INTRO

**Gene expression important for phenotypic evolution**

Gene regulation guides spatiotemporal gene expression throughout development and organismal life. As such, changes in gene regulation that produce gene expression variation often underlie the evolution of higher-order phenotypes, such as morphology and behavior (cite). These mechanistic changes often take place specifically during transcription activation, during which *trans*-acting elements bind to *cis­*-regulatory elements to recruit transcriptional machinery, although this may in part be biased by our lagging knowledge of the other steps. However, our rapidly growing understanding of gene regulation in its entirety now allows us to survey the role and/or contribution of other mechanistic steps to gene expression evolution.

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

Chromatin remodeling is an essential early step of gene regulation. To allow for regulatory machinery to activate transcription, DNA must first be made accessible by evicting nucleosomes and then epigenetically modifying regions to maintain accessibility. The extent to which regions are made and maintain accessible is known as chromatin accessibility. The molecular mechanisms that give rise to chromatin accessibility are varied and can depend on the functional category (e.g. promoter or enhancer) of a given region, and can even vary within these categories. For example, the accessibility of some *Drosophila* enhancers seems to be primarily determined by the binding of a single transcription factor with the ability to bind nucleosome-bound DNA (known as a pioneer factor), whereas other enhancers’ accessibility is determined by the collective binding of numerous transcription factors. Additionally, promoters are generally known to differ in nucleosome affinity based on whether they are “broad” or “narrow” peak promoters. The fact that chromatin remodeling is carried out differently across the genome to produce different accessibility patterns presents an interesting evolutionary question: does the remodeling mechanism impact the evolutionary patterns of chromatin accessibility and its relationship to expression variation?

Firstly, previous work has found that chromatin accessibility variation can explain TF binding variation unaccounted for by motif sequence variation, suggesting that chromatin variation can causally lead to gene expression variation. Interestingly, distal enhancer regions are more variable than proximal promoter elements, although variation in the latter better correlates with gene expression variation, suggesting that variation in promoters is a better predictor of gene expression variation. Furthermore, tests with F1 hybrids have shown that most chromatin accessibility variation results from changes in *cis* (e.g. nearby transcription factor binding sites) rather than *trans* (e.g. diffusible molecule, such as chromatin remodeling protein or transcription factor). These findings shed light on the extent to which chromatin accessibility is variable and the mechanistic link between chromatin remodeling and gene expression variation, but yet to be studied is how these patterns may vary based on chromatin remodeling mechanism.

To address this knowledge gap, we collected chromatin accessibility and gene expression datasets of imaginal wing disc tissue from a North American and an African population of *D. melanogaster* and their hybrids. We then estimated the extent and mode of chromatin variation at transcription start sites, transcription end sites, and called peaks that fell either within or outside of genic regions. To test the relationship between chromatin accessibility and gene expression variation, we ran a generalized linear model with the different promoters and called peaks labelled by promoter type and remodeling mechanism, respectively. These data allowed us to describe chromatin accessibility variation at a more finely resolved resolution, as well as incorporate current molecular biology into our understanding of how variation across mechanistic layers is related.

*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.

FIGURES NEEDED: HEATMAPS OF EACH CATEGORY AND REVISED PHYLOGENY (INCLUDE RNA-SEQ NOW?)

*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.

FIGURES NEEDED: (A) CISxTRANS (B) PERCENT CIS DISTRIBUTION (C) EXTENT AND (D) PERCENT CIS ACROSS CATEGORIES

*Extent and mode of chromatin accessibility divergence differs for TSS regions with different promoter types.*

*Extent and mode of chromatin accessibility divergence differs for intergenic regions made accessible by pioneer factor or non-pioneer factor mechanisms.*

To contrast the extent and mode of chromatin accessibility divergence by regions made accessible by different molecular mechanisms, we took advantage of previous work that identified regions necessary for accessibility in imaginal disc epithelial tissue. More specifically, regions for which Grainyhead KOs resulted in a loss of accessibility, had Grainyhead binding, and overlapped with accessible regions in the data collected here, and were defined as pioneer factor accessible. Interestingly, these regions are overall less divergent but more often due to cis-regulatory changes. Furthermore, we verified that these differences in chromatin accessibility are associated with Grh motif differences more frequently than expected by chance. Taken together, these results are consistent with our understanding that 1) Grainyhead primes constitutively active regulatory regions in epithelial cells which may be under more constraint, and 2) the Grainyhead protein is essential for genome-wide chromatin remodeling and therefore trans changes may be under more constraint than regions that are remodeled by a multitude of TFs with more specific functionalities. These findings highlight the relevance of understanding the molecular mechanisms to predict evolutionary outcomes.

FIGURES NEEDED: (A) SCHEMATIC FOR CATEGORIZING GRH-BOUND KO NECESSARY VS GRH-BOUND KO NOT NECESSARY (B) PIE CHART FOR PROPORTION OF GRH-KO REGIONS IN EACH CATEGORY (C) ROLE OF VARIATION IN GRH BINDING SITES IN ACCESSIBILITY DIVERGENCE (D) EXTENT AND (E) PERCENT CIS BETWEEN GRH-KO AND NON-KO.

*The relationship between gene expression and chromatin accessibility divergence based on chromatin remodeling mechanism.*

We next sought to contrast the relationship between gene expression and chromatin accessibility divergence between regions remodeled by a pioneer factor or not. Previous work has shown weak to modest correlations between the two mechanistic layers, although for some cases the correlation seems to be non-overlapping with motif variation. Based on our current molecular model for enhancer remodeling, changes to pioneer factor binding could disrupt the binding of all subsequent activators by altering the overall accessibility, whereas alterations to binding for other enhancers may not be as large. Are changes in accessibility for pioneer factor regions associated with greater differences in gene expression?

FIGURES NEEDED: (A) SCHEMATIC FOR LINKING REGIONS (B) SCATTERPLOT OF ACCESSIBLITY VS EXPRESSOIN (C) EXPRESSION DIVERGENCE AND (D) PERCENT CIS FOR ACCESSIBILITY OVER ALL AND (E) GRH KO NON KO.