression.

There are several possibilities that might explain these apparently contradictory results. Some studies [8] only assess the conservation of TAD boundaries, rather than TADs themselves. It is possible that boundary elements evolve at a different rate than TADs. It is also possible that there are distinct functional subtypes of TADs, with some being more tolerant of reorganization than others. Consistent with such a possibility, a recent study identified a subset of ancient, highly-conserved

TADs in both vertebrates and flies that are enriched for conserved noncoding elements and developmental genes [20].

Here, we have compared 3D genome organization between *Drosophila melanogaster* and *Drosophila triauraria*, which diverged ∼28 million years ago [3]. We chose this species pair because they have accumulated extensive chromosomal rearrangements since their divergence, yet maintain blocks of synteny that are roughly the same size as TADs (∼120 Kb on average, see Results). We have improved a previously published *D. triauraria* genome assembly [35] by performing additional nanopore sequencing and Hi-C scaffolding which yielded chromosome-length scaffolds. We then used two biological replicates of Hi-C sequencing data to identify high-confidence TADs and TAD boundaries in each species. We separately assessed the evolutionary conservation of boundary elements and complete TADs and found that TAD boundaries are significantly more conserved between these two species than the TADs themselves. Overall, we find that only 25% of TADs are orthologous and these conserved TADs are enriched for Polycomb-repressed chromatin, similar to what has been observed by Harmston et al. [20]. Our results show that TADs have been reorganized since the common ancestor of *D. melanogaster* and *D. triauraria*, yet the sequence elements that specify TAD boundaries remain highly conserved. We propose that evolutionary divergence in 3D genome organization results from shuffling of conserved boundary elements across chromosomes, breaking old TADs and creating new TAD architectures. Our results also support the existence of functionally distinct TADs subtypes: many TADs may be evolutionarily flexible and able to be reorganized without perturbing gene expression, whereas there may also be a distinct set of TADs subtypes that remain highly conserved due to their importance in restricting gene-regulatory element interactions.

2 Results

2.1 *D. triauraria* genome assembly

A recently published genome assembly for D. triauraria was made using relatively low-coverage (depth 18.8x, N50 = 0.72 Mb) nanopore sequencing data. In order to create an improved assembly, we performed additional long-read nanopore sequencing of genomic DNA extracted from ∼30 adult females from *D. triauraria* strain 14028-0651.00 (National Drosophila Species Stock Center at Cornell). We used three r9.4 flow cells to generate a total of 633,844 reads (10,287 bp mean length) and 6.5 Gb of sequencing data. We combined our data with previously published nanopore data from *D. triauraria* [35] for a final dataset of 1,043,600 reads (10.5 Gb). We basecalled the raw signal data with *Albacore* and assembled the basecalled reads with *Canu* [25], which produced an assembly with contig N50 of 1.3 Mb (1098 total contigs, 269 Mb total size). We then polished the assembly by using *Nanopolish* [31] with raw nanopore signal data and *Pilon* [51] with Illumina data, which corrected a total of 1,185,510 assembly errors. Next, we used *Purge Haplotigs* [42] to identify allelic contigs, where highly heterozygous haplotypes were assembled as separate contigs rather than collapsed. After removing secondary haplotigs and bacterial contigs, our final contig assembly consisted of a total of 294 contigs which sum to ∼200 Mb and have an N50 of 1.7 Mb, which is 240% larger than that generated in a previous study [35] (Table S1).

We next performed Hi-C scaffolding using the polished nanopore contigs and the software packages *Juicer* and *3D-DNA* [10, 9] (Figure 2a,b). In order to assign the *D. triauraria* scaffolds to Muller elements, we performed a translated BLAST search of our scaffolds using *D. melanogaster* peptides as queries and keeping only the best hit for each query sequence. We found that each scaffold was highly enriched for *D. melanogaster* peptides from a specific Muller element (Figure S1) and we successfully identified scaffolds corresponding to Muller elements A-F for downstream analysis.

D. triauraria was previously sequenced unintentionally because it was mislabeled as *D. kikkawai*, which means that the published *D. triauraria* nanopore data are from an unknown strain. The strain we sequenced is from the same stock center,

making it likely that the contaminant and our strain, are in fact the same strains. To test this we aligned the Illumina data from Miller et. al. [35], in conjunction with our uninformative Hi-C reads, to our nanopore assembly and called SNPs using *FreeBayes* [17]. We compared the genotypes from the Miller et. al. [35] assembly to our nanopore assembly at ∼93.7 million sites and found that 93.5 million sites (∼99.8%) were homozygous reference in both datasets, while at 215,000 sites, both assemblies had the same heterozygous genotype. For another 37,000 sites, one dataset was identifed as homozygous and the other heterozygous. The assemblies were in complete disagreement (i.e. they were homozygous for different alleles) at only 3 sites. From this analysis, we concluded that the *D. triauraria* strain mislabeled *D. kikkawai* was in fact the same strain we sequenced.

2.2 Genome Synteny

We next sought to identify synteny blocks and assess the degree of chromosomal rearrangements between *D. melanogaster* and *D. triauraria*. We created an orthology map between the genome assemblies for these two species using *Mercator* [7] and identified a total of 991 synteny blocks with average size of ∼117 Kb in *D. melanogaster* and ∼140 Kb in *D. triauraria*. The larger size of the synteny blocks in *D. triauraria* is consistent with the larger genome size for this species. We visualized synteny by using the promer tool from the mummer pipeline [27] to produce a dotplot (Figure 2c), which shows that there have been extensive chromosomal rearrangements since the divergence of these two species, with the majority of rearrangements occurring within Muller elements.

2.3 TAD Boundary and Domain Annotation

In order to determine how the large number of chromosomal rearrangements present between these two species has affected 3D genome organization, we identified TAD boundaries as well as complete contact domains in both species. We used *HiCExplorer* [39] to identify TAD boundaries at 5 kb resolution for both *D. melanogaster* and *D. triauraria*. HiCExplorer calculates the TAD separation score for each 5 kb bin in the genome and identifies TAD boundaries as those bins whose score shows significantly larger contact insulation compared to neighboring bins. The TAD separation scores were highly correlated between replicate datasets for each species (Spearman's rho: 0.993 [*D. melanogaster*] and 0.986 [*D. triauraria*] (Figure S2a,b) and the majority of predicted boundaries were identified in each replicate independently (74% [*D. melanogaster*] 70% [*D. triauraria*]). We refer to the boundaries that were identified in both replicates as high confidence boundaries, and those identified in only one of the two replicates as low confidence boundaries. In total, we identified 701 and 843 high confidence TAD boundaries for *D. melanogaster* and *D. triauraria*, respectively, and 249 and 355 low confidence boundaries (Table S2).

HiCExplorer links TAD intra-boundary regions together into full contact domains. Similar to our approach with boundary elements, we identified contact domains that were found independently in both replicate datasets as high confidence domains and those found only in one replicate as low confidence domains. In total, we identified 552 and 639 high confidence TAD domains for *D. melanogaster* and *D. triauraria*, respectively, and 593 and 811 low confidence domains (Table S2).

2.4 Boundary Motif Enrichment

In *Drosophila*, TAD boundaries are highly enriched for motifs recognized by the insulator proteins M1BP and BEAF-32/DREF [39, 19]. To validate boundary calls made by HiCExplorer, we used *Homer* software to search the identified boundaries for known insulator motifs. Boundaries were enriched for motifs recognized by M1BP (p-value $= 1e-17$) and BEAF-32 (p-value $= 1e-18$) in *D. melanogaster* and M1BP (p-value $= 1e-42$) and BEAF-32/DREF (p-value $= 1e-15$) in *D. triauraria*

boundaries which supports the accuracy of our boundary calls. Motifs available in Figure S3.

2.5 Domain and Boundary Conservation

We assessed the evolutionary conservation of TAD boundaries between *D. melanogaster* and *D. triauraria* by lifting over the high confidence *D. melanogaster* boundary coordinates to the *D. triauraria* genome coordinates. We considered boundaries to be orthologous when high confidence boundary regions lifted over from *D. melanogaster* to *D. triauraria* overlapped either a high or low confidence boundary that was independently identified in *D. triauraria*. Out of a total of 701 boundaries identified in *D. melanogaster*, 654 were successfully lifted over to a corresponding region in *D. triauraria*. Of the lifted over boundaries, 492 (∼75%) are orthologous between the two species and 163 (∼25%) are melanogaster-specific (Table 1, Figure S4a). Considering 75% of boundaries are orthologous, our results suggest that the sequences that specify these boundaries are highly conserved over evolutionary time.

There are two possibilities that would explain the high degree of boundary conservation in spite of the large number of chromosomal rearrangements between these species. One possibility is that the boundaries are conserved because the chromosomal rearrangements occur in such a way that TADs are shuffled as intact units (Model 1, Figure 1). The other possibility is that the sequences that specify boundaries remain conserved while chromosomal rearrangements shuffle these sequence elements in ways that lead to widespread TAD reorganization (Model 2, Figure 1). To differentiate between these possibilities, we identified orthologous contact domains between these two species. Similar to our approach with boundaries, we considered contact domains to be orthologous when high confidence domain regions from *D. melanogaster* lifted over as a continuous block (allowing for internal rearrangements) to *D. triauraria* and overlapped either a high or low confidence TAD domain that was independently identified in *D. triauraria*. We required that the domains were reciprocally overlapping by at least 90% of their lengths. Out of a total of 552 domains identified in *D. melanogaster*, 544 were successfully lifted over to a corresponding region in *D. triauraria*. Of the lifted over domains, we found that 134 (25%) are orthologous between the two species, whereas 410 (75%) of the *D. melanogaster* TADs do not show a one-to-one relationship with a *D. triauraria* TAD (Table 1, Figure S4a). Most of the non-orthologous TADs are due to cases where TADs have been split (Figure 3b). Of the orthologous domains, 84 (∼63%) also shared orthologous boundary regions. Given that only 25% of domains are orthologous between the two species, we conclude that chromosomal rearrangements have reorganized the majority of TADs present in each of these species since their common ancestor. Comparing the boundary and domain data, the rate of conservation of TAD boundaries is much higher than domains (Fisher's Exact Test p-value = $4.78e-63$) (Figure S4). Our results suggest that Model 2 (Figure 1) is the most likely scenario for TAD evolution: boundary regions are conserved but contact domains are reorganized between species. For consistency, we repeated these analyses by performing the liftover in the opposite direction, from *D. triauraria* to *D. melanogaster*, and obtained similar results (Table S3, Figure S4b).

2.6 Gene Expression

Enhancer-promoter contacts regulate gene expression. We hypothesized that TADs rearranged in *D. triauraria* compared to *D. melanogaster* might reorganize enhancer-promoter contacts and result in altered gene expression profiles. We compared the expression of genes within orthologous and non-orthologous TAD domains between the two species and found that, while nonconserved TADs show a slightly higher percentage of differentially expressed genes (10.5% versus 9.1%), this difference is not significant (Figure 4, Fisher's Exact Test p-value = 0.151). We did, however, find that orthologous TADs are enriched for homeobox domain-containing genes (FlyMine protein domain enrichment test, Benjamini-Hochberg corrected p-value = 0.02) and, in comparison to the genes within non-orthologous TADs, are also highly enriched for genes predicted to be regulated

by Polycomb-group proteins (Fisher's Exact Test p-value = $2.6e-5$) [5].

2.7 Chromatin State

Finding that genes in orthologous TADs are enriched for Polycomb-group proteins, which are characterized by the *blue* chromatin state described by Filion et al. [14], led us to compare chromatin states between genes in orthologous and non-orthologous TADs. We quantified the number of genes in each of five chromatin states within orthologous and non-orthologous TAD regions (Table S4). Orthologous TADs show significant enrichment of the *black* (transcriptionally silent) and *blue* (Polycomb-repressed) chromatin states and significant depletion of the *green* (constitutive heterochromatin) and *yellow* (constitutively active) chromatin states, compared to non-orthologous TADs (Fisher's Exact Test p-values: $1.225e-25$ [*black*], $4.322e-4$ [*blue*], $1.552e-15$ [*green*], $8.375e-23$ [*yellow*]) (Figure 5). The chromatin state tracks in Figure 3a and 3b support our findings. The majority of genes in the conserved TADs in Figure 3a are black and blue, while the genes within the split TAD in Figure 3b are predominantly *red* and *yellow*. These results largely mirror the chromatin states of the ancient and highly conserved contact domains identified by Harmston et. al. [20], which contain clusters of conserved non-coding elements and developmental genes .

3 Discussion

In this study, we sought to examine the evolutionary conservation of 3D genome organization in *Drosophila*. We selected *D. melanogaster* and *D. triauraria* for this comparison because they are separated by ∼28 million years of evolution [3]. We predicted that this level of divergence would be long enough that large-scale chromosomal rearrangements would have occurred between the two species but short enough that conservation at the nucleotide level would allow for an accurate whole-genome alignment. We used a combination of nanopore and Illumina Hi-C sequencing data to improve a recently published *D. triauraria* genome assembly produced from relatively low-coverage (depth 18.8x) nanopore sequencing data [35]. We have previously shown that Hi-C data can be used to scaffold *Drosophila* nanopore contigs with high accuracy, and even correct contig misassemblies [11]. Our improved *D. triauraria* assembly resulted in chromosome-length scaffolds highly enriched for genes corresponding to a single Muller element (Figure S1), further supporting the efficacy of this approach. We were able to align ∼87% of our *D. triauraria* assembly to the *D. melanogaster* reference assembly and we found extensive chromosomal rearrangements (Figure 2c), consistent with our initial prediction that *D. triauraria* and *D. melanogaster* represent an ideal species pair for use in a comparative study of 3D genome organization.

Previous research has yielded conflicting results on the conservation of TAD domains. While the overarching opinion is that TADs are highly conserved due to their role in restricting gene-enhancer interactions and therefore, gene expression, more current research suggests that TADs may actually diverge quickly and that TAD reorganization is not necessarily associated with divergence in gene expression [13, 18, 29, 34, 6]. Given the high degree of evolutionary conservation despite extensive chromosomal rearrangement between *D. melanogaster* and *D. triauraria* (Figure 2c), we expected that entire TAD contact domains including boundary regions would be conserved (i.e. Model 1, Figure 1). However, we found that TAD boundary elements are much more conserved compared to the TADs themselves (Fisher Exact Test p-value = $4.78e-63$), which suggests that shuffling of boundary elements results in reorganization of TAD architecture, depicted by Model 2 (Figure 1). Previous studies have identified inconsistencies in TAD-calling software packages [53] and have raised the possibility that TAD conservation results may depend on the direction of the liftover comparison [13]. For example, studies report conservation estimates by first calling TADs in the species for which they have less data and then identifying the orthologous domains in

the species for which they have more data [40, 8, 41]. When reversing the analysis the conservation rate can be reduced by up to 25%. However, in our study, we used biological replicates to demonstrate that the identification of TAD boundaries and TAD units is highly reproducible. We also performed bilateral assessments of TAD conservation and obtained similar results regardless of the direction of comparison. Furthermore, our estimates of conservation, if biased at all, should be biased towards inferring higher levels of conservation. We only considered TADs for our liftover step if they were independently identified in both biological replicates, which should enrich for stronger TADs. Furthermore, after liftover, we considered the TAD to be orthologous if it overlapped *either* a strong (i.e. high-confidence) TAD *or* a weak TAD (i.e. low-confidence TAD identified in only single replicate). We also did not require orthologous TADs to have orthologous boundaries. Instead, they were only required to have a reciprocal overlap of at least 90% of their lengths. We would expect these relatively low-stringency criteria to potentially result in an over-estimate of TAD conservation, yet we still only find ∼25% of TADs to be orthologous between species.

One potential function of TADs is to reduce aberrant enhancer-promoter contacts by constraining 3D interactions to specific segments of the chromosome. Some genome editing experiments have shown that deletion of TAD boundaries results in the fusion of neighboring TADs and is accompanied by misregulation of gene expression [36, 32]. However, other studies show that removal of a TAD boundary or rearrangement does not affect gene regulation [6, 18]. We compared the gene expression profiles within orthologous TADs to non-orthologous TADs. Slightly (∼2%) more of the genes in non-orthologous domains are differentially expressed, however this difference is not significant (Fishers Exact Test p-value = 0.151). One possibility that would explain the lack of association between TAD reorganization and gene expression divergence is that, in the relatively compact *Drosophila* genome, most gene regulatory sequences are directly adjacent to the genes they regulate and are therefore rarely separated by chromosomal rearrangements or TAD reorganization [29].

Our chromatin evaluation revealed that orthologous TADs are significantly enriched for inactive and Polycomb-repressed chromatin states, as well as homeobox-domain containing genes and genes predicted to be regulated by Polycomb-group (PcG) proteins. PcG proteins play an important role in regulation of developmental genes [5], bringing our results in line with previous studies showing that TADs enriched for developmental genes are more likely to be conserved [20]. It has been known that developmental genes are surrounded by clusters of conserved noncoding elements (CNEs) Sandelin2004-ug. These clusters of CNEs have been termed genomic regulatory blocks (GRBs) [12] and are highly-conserved [43, 12]. It has been recently discovered that GRBs coincide strongly with a subset of highly-conserved TADs in vertebrates and invertebrates ranging from *Drosophila* to humans [20, 26]. This high level of conservation may be due to the fact that disruption of gene-regulatory element contacts for developmental genes is likely to cause aberrant gene expression and phenotypic abnormalities [20, 26].

To summarize, our results show that TAD structures diverge rapidly and this divergence in 3D organization is not associated with gene expression divergence. Instead, highly conserved TADs are enriched for Polycomb-repressed genes, lending more support for an important link between Polycomb-repressed chromatin and 3D genome organization in *Drosophila*. One possibility that would explain the apparent contradictions in results from experiments testing the role of TADs in gene expression and the data showing that conserved and non-conserved TADs are enriched for different gene and chromatin types is that there are different subtypes of TADs with variable tolerance for disruption. Disruption of some types of TADs, such as those containing developmental genes, may detrimentally affect gene expression profiles and are therefore highly conserved. Other types of TADs could potentially be altered without any major effects on gene expression and these are the ones found to diverge quickly between species. If this is true, previous studies reporting contradictory effects of TAD rearrangement on gene expression may simply be due to the differences in the subtypes of TAD being tested.

4 Methods

4.1 Sequencing

Using the Qiagen DNAeasy Blood and Tissue Kit we extracted DNA from ∼30 *D. triauraria* females. We used the Oxford Nanopore Technologies (ONT) SQK-LSK 108 library preparation kit to construct three replicate PCR-free libraries for each species according to the ONT 1D Genomic DNA by Ligation protocol. The libraries were sequenced on a three MinION r9.4 flow cells. Raw signal data were basecalled using the ONT *Albacore* software package with default parameters.

4.2 Hi-C Chromosome Conformation Capture

D. triauraria and *D. melanogaster* strains were maintained in population cages on molasses agar with yeast paste. Embryos (8-16 h) for each species were collected and dechorionated in 50% commercial bleach for 2.5 min. Two biological replicate sets of nuclei from each species (∼100g/set) were isolated and fixed in 1.8% formaldehyde for 15 minutes according to the protocol [44]. Replicate sets for each species were sequenced using the *in situ* DNase Hi-C protocol described by [38]. Hi-C sequencing was performed on an Illumina HiSeq 2500 at the Rutgers Human Genetics Institute.

4.3 Genome Assembly and Visualization

Nanopore reads from each *D. triauraria* replicate were assembled using *Canu*. Genome quality control was performed using a series of software packages. First, *Purge Haplotigs* identifies regions of heterozygosity and determines a consensus sequence [42]. Second, *Nanopolish* accounts for errors in nanopore sequencing [31]. Third, the *Pilon* package combines data from high-quality short read Illumina sequences with long-read sequences to correct bases, fix mis-assemblies, and fill gaps in the assembly [51]. Our assembly was combined with the draft assembly from Miller et. al. [35] for downstream analysis.

The *Juicer* pipeline scaffolded the *D. triauraria* nanopore reads using Hi-C sequencing data. *Juicer* computed contact frequencies between nanopore contigs. *3D-DNA* software [9] then utilized Juicer contact frequencies to organize and concatenate the assembly fragments into 'megascaffolds' which are then split into full length chromosome arms. To assign the *D. triauraria* megascaffolds to corresponding to the Muller elements A-F, we performed a reverse BLAST analysis using the *D. melanogaster* iso-1 genome [22] (Figure 2a,b). *Juicebox* software was used to visualize the alignment, manually correct any minor misalignments, and export a finalized reference sequence for downstream analysis. *D. melanogaster* Hi-C data was also aligned to the reference assembly using the *Juicer* pipeline but only for visualization using *Juicebox*. [2]. Dotplot comparison of *D. melanogaster* and *D. triauraria* references was produced using *Mummer* software [27] (Figure 2c).

4.4 Identifying TAD Boundaries and Domains

We removed adapter sequences from replicate Hi-C reads for each species using *Trimmomatic* and used a custom perl script to split reads that contain a ligation junction. We used BWA software to align the split forward and reverse Hi-C reads to each species' reference assembly (D.triauraria assembly generated in this study combined with assembly from Miller et. al. [35], *D. melanogaster* assembly from Flybase [48]). We used *HiCExplorer* software to build a corrected contact frequency matrix. To find TAD boundaries and domains for each species we ran the *hicFindTads* option on each replicate matrix. We used a custom python script to correlate the TAD separation scores for each replicate. Boundaries and domains identified in both replicates were considered high confidence while those identified in one replicate are low confidence.

4.5 Defining and identifying orthologous boundaries between *D. melanogaster* and *D. triauraria*

We softmasked the *D. melanogaster* and *D. triauraria* genomes using *Repeatmasker* and aligned them using *Cactus* to generate a hal file. We input the high confidence boundaries for D. melanogater to *halLiftover* software to identify the corresponding genomic coordinates in *D. triauraria*. We merged 'lifted over' boundary locations within 5000 bp. Lifted over boundaries less than 500 bp in size were removed as this is 1/10 the size of an average boundary and so we believed those to be artifacts. Merged and filtered boundaries are defined as unique boundaries. Lifted over boundary locations in *D. triauraria* that were independently identified as high or low confidence boundaries by HiC explorer software are orthologous boundaries. Unique, lifted over boundary locations in *D. triauraria* that were not identified as boundaries by HiC explorer are non-orthologous. We implemented the same pipeline for *D. triauraria* to *D. melanogaster*.

4.6 Boundary Motif Enrichment

To confirm boundary identification using *HiCExplorer*, we used *Homer* software to test for insulator motif enrichment in HiC explorer output boundaries. We used the corresponding fasta sequences for each high confidence boundary location in *D. melanogaster* as the target. The entire genome broken into 5kb sequences was used as the background. Additionally, we used *Homer* to test for motif enrichment at non-orthologous boundary regions. The target sequences were regions identified as a boundary in *D. melanogaster* and the background regions were the corresponding genomic locations in *D. triauraria* that were not identified as boundaries by *HiCExplorer*. The same two analyses were completed using *D. triauraria* as the target.

4.7 Defining and identifying orthologous domains between *D. melanogaster* and *D. triauraria*

To assess domain conservation between *D. melanogaster* and *D. triauraria* we used halLiftover. Lifted over domain locations were merged if within 20 kb. Lifted over features less than 5000 bp in size were removed since this is about 1/10 the size of an expected domain. Merged and filtered domains are defined as unique boundaries. Lifted over domains in *D. triauraria* overlapping >90% with HiC explorer identified high or low confidence domains are orthologous domains (Figure S5a). Unique, lifted over domain locations in *D. triauraria* that were not identified as domains by HiC explorer are non-orthologous. Additionally, to identify orthologous domains between the two species that also share boundaries, we filtered orthologous boundaries for those whose endpoints are within 5kb. We implemented the same pipeline for *D. triauraria* to *D. melanogaster* (Figure S5b).

4.8 Gene Expression

We compared gene expression between orthologous and non-orthologous TAD domains categorized using the *D. melanogaster* to *D. triauraria* liftover analysis. To identify genes within each TAD category we intersected the *D. melanogaster* genomic coordinates of the TAD domains with the *D. melanogaster* reference [22]). RNA-seq data was analyzed using *DESeq* R software package [1]. Orthologous genes and non-orthologous genes were extracted from *DESeq* output. Fisher's Exact Test was used to evaluate whether genes within orthologous and non-orthologous TADs are differentially expressed. We applied a Gene Ontology (GO) enrichment analysis to the genes from orthologous domains to evaluate whether TADs with genes of a certain function are more likely to be conserved.

4.9 Chromatin State

We assessed the chromatin state of genes within orthologous and non-orthologous TADs categorized using the *D. melanogaster* to *D. triauraria* liftover analysis. We intersected our lists of orthologous and non orthologous genes with the chromatin state reference [22] to assign each gene one of five chromatin states (*black*, *blue*, *green*, *red*, *yellow*). We implemented a custom perl script to count the number of orthologous and non orthologous genes that fall into each of five chromatin states. Output was analyzed using one-sided Fisher's Exact Tests for each chromatin category.

References

[1] S. Anders and W. Huber. Differential expression analysis for sequence count data. *Genome Biol.*, 11(10):R106, Oct. 2010.

[2] J. Armstrong, G. Hickey, M. Diekhans, A. Deran, Q. Fang, D. Xie, S. Feng, J. Stiller, D. Genereux, J. Johnson, V. D. Marinescu, D. Haussler, J. Alföldi, K. Lindblad-Toh, E. Karlsson, G. Zhang, and B. Paten. Progressive alignment with Cactus: a multiple-genome aligner for the thousand-genome era. Aug. 2019.

[3] O. Barmina and A. Kopp. Sex-specific expression of a HOX gene associated with rapid morphological evolution. *Dev. Biol.*, 311(2):277–286, Nov. 2007.

[4] B. Bonev, N. Mendelson Cohen, Q. Szabo, L. Fritsch, G. L. Papadopoulos, Y. Lubling, X. Xu, X. Lv, J.-P. Hugnot, A. Tanay, and G. Cavalli. Multiscale 3D genome rewiring during mouse neural development. *Cell*, 171(3):557–572.e24, Oct. 2017.

[5] B. A. Bredesen and M. Rehmsmeier. DNA sequence models of genome-wide drosophila melanogaster polycomb binding sites improve generalization to independent polycomb response elements. *Nucleic Acids Res.*, 47(15):7781–7797, Sept. 2019.

[6] A. Despang, R. Schöpflin, M. Franke, S. Ali, I. Jerković, C. Paliou, W.-L. Chan, B. Timmermann, L. Wittler, M. Vingron, S. Mundlos, and D. M. Ibrahim. Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat. Genet.*, 51(8):1263–1271, Aug. 2019.

[7] C. N. Dewey. Aligning multiple whole genomes with mercator and MAVID. *Methods Mol. Biol.*, 395:221–236, 2007.

[8] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380, Apr. 2012. PMC3356448.

[9] O. Dudchenko, S. S. Batra, A. D. Omer, S. K. Nyquist, M. Hoeger, N. C. Durand, M. S. Shamim, I. Machol, E. S. Lander, A. P. Aiden, and E. L. Aiden. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, 356(6333):92–95, Apr. 2017. PMC5635820.

[10] N. C. Durand, M. S. Shamim, I. Machol, S. S. P. Rao, M. H. Huntley, E. S. Lander, and E. Lieberman Aiden. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell systems*, 3:95–98, July 2016.

[11] C. E. Ellison and W. Cao. Nanopore sequencing and Hi-C scaffolding provide insight into the evolutionary dynamics of transposable elements and piRNA production in wild strains of drosophila melanogaster. *Nucleic Acids Res.*, 48(1):290–303, Jan. 2020.

[12] P. G. Engström, S. J. Ho Sui, O. Drivenes, T. S. Becker, and B. Lenhard. Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Res.*, 17(12):1898–1908, Dec. 2007.

[13] I. E. Eres, K. Luo, C. J. Hsiao, L. E. Blake, and Y. Gilad. Reorganization of 3D genome structure may contribute to gene regulatory evolution in primates. *PLoS Genet.*, 15(7):e1008278, July 2019.

[14] G. J. Filion, J. G. van Bemmel, U. Braunschweig, W. Talhout, J. Kind, L. D. Ward, W. Brugman, I. J. de Castro, R. M. Kerkhoven, H. J. Bussemaker, and B. van Steensel. Systematic protein location mapping reveals five principal chromatin types in drosophila cells. *Cell*, 143(2):212–224, Oct. 2010.

[15] W. A. Flavahan, Y. Drier, B. B. Liau, S. M. Gillespie, A. S. Venteicher, A. O. Stemmer-Rachamimov, M. L. Suvà, and B. E. Bernstein. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*, 529(7584):110–114, Jan. 2016. PMC4831574.

[16] M. Franke, D. M. Ibrahim, G. Andrey, W. Schwarzer, V. Heinrich, R. Schöpflin, K. Kraft, R. Kempfer, I. Jerković, W.-L. Chan, M. Spielmann, B. Timmermann, L. Wittler, I. Kurth, P. Cambiaso, O. Zuffardi, G. Houge, L. Lambie, F. Brancati, A. Pombo, M. Vingron, F. Spitz, and S. Mundlos. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*, 538(7624):265–269, Oct. 2016.

[17] E. Garrison and G. Marth. Haplotype-based variant detection from short-read sequencing. *arXiv [q-bio.GN]*, pages 1–9, July 2012.

[18] Y. Ghavi-Helm, A. Jankowski, S. Meiers, R. R. Viales, J. O. Korbel, and E. E. M. Furlong. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.*, 51(8):1272–1282, Aug. 2019.

[19] B. V. Gurudatta, J. Yang, K. Van Bortle, P. G. Donlin-Asp, and V. G. Corces. Dynamic changes in the genomic localization of DNA replication-related element binding factor during the cell cycle. *Cell Cycle*, 12(10):1605–1615, May 2013.

[20] N. Harmston, E. Ing-Simmons, G. Tan, M. Perry, M. Merkenschlager, and B. Lenhard. Topologically associating domains are ancient features that coincide with metazoan clusters of extreme noncoding conservation. *Nat. Commun.*, 8(1):441, Sept. 2017. PMC5585340.

[21] D. Hnisz, A. S. Weintraub, D. S. Day, A.-L. Valton, R. O. Bak, C. H. Li, J. Goldmann, B. R. Lajoie, Z. P. Fan, A. A. Sigova, J. Reddy, D. Borges-Rivera, T. I. Lee, R. Jaenisch, M. H. Porteus, J. Dekker, and R. A. Young. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*, 351(6280):1454–1458, Mar. 2016.

[22] R. A. Hoskins, J. W. Carlson, K. H. Wan, S. Park, I. Mendez, S. E. Galle, B. W. Booth, B. D. Pfeiffer, R. A. George, R. Svirskas, M. Krzywinski, J. Schein, M. C. Accardo, E. Damia, G. Messina, M. Méndez-Lago, B. de Pablos, O. V. Demakova, E. N. Andreyeva, L. V. Boldyreva, M. Marra, A. B. Carvalho, P. Dimitri, A. Villasante, I. F. Zhimulev, G. M. Rubin, G. H. Karpen, and S. E. Celniker. The release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Res.*, 25(3):445–458, Mar. 2015. PMC4352887.

[23] C. Hou, L. Li, Z. S. Qin, and V. G. Corces. Gene density, transcription, and insulators contribute to the partition of the drosophila genome into physical domains. *Mol. Cell*, 48(3):471–484, Nov. 2012.

[24] C. B. Hug, A. G. Grimaldi, K. Kruse, and J. M. Vaquerizas. Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell*, 169(2):216–228.e19, Apr. 2017.

[25] S. Koren, B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, and A. M. Phillippy. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.*, 27(5):722–736, May 2017.

[26] J. Krefting, M. A. Andrade-Navarro, and J. Ibn-Salem. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol.*, 16(1):87, Aug. 2018. PMC6091198.

[27] S. Kurtz, A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S. L. Salzberg. Versatile and open software for comparing large genomes. *Genome Biol.*, 5(2):R12, Jan. 2004.

[28] N. H. Lazar, K. A. Nevonen, B. O'Connell, C. McCann, R. J. O'Neill, R. E. Green, T. J. Meyer, M. Okhovat, and L. Carbone. Epigenetic maintenance of topological domains in the highly rearranged gibbon genome. *Genome Res.*,

28(7):983–997, July 2018. PMC6028127.

[29] H. Lee, D.-Y. Cho, C. Whitworth, R. Eisman, M. Phelps, J. Roote, T. Kaufman, K. Cook, S. Russell, T. Przytycka, and B. Oliver. Effects of gene dose, chromatin, and network topology on expression in drosophila melanogaster. *PLoS Genet.*, 12(9):e1006295, Sept. 2016.

[30] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, Oct. 2009. PMC2858594.

[31] N. J. Loman, J. Quick, and J. T. Simpson. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods*, 12(8):733–735, Aug. 2015.

[32] D. G. Lupiáñez, K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H. Kayserili, J. M. Opitz, R. Laxova, F. Santos-Simarro, B. Gilbert-Dussardier, L. Wittler, M. Borschiwer, S. A. Haas, M. Osterwalder, M. Franke, B. Timmermann, J. Hecht, M. Spielmann, A. Visel, and S. Mundlos. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 161(5):1012–1025, May 2015. PMC4791538.

[33] D. G. Lupiáñez, M. Spielmann, and S. Mundlos. Breaking TADs: How alterations of chromatin domains result in disease. *Trends Genet.*, 32(4):225–237, Apr. 2016.

[34] L. A. Meadows, Y. S. Chan, J. Roote, and S. Russell. Neighbourhood continuity is not required for correct testis gene expression in drosophila. *PLoS Biol.*, 8(11):e1000552, Nov. 2010.

[35] D. E. Miller, C. Staber, J. Zeitlinger, and R. S. Hawley. Highly contiguous genome assemblies of 15 drosophila species generated using nanopore sequencing. *G3*, 8(10):3131–3141, Oct. 2018.

[36] V. Narendra, P. P. Rocha, D. An, R. Raviram, J. A. Skok, E. O. Mazzoni, and D. Reinberg. CTCF establishes discrete functional chromatin domains at the hox clusters during differentiation. *Science*, 347(6225):1017–1021, Feb. 2015. PMC4428148.

[37] J. E. Phillips-Cremins, M. E. G. Sauria, A. Sanyal, T. I. Gerasimova, B. R. Lajoie, J. S. K. Bell, C.-T. Ong, T. A. Hookway, C. Guo, Y. Sun, M. J. Bland, W. Wagstaff, S. Dalton, T. C. McDevitt, R. Sen, J. Dekker, J. Taylor, and V. G. Corces. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, 153(6):1281–1295, June 2013.

[38] V. Ramani, D. A. Cusanovich, R. J. Hause, W. Ma, R. Qiu, X. Deng, C. A. Blau, C. M. Disteche, W. S. Noble, J. Shendure, and Z. Duan. Mapping 3D genome architecture through *in situ* DNase Hi-C. *Nat. Protoc.*, 11(11):2104–2121, Nov. 2016. PMC5547819.

[39] F. Ramírez, V. Bhardwaj, L. Arrigoni, K. C. Lam, B. A. Grüning, J. Villaveces, B. Habermann, A. Akhtar, and T. Manke. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.*, 9(1):189, Jan. 2018. PMC5768762.

[40] S. S. P. Rao, S.-C. Huang, B. Glenn St Hilaire, J. M. Engreitz, E. M. Perez, K.-R. Kieffer-Kwon, A. L. Sanborn, S. E. Johnstone, G. D. Bascom, I. D. Bochkov, X. Huang, M. S. Shamim, J. Shin, D. Turner, Z. Ye, A. D. Omer, J. T. Robinson, T. Schlick, B. E. Bernstein, R. Casellas, E. S. Lander, and E. L. Aiden. Cohesin loss eliminates all loop domains. *Cell*, 171(2):305–320.e24, Oct. 2017. PMC5846482.

[41] G. Renschler, G. Richard, C. I. K. Valsecchi, S. Toscano, L. Arrigoni, F. Ramirez, and A. Akhtar. Hi -C guided assemblies reveal conserved regulatory topologies on X and autosomes despite extensive genome shuffling. *BioRxiv*, Mar. 2019.

[42] M. J. Roach, S. A. Schmidt, and A. R. Borneman. Purge haplotigs: allelic contig reassignment for third-gen diploid

genome assemblies. *BMC Bioinformatics*, 19(1):460, Nov. 2018.

[43] A. Sandelin, P. Bailey, S. Bruce, P. G. Engström, J. M. Klos, W. W. Wasserman, J. Ericson, and B. Lenhard. Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics*, 5(1):99, Dec. 2004.

[44] T. Sandmann, J. S. Jakobsen, and E. E. M. Furlong. ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in drosophila melanogaster embryos. *Nat. Protoc.*, 1(6):2839–2855, 2006.

[45] T. Sexton, E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, and G. Cavalli. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell*, 148(3):458–472, Feb. 2012.

[46] M. R. Stadler, J. E. Haines, and M. B. Eisen. Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early *Drosophila melanogaster* embryo. *Elife*, 6, Nov. 2017. PMC5739541.

[47] O. Symmons, V. V. Uslu, T. Tsujimura, S. Ruf, S. Nassari, W. Schwarzer, L. Ettwiller, and F. Spitz. Functional and topological characteristics of mammalian regulatory domains. *Genome Res.*, 24(3):390–400, Mar. 2014.

[48] J. Thurmond, J. L. Goodman, V. B. Strelets, H. Attrill, L. S. Gramates, S. J. Marygold, B. B. Matthews, G. Millburn, G. Antonazzo, V. Trovisco, T. C. Kaufman, B. R. Calvi, and FlyBase Consortium. FlyBase 2.0: the next generation. *Nucleic Acids Res.*, 47(D1):D759–D765, Jan. 2019.

[49] S. V. Ulianov, E. E. Khrameeva, A. A. Gavrilov, I. M. Flyamer, P. Kos, E. A. Mikhaleva, A. A. Penin, M. D. Logacheva, M. V. Imakaev, A. Chertovich, M. S. Gelfand, Y. Y. Shevelyov, and S. V. Razin. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.*, 26(1):70–84, Jan. 2016. PMC4691752.

[50] M. Vietri Rudan, C. Barrington, S. Henderson, C. Ernst, D. T. Odom, A. Tanay, and S. Hadjur. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.*, 10(8):1297–1309, Mar. 2015. PMC4542312.

[51] B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, and A. M. Earl. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, 9(11):e112963, Nov. 2014.

[52] J. Weischenfeldt, T. Dubash, A. P. Drainas, B. R. Mardin, Y. Chen, A. M. Stütz, S. M. Waszak, G. Bosco, A. R. Halvorsen, B. Raeder, T. Efthymiopoulos, S. Erkek, C. Siegl, H. Brenner, O. T. Brustugun, S. M. Dieter, P. A. Northcott, I. Petersen, S. M. Pfister, M. Schneider, S. K. Solberg, E. Thunissen, W. Weichert, T. Zichner, R. Thomas, M. Peifer, A. Helland, C. R. Ball, M. Jechlinger, R. Sotillo, H. Glimm, and J. O. Korbel. Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat. Genet.*, 49(1):65–74, Jan. 2017.

[53] M. Zufferey, D. Tavernari, E. Oricchio, and G. Ciriello. Comparison of computational methods for the identification of topologically associating domains. *Genome Biol.*, 19(1):217, Dec. 2018. PMC6288901.

5 Tables

Category	**Boundaries**	**Domains**
Total in *D. melanogaster*	701	552
Unique lifted over to *D. triauraria*	654	544
Orthologous	492	134
Non-orthologous	163	410

Table 1: Summary of results from *D. melanogaster* to *D. triauraria* liftover analysis.

6 Figures

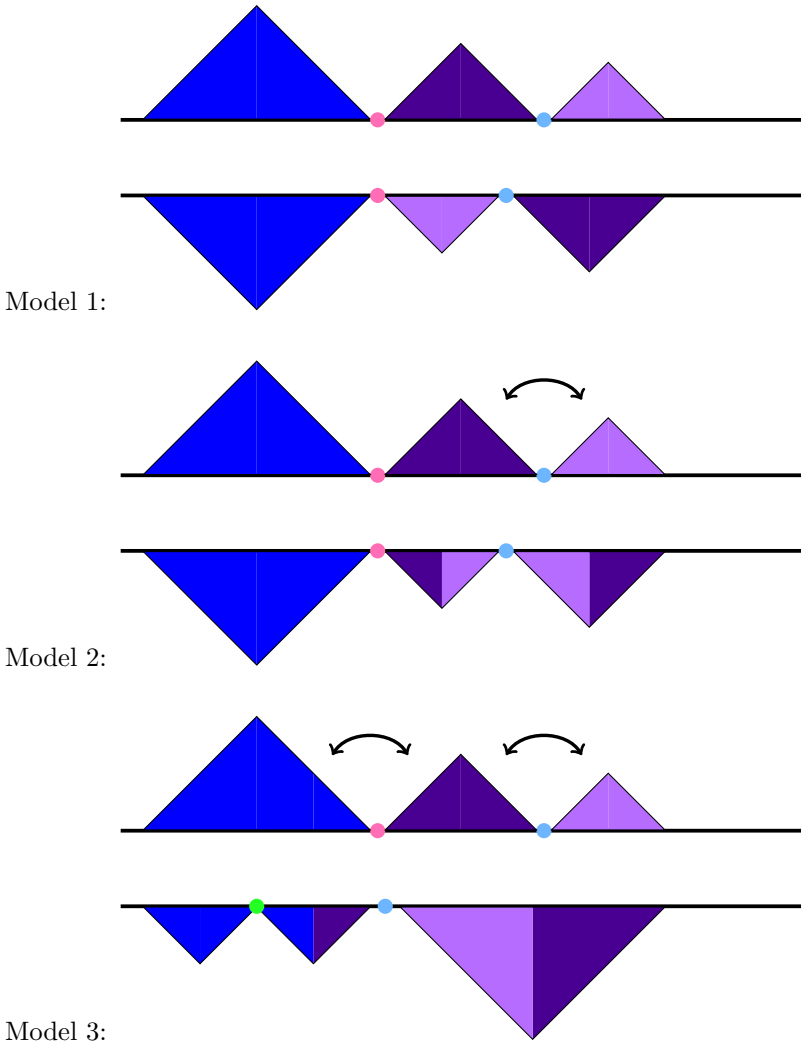


Figure 1: Models depicting possible TAD and boundary rearrangement scenarios. Model 1 shows TAD contact domains and boundaries conserved in two species, though domains and boundaries may shuffle as a unit, as represented by the purple and lilac TADs. Model 2 shows conservation of TAD boundaries but disruption of contact domains. Model 3 shows a scenario where the TAD domains and boundaries are disrupted with the creation of new TADs and boundaries as well as the loss of TADs and boundaries.

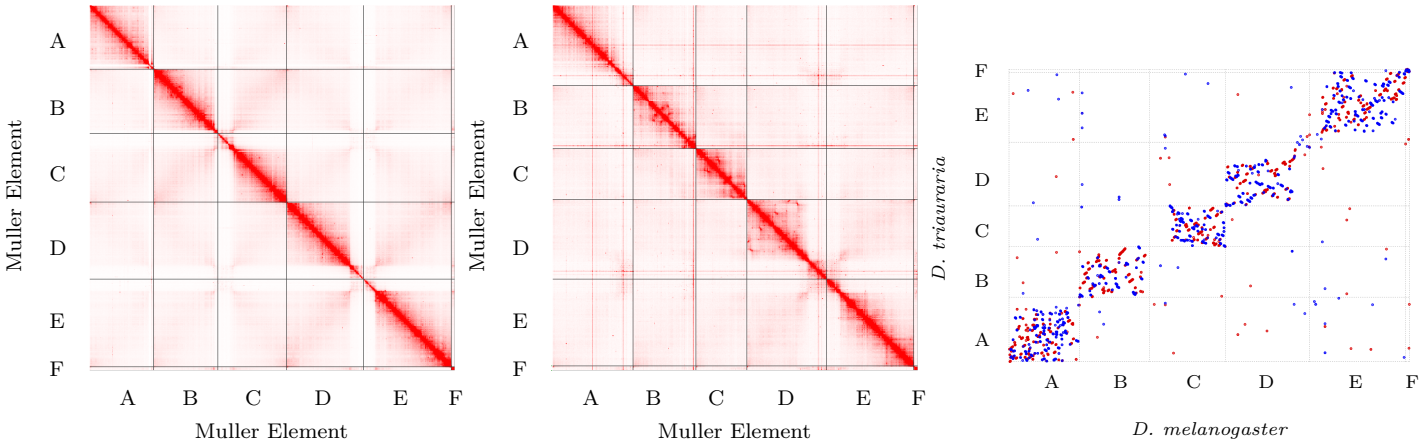


Figure 2: a) *D. melanogaster* contact map; b) *D. triauraria* contact map; c)MUMMER dotplot depicting genome rearrangement between *D. melanogaster* and *D. triauraria*.

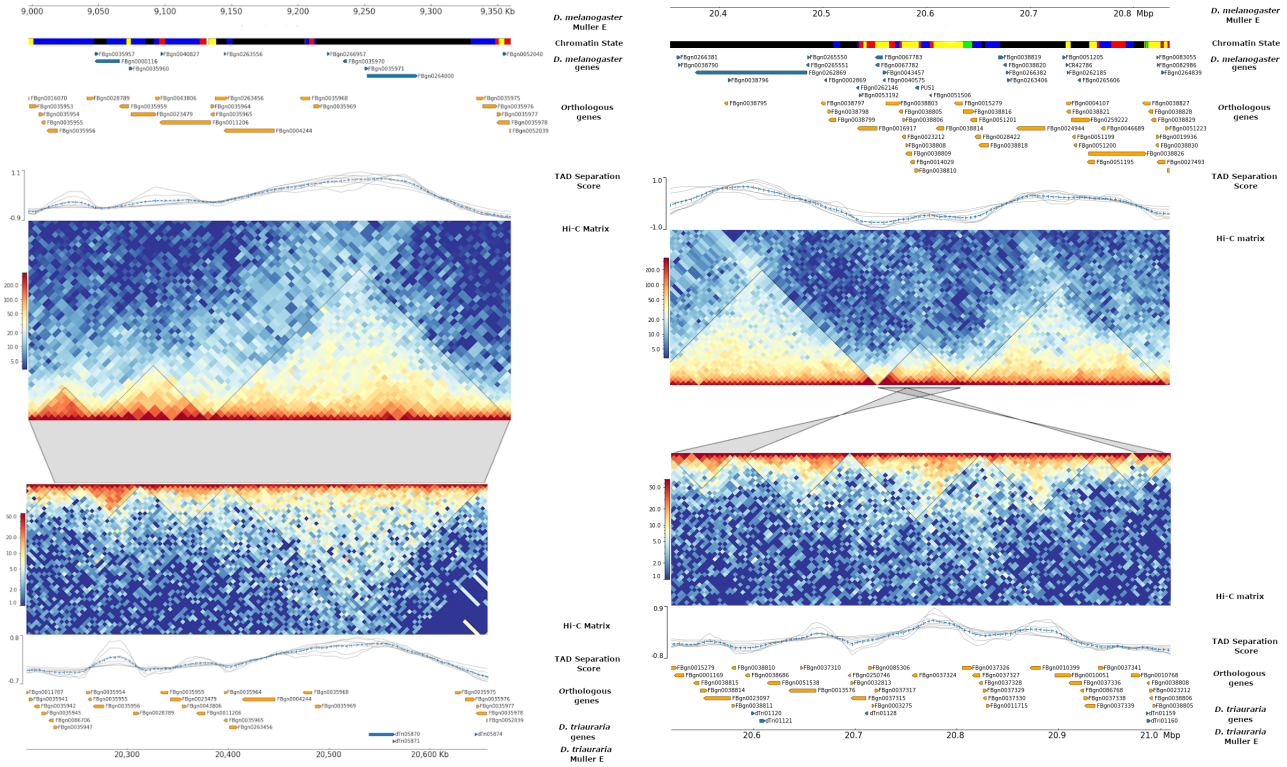
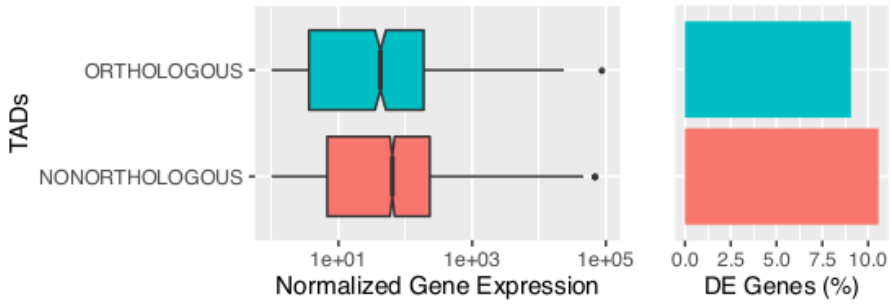


Figure 3: a) Three orthologous, conserved TADs on Muller E between *D. melanogaster* and *D. triauraria*. b) Non-orthologous, spilt TAD on Muller E between *D. melanogaster* and *D. triauraria*. Panels as indicated. Chromatin state of *D. melanogaster* genes based on Hoskins et al. [22]. *Black*, *blue*, and *green* are inactive. *Red* and *yellow* are active. Orthologous genes labeled by FlyBase IDs. Hi-C matrices generated by *HiCExplorer*. Grey blocks connecting matrices indicate syntenic regions. In b) gene tracks show that genes such as FBgb0038805, FBgn0038806, and FBgn0038814 are split between different TADs in *D. triauraria* and are in the reverse orientation compared to *D. melanogaster*.



(a) (b)

Figure 4: a) Normalized gene expression of genes within orthologous and non-orthologous TADs. b) Percent of differentially expressed genes within orthologous (9.1%) and non-orthologous (10.5%) TADs. Differences in expression are not significant (p-value = 0.151)

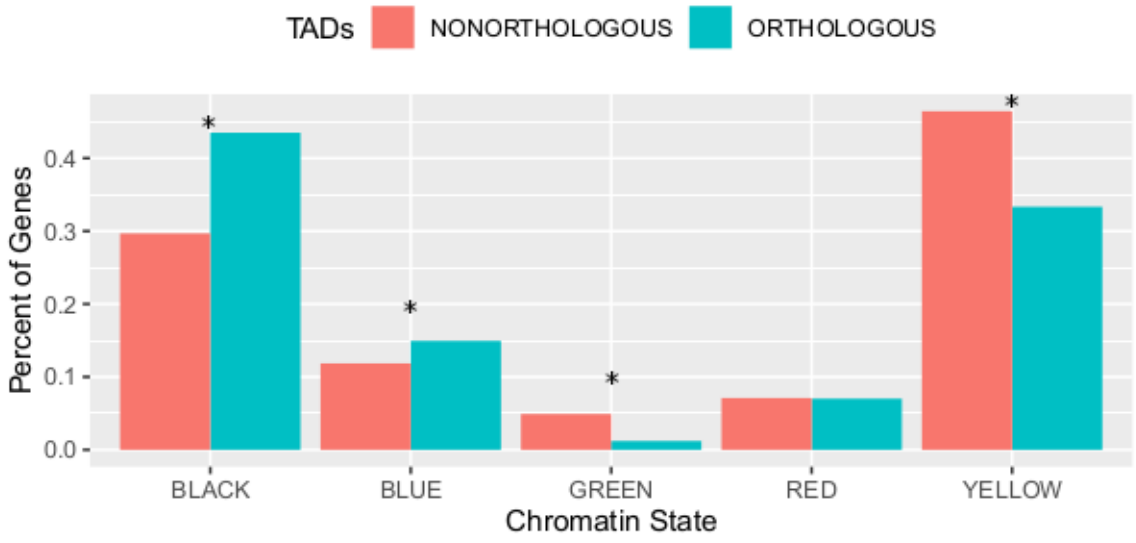


Figure 5: Percent of genes in each chromatin state in orthologous and non-orthologous TADs. Asterisk $*$ indicates significant difference as calculated by Fisher's Exact Test (p-values: Black $= 1.225e-25$; Blue $= 4.322e-4$; Green $= 1.552e-15$; Yellow $= 8.375e-23$; Red $= 0.398$. Results revealed that orthologous TADs are significantly enriched for black and blue chromatin regions. Non-orthologous TADs are significantly enriched for green and yellow chromatin regions.