INTRO

Outline:

**Gene expression important for phenotypic evolution**

Gene regulation is an essential molecular process that guides spatiotemporal gene expression throughout development and organismal life. Given the role of gene regulation in organismal development and functionality, variation in gene regulatory mechanisms often underlie the evolution of phenotypes such as morphology, behavior, etc. Historically, many of phenotypic changes have been mapped or unraveled at the stage of transcriptional activation, where transcription factors bind to cis-regulatory elements. However, more recent comparative work has found other steps, such as chromatin remodeling, to play a causal role in gene expression evolution.

**Chromatin remodeling increasingly well studied and being integrated into our understanding of evolutionary mechanisms**

Chromatin remodeling entails making regions of DNA accessible and epigenetically marked to allow for subsequent binding of regulatory proteins for transcriptional activation. The molecular mechanisms by which chromatin is made accessible are complex and multifaceted but seem to occur on a chromosome-wide level, in which larger regions are remodeled by X and X, and 2), as well as at a more locus-specific level, in which various transcription factor-based mechanisms determine the accessibility of a regulatory region. The distinction between these various mechanisms is important for understanding the role of chromatin remodeling in gene expression evolution because the gene expression may be altered in different ways, allowing for different types of variation. Past work, however, has focused more on broader trends of chromatin accessibility evolution and results are varied.

One predominant trend has emerged from previous work focusing on chromatin accessibility evolution: genome-wide chromatin accessibility divergence is largely due to changes in *cis* rather than *trans.* Here, changes in *cis* may be any changes that alter chromatin accessiblity in an allele-specific manner, such as alterations to transcription factor binding sites; whereas changes in *trans* could be changes in any diffusible molecule that is a determinant of chromatin accessibility. However, there are variable results on the extent to which changes in chromatin accessibility correlate with changes in gene expression. In yeast, multiple studies have found little correlation between changes to these two mechanistic layers, although more notable correlations were found with *Drosophila* population variation. Notably, a modelling approach between *Drosophila* species found that chromatin accessibility variation explained a non-overlapping amount of TF binding compared to sequence changes, suggesting that chromatin accessibility changes can play a causal role in downstream regulatory changes.

**Gap and we did this to fill it.**

Much of chromatin remodeling remains unclear in both a functional and evolutionary context, and a more direct integration of the two is needed to advance both. More specifically, chromatin remodeling does not occur in the same way across the genome: how does this impact evolutionary patterns? Here, we collect chromatin accessibility and gene expression data from *Drosophila* species and their hybrids in a tissue with more extensive molecular understanding of chromatin remodeling mechanisms to better interrogate the evolutionary consequences of molecular differences in chromatin remodeling. We find that, consistent with previous results,

RESULTS

*Experimental schematic*

To interrogate the mode of chromatin accessibility and gene expression evolution, we collected ATAC- and RNA-seq data from the imaginal wing discs of D. melanogaster and D. simulans and the F1 hybrid. These strains were chosen because of 1) their capacity to interbreed and 2) the extensive previous work to catalog the genomic variation between these species. Following alignment and processing, accessibility was measured for four genomic categories: +/- 500bp transcriptions start sites (TSS), +/- 500bp transcriptions end sites (TES), +/- 500bp center of called peaks that fall within coding regions (intragenic), and +/- 500bp center of called peaks that fall between coding regions (intergenic). As expected, variation was greatest for intergenic regions, then intragenic, and then TSS and TES. After filtering regions with low read counts, we had X, X, X, and X numbers per group. We then used an empirical Bayes method to estimate the difference between parental strains, hybrid alleles, and the parental vs hybrid difference. This provided us with divergence estimates as well as statistical metrics to determine cis and trans categories.

*High divergence due to cis changes is constant across genomic categories*

To quantify the amount of chromatin accessibility divergence due to *cis* changes, we used the **percent *cis*** metric which infers such from parental and hybrid allele differences. With this metric, chromatin accessibility regions are, on average, 75% due to *cis* changes, consistent with prior work in fly embryos, yeast, and stickleback fish. We then contrasted percent cis across the different functional groups to see if there were any differences between groups since the mechanisms giving rise to accessibility between these groups are likely different. Interestingly, percent cis remained within +/- for all groups, with only statistically significant differences between X and Y, with divergence being X and Y.

*Expression divergence is most extreme and more due to cis changes when correlated with accessibility changes.*

To assess the relationship between expression and accessibility divergence with an emphasis on mode of divergence, we integrated our allele-specific ATAC- and RNA-seq datasets by either matching TSS or TES sites with their annotated gene or matching inter- and intra-genic regions with the closest expressed gene. Next, we grouped accessible region-gene pairs based on whether the parental strains were significantly different from each other to then contrast patterns of divergence for instances when accessibility and expression were or were not both divergent or conserved. Interestingly, gene expression divergence is greater when associated regions also have divergent accessibility, consistent with these two mechanistic layers being correlated. Furthermore, gene expression changes are much more due to cis changes when associated regions also have divergent accessibility.

*Regions remodeled by pioneer factors evolve differently than those not*

Chromatin remodeling is not mechanistically homogenous across the genome. In particular, past work in *Drosophila* imaginal discs has shown that many regions are made accessible by the pioneer factor, *Grainyhead*, whereas others seem to be made accessible and activated by collective TF binding. The discrepancy between remodeling mechanisms raises an interesting question: do we see different modes of chromatin accessibility evolution based on remodeling mechanism? We used previously existing ATAC-seq data from *Grainyhead* knockout and control imaginal discs to characterize regions for which pioneer factor activity is *necessary* for accessibility. We took these 300 regions to be putatively remodeled by a pioneer factor for the subsequent analyses.