The genetic mechanisms that coordinate the development and morphological divergence of *Drosophila* genital structures

Amber Mae Ridgway

Degree awarded by
Oxford Brookes University

Thesis submitted in partial fulfilment of the requirements of the award of Doctor of Philosophy

November 2023

Acknowledgements

In the last 4 years, I became low key obsessed with *Drosophila* male genitalia whilst living through the COVID-19 pandemic. Despite the obstacles, I can honestly say that this PhD has been brilliant. *C15* is a fascinating gene that I now feel responsible for and *Sox21b*, although you took me over a year to figure out your expression pattern (scarred by colorimetric *ISHs* for at least a decade now), you got me my first ever first-authored paper, so cheers to you.

Firstly, I would like to thank Alistair McGregor for believing in me and giving me this opportunity. Although you weirdly wanted to move to a place called Durham that is an average of two degrees colder than Oxford and is the birthplace of English mustard, you have provided ongoing support and guidance throughout, and for that I say thank you! Thank you to Grace and Amber, my spider friends I now cannot imagine my life without. Thank you for the endless laughter. Thank you to Alex, my midlands (ish) life companion, for always being so supportive and incredibly knowledgeable, could not have done this without you. Javier, your dark humour is one of a kind, as well as our shared love for *Drosophila* male genitalia, thank you for your patience and exchange of knowledge. I'd like to also thank the rest of the McGregor lab, Dani and Emily, Seb and Maike, as well as the *Drosophila* community at Brookes. The academic environment here has been hugely supportive, and I will miss it.

Jake Melles. You deserve at least 23% of this PhD. Thank you for always believing in me, for driving me to university late at night and in the rainiest of rains, for cooking most nights these past four years (even though I've eaten enough dhal in the past 4 years to fill Warwick Castle) and for being endlessly proud of me, it did not go unnoticed. You have seen my academic journey since A-Levels, witnessed my stress episodes and nerdy excitement, and even became my husband in 2023. Thank you from the bottom of my heart.

Eastenders – you have been a weekly essential when getting ready for work in the mornings. Although, I still do not how many dun dun duns there are in the theme tune.

Thank you to mom and dad for telling everyone I work on genitalia with no context. You have both provided such support and encouragement from the very start and I cannot thank you enough. Poppy, I feel like I need to name you here. Nan, for being my fashion icon and a beam of sunshine. And Grampy, although you passed just a week before I handed my thesis in, I know that you were always so proud of me. I miss you every day and miss our phone calls with the iconic ‘Hiya Darling’ start lines. This thesis is dedicated to you Gramps (Christopher Michael Finch), I love you lots.

Abstract

External male genitalia are one of the most rapidly evolving insect body parts. Although the developmental programmes required to form some of these structures are known to some extent, the genetic differences causing phenotypic divergence are not well understood. *Drosophila simulans* and *D. mauritiana* exhibit striking morphological differences in the size, shape, and bristle composition of the male periphallic genitalia, despite only diverging 240,000 years ago. In my PhD project, I first investigated the genetic components underlying male posterior lobe shape and size differences between *D. simulans* and *D. mauritiana*. By performing RNAi knockdown against differentially expressed transcription factor encoding genes, I identified *Sox21b*. I found that posterior lobe size is repressed by Sox21b during larval genital development. To investigate whether it had evolved between species, I conducted reciprocal hemizygosity tests and found that posterior lobe size and shape differences are causative of *Sox21b* evolution between *D. simulans* and *D. mauritiana*. I then isolated an enhancer driving *Sox21b* expression in the posterior lobe primordia, which is positively regulated by Abd-B. By comparing this enhancer's activity between *D. simulans* and *D. mauritiana* I found differences consistent with *Sox21b* gene expression differences. In parallel, I aimed to better understand the developmental programme coordinating genital development. It has been shown that the genital structures are a derived walking appendage body part, yet gene regulatory network conservation between different developmental contexts is not well understood. I focused on the leg pretarsal gene regulatory network coordinated by the transcription factor C15, and its similarities and differences during male clasper and female epiproct genital development. I characterised the expression and function of *C15* and its interacting genes in both structures and described network similarities and differences. I also found that during male clasper development, C15 uniquely represses bristle growth in contrast to its activator role elsewhere in the fly. I identified the enhancer coordinating *C15* male genital expression and predicted upstream TFs that may regulate its male-genital specific expression. Overall, my study has contributed new insights into the causative genetic basis for the rapid evolution of male *Drosophila* genitalia and the gene regulatory networks underlying the development of these structures.

Publications relevant to thesis

Publication

Ridgway AM, Hood E, Jimenez JF, Nunes MDS, McGregor AP. *Sox21b* underlies the rapid diversification of a novel male genital structure between *Drosophila* species. *Current Biology.* (2024) https://doi.org/10.1016/j.cub.2024.01.022 Chapter 3

Publication associated with this thesis

Hagen JFD, Mendes CC, Booth, SR, Jimenez JF, Tanaka KM, Franke FA, Baudouin-Gonzalez L, **Ridgway AM**, Arif S, Nunes MDS, McGregor AP. Unravelling the genetic basis for the rapid diversification of male genitalia between *Drosophila* species. *Molecular Biology and Evolution.* 2021;38(2):437-448. doi:10.1093/molbev/msaa232 Chapter 3, 4 and 5

Table of Contents

1. Chapter One: General Introduction 1

1.1 Morphological Diversity 1

1.1.1 Enhancer evolution drives phenotypic variation 3

1.1.2 Hotspots for phenotypic evolution 4

1.2 Rapid evolution of *Drosophila* external male genital structures 6

1.2.1 The posterior lobes are a morphological novelty 9

1.2.2 Sexual selection and male posterior lobe diversity 9

1.2.3 Importance of posterior lobe shape and size during copulation 10

1.3 Thesis Aims 11

2. Chapter Two: Materials and Methods 13

2.1 GAL4-UAS: RNAi knockdown and overexpression 13

2.1.1 Phenotyping 13

2.1.2 Statistical analysis 14

2.1.3 Scanning electron microscopy 14

2.2 Immunohistochemistry staining 15

2.2.1 Collection, fixation and staining of embryos 15

2.2.2 Larval discs 15

2.2.3 Pupal terminalia 16

2.2.4 Imaging 17

2.3 Hybridisation chain reaction *in situs* (HCR-IS) 17

2.4 Generation of *Sox21b* mutants using CRISPR-Cas9 17

2.4.1 CRISPR design and cloning 17

2.4.2 Embryo injections 18

2.4.3 Validation of DsRed insertion 18

2.4.4 Reciprocal hemizygosity test and phenotyping 19

2.4.5 Morphometric analysis of posterior lobe shape 19

2.5 Janelia GAL4 enhancer reporter lines 19

2.5.1 Pupa preparation and live imaging 20

2.6 *Sox21b* Enhancer 3 GAL4 reporter lines 20

2.6.1 Generation of pBPGUw Enhancer 3 constructs 20

2.6.2 Injections and screening 21

2.7 Generation of UAS lines 21

2.7.1 Cloning strategy 21

2.7.2 Injections 21

2.8 Bioinformatic analysis 22

2.8.1 ATAC-seq processing 22

2.8.2 TF binding site search 22

2.8.3 Multiple sequence alignment 23

2.8.4 Pairwise sequence alignment 23

3. Chapter Three Results 24

Sox21b contributes to the rapid diversification of a novel male genital structure between Drosophila simulans and Drosophila mauritiana 24

3.1 Background: Mapping studies to resolve genetic regions contributing to morphological diversity in *Drosophila* male genital structures 24

3.1.1 Chapter three aims 25

3.2 Results: Transcription Factors differentially expressed between species contributes to the development of periphallic male genital structures 26

3.2.1 Posterior lobe RNAi knockdown effect genes 28

3.2.2 Clasper RNAi knockdown effect genes 29

3.2.3 Cercus RNAi knockdown effect genes 30

3.3 *Sox21b*-coordinated posterior lobe development and evolution 32

3.3.1 RNAi knockdown of *Sox21b* in *D. melanogaster* 32

3.3.2 *Sox21b* RNAi knockdown posterior lobe base width and lateral plate effects 34

3.3.3 *Sox21b* HCR-IS in *D. melanogaster* 35

3.3.4 Co-option of posterior spiracle GRN in posterior lobes 37

3.4 Using CRISPR to test if *Sox21b* has evolved between *D. simulans* and *D. mauritiana* 38

3.4.1 Reciprocal hemizygote posterior lobe size differences 39

3.4.2 Reciprocal hemizygote posterior lobe shape differences 41

3.4.3 *Sox21b* HCR-IS in *D. simulans* and *D. mauritiana* male genital discs 42

3.5 Group B Sox Genes 43

3.5.1 Group B Sox genes expression levels in *D. mauritiana* and *D. simulans* developing male genitalia 44

3.5.2 Dichaete protein localisation across development of the male terminalia compared to *Sox21b* 45

3.5.3 *Dichaete* RNAi knockdown and overexpression 46

3.6 Chapter Three Discussion 48

3.6.1 Group B Sox genes and the regulation of posterior lobe development 49

3.6.2 *Sox21b* was probably not co-opted from the ancestral posterior spiracle GRN 49

3.6.3 Evolution of *Sox21b* between *D. simulans* and *D. mauritiana* 50

4. Chapter Four Results 51

Characterisation of Sox21b enhancers and trans-regulatory landscape between species 51

4.1 Background: Enhancers and their evo-devo importance 51

4.1.1 Chapter four aims 52

4.2 Results: *Sox21b* enhancer screening in *D. melanogaster* 53

4.2.1 Live pupa screening of putative *Sox21b* enhancers 54

4.2.2 Live pupa screening of GMR73B05 (Enhancer 1) and GMR73A06 (Enhancer 2) 56

4.2.3 *Sox21b* Enhancer 1 57

4.2.4 *Sox21b* Enhancer 2 60

4.3 *Sox21b* Enhancer 3: cis-regulatory sequence shared between Enhancer 1 and Enhancer 2 63

4.3.1 *D. melanogaster Sox21b* Enhancer 3 64

4.3.2 Comparison of *D. simulans* and *D. mauritiana* Enhancer 3 activity 65

4.4 Transcription factor binding sites in *Sox21b* Enhancer 3 67

4.4.1 Footprint mapping of Enhancer 3 67

4.4.2 TF motif mapping in *D. melanogaster* Enhancer 3 68

4.4.3 Functional testing of Abd-B binding sites in *D. melanogaster* Enhancer 3 70

4.5 Chapter Four Discussion 72

4.5.1 ATAC-seq and enhancer activity comparisons 72

4.5.2 Pupal genital activity of *Sox21b* enhancers 74

4.5.3 Species-specific activity of *Sox21b* Enhancer 3 74

4.5.4 Abd-B regulates *Sox21b* expression 75

5. Chapter Five: Results 77

A comparative assessment of C15's role during the development of the terminalia compared to the pretarsal segment of the leg 77

5.1 Background: *Drosophila* male genitalia and legs 77

5.1.1 C15 is a pretarsus determination factor 79

5.1.2 Chapter five aims 82

5.2 Results: Characterisation of the pretarsal *C15* GRN in the male claspers 82

5.2.1 *wg* and *dpp* expression in the male pupal claspers 82

5.2.2 *dac* and *Dll* expression hubs in the male claspers 84

5.2.3 C15 protein localisation across development of the male terminalia 86

5.2.4 RNAi knockdown of *C15* in male terminalia 87
5.2.5 Al and Lim1 expression during male pupal genital development....89
5.2.6 Al and Lim1 function during male pupal genital development 90
5.2.7 C15, Lim1 and Al interactions during male clasper development....91
5.2.8 The role of Bab2, Bowl and Ap during male clasper development 95

5.3 Characterisation of the pretarsal *C15* GRN in the female epiproct 97
5.3.1 *wg* and *dpp* expression in the female pupal genitalia....97
5.3.2 Dll regulatory role in the female epiproct 98
5.3.3 C15 protein localisation during female terminalia development 99
5.3.4 RNAi knockdown of *C15* in female terminalia 100
5.3.5 Al and Lim1 expression during female pupal genital development....102
5.3.6 Al and Lim1 function during female pupal genital development 103
5.3.7 C15 and Al interaction during female epiproct development 105

5.4 Regulation of *C15* expression in the male claspers 106
5.4.1 ATAC-seq comparisons of the *C15* locus in leg and male genital samples....106
5.4.2 *C15* Intron Enhancer screen in male and female live pupae....107
5.4.3 Motif mapping of genital expressed TFs in *C15* intron 2 enhancer 1 109
5.4.4 Motif mapping of identified clasper-acting TFs in *C15* intron 2 enhancer 1.... 111

5.5 Chapter Five Discussion 111
5.5.1 Pretarsal genes and their deployment compared to male genitalia.... 112
5.5.2 C15 and Lim1 boundary formation in male genital structures 112
5.5.3 *C15* intron 2 enhancer 1 113
5.5.4 Sex-specific C15 expression in the cerci.... 114

6. General Discussion 116

6.1 *Sox21b* evolutionary implications 116
6.1.1 *Sox21b* and copulation 117
6.1.2 *Sox21b*'s posterior lobe GRN 117

6.2 *Sox21b* enhancer activity divergence between species....118
6.2.1 Enhancer 3 pleiotropy.... 118
6.2.2 Enhancer 3 evolution.... 118
6.2.3 Enhancer activity quantification 119

6.3 C15 disc-dependent regulatory role 119
6.3.1 C15 target gene differences 121

6.4 Evolutionary perspective of *C15*....121
6.4.1 *Tribolium C15* and *al*.... 122
6.4.2 Clasper diversity amongst *Drosophila* species.... 122

6.5 Conclusions **122**

7. Bibliography ***124***

8. Supplementary Materials ***144***

8.1 Chapter Two Supplementary Materials **144**

8.2 Chapter Three Supplementary Materials **152**

8.3 Chapter Four Supplementary Materials **158**

8.4 Chapter Five Supplementary Materials **166**

Table of figures

Figure 1: Examples of morphological diversity across the metazoan division of the animal kingdom. 2
Figure 2: Structural diversity of the periphallic male genitalia among *Drosophila* species. 7
Figure 3: RNAi knockdown posterior lobe effects. 28
Figure 4: RNAi knockdown clasper effects. 29
Figure 5: RNAi knockdown cercus effects. 31
Figure 6: Posterior lobe size effects of *Sox21b* RNAi knockdown. 34
Figure 7: Effects of *Sox21b* RNAi knockdown on the posterior lobe base width and lateral plate area. 35
Figure 8: HCR-IS of *Sox21b* in the male genital disc of *D. melanogaster*. 35
Figure 9: Drop staining in *D. melanogaster* developing male genital structures. 36
Figure 10: *Sox21b* HCR-IS and Anti-Eya staining in *D. melanogaster* embryos. 37
Figure 11: Generation of *; ; Sox21b-DsRed* transgenic flies in a pure *D. simulans*^{w501} *nos>cas9;;* line and in an *nos>cas9;;IL108 D. mauritiana* introgression line. 39
Figure 12: Effects of species-specific *Sox21b* alleles on posterior lobe size. 40
Figure 13: Effects of species-specific *Sox21b* alleles on posterior lobe shape. 42
Figure 14: Dynamic developmental and evolutionary expression of *Sox21b* in *D. simulans* and *D. mauritiana*. (A,B) 43
Figure 15: Species comparison of the Group B Sox locus on chromosomal arm 3L. 44
Figure 16: Dichaete antibody staining throughout male genital development. 46
Figure 17: Changes to *Dichaete* expression levels generates significant posterior lobe size phenotypes. 47
Figure 18: Janelia GAL4 enhancer driver lines tested at the *Sox21b* locus. 54
Figure 19: *Sox21b* GAL4-Enhancer driver lines that showed no visible genital activity in live pupae. 55
Figure 20: The GMR73B04 GAL4 line is active in isolated cells of the claspers and lateral plates. 56
Figure 21: *Sox21b* GAL4-Enhancer driver lines with genital-specific activity in live pupa. 57
Figure 22: GMR73B05 (Enhancer 1) is active in posterior lobe specifying cells throughout *D. melanogaster* male genital development. 59
Figure 23: *Sox21b* Enhancer 1 driven *Sox21b* RNAi knockdown posterior lobe size effect. 60
Figure 24: GMR73A06 (Enhancer 2) is active in posterior lobe primordia throughout *D. melanogaster* male genital development. 61
Figure 25: *Sox21b* Enhancer 2 driven *Sox21b* RNAi knockdown posterior lobe size effect. 63
Figure 26: New *Sox21b* Enhancer (Enhancer 3) consisting of the 811 bp region shared between Enhancer 1 and Enhancer 2. 64

Figure 27: *D. melanogaster* Enhancer 3 activity in live pupa and fixed larval genital discs. 65
Figure 28: Comparison of *D. simulans* and *D. mauritiana* Enhancer 3 sequences driving GFP expression in live pupae and fixed larval genital discs. 66
Figure 29: Footprint analysis of *Sox21b* enhancers in *D. simulans* and *D. mauritiana* ATAC-seq data. 68
Figure 30: Mapped TF binding sites in *D. melanogaster* Enhancer 3. 69
Figure 31: Mutant Abd-B binding sites in *D. melanogaster* Enhancer 3. 71
Figure 32: Enhancer 3 Abd-B binding sites in non-lobed species compared to lobed *D. melanogaster*. 72
Figure 33: Expression patterns of key signalling and patterning genes in the developing leg versus the developing male genitalia. 78
Figure 34: A simplified schematic highlighting the pivotal interactions originating from *C15* expression in the pretarsal segment of the leg. 81
Figure 35: *wg* and *dpp* are both expressed in the male claspers during male pupal genital development. 83
Figure 36: *dac* and Dll expression and function during male clasper development. 85
Figure 37: C15 expression in male genital and cercus structures during development. 86
Figure 38: RNAi knockdown of *C15* in male *D. melanogaster* developing genitalia. (A) 88
Figure 39: Expression of Al and Lim1 in the developing male genitalia. 89
Figure 40: RNAi knockdown of *al* and *lim1* in *D. melanogaster* male claspers and posterior lobes. 91
Figure 41: *al*, *lim1* and *C15* RNAi knockdown and reciprocal stainings to test known pretarsal interacting partners during male pupal genitalia development. 92
Figure 42: C15 expression upon RNAi knockdown of known male genitalia factors. 94
Figure 43: Pretarsal factors Bab2 and *bowl* expression and function during clasper development. 96
Figure 44: *wg* and *dpp* are both expressed in the female epiproct during pupal genital development. 98
Figure 45: RNAi knockdown of *Dll* in the developing *D. melanogaster* female epiproct. 99
Figure 46: Anti-C15 staining in female genital and cercus structures during development. 100
Figure 47: RNAi knockdown of *C15* in female *D. melanogaster* terminalia. 101
Figure 48: Expression of Al and Lim1 in the developing female genitalia. 102
Figure 49: RNAi knockdown of *al* and *lim1* in female *D. melanogaster* terminalia. 104
Figure 50: RNAi knockdown of *al* and *C15* in the female developing terminalia alters the expression of C15 and Al, respectively. 105
Figure 51: ATAC-seq for *C15* locus in larval leg discs and male pupal terminalia tissue. 107
Figure 52: C15 Intron enhancer reporter line activities in live and fixed pupae structures. 108
Figure 53: Mapped TF binding sites within *C15* Intron 2 Enhancer 1. 110

Figure 54: *Sox21b* expression differences contributing to posterior lobe divergence between *D. mauritiana* and *D. simulans*. 116
Figure 55: Summary of C15-central GRNs in the pretarsus of the leg, claspers of the male genitalia, and epiproct of the female cercus. 120

Supplementary Figure 1: An alternative GAL4 driver and increased temperature also generates an increase in cercus and posterior lobe area upon RNAi knockdown of *E(spl)m3-HLH*. ..153
Supplementary Figure 2: *Sox21b* HCR-IS in *D. melanogaster* female genital discs and developing male pupal terminalia. ..153
Supplementary Figure 3: Shape variation between *Sox21b* reciprocal hemizygotes posterior lobes captured by principal components. ..155
Supplementary Figure 4: *Sox21b* HCR-IS in *D. simulans* and *D. mauritiana* male pupal terminalia. (A,B) ..155
Supplementary Figure 5: Multiple sequence alignment of Group B Sox protein sequences from *D. melanogaster.* Sequence sourced from *D. melanogaster* R6.48. ..157
Supplementary Figure 6: Pairwise alignment of Dichaete and Sox21b *D. melanogaster* protein sequences. Both sequence were sourced from *D. melanogaster* R6.24. The Dichaete sequence shown in this figure is the region of the protein targeted by the Anti-Dichaete antibody. This sequence was compared to Sox21b to assess off-target effects. Alignment generated using EBLOSUM62 matrix. ..158
Supplementary Figure 7: *Sox21b* GAL4 enhancer driver lines not expressed during *D. melanogaster* male genitalia development. ..159
Supplementary Figure 8: Comparison of *D. melanogaster*, *D. simulans* and *D. mauritiana* Enhancer 3 GAL4 sequences driving *UAS-GFP* expression during male pupal genitalia development. ..160
Supplementary Figure 9: Abd-B and Eya stains in *D. melanogaster* developing male genital structures. (A) Abd-B is present in all structures apart from the cerci. (B) Eya is expressed in the lateral plate and the hinge region between the posterior lobe and lateral plate. ..161
Supplementary Figure 10: Multiple sequence alignment of *Sox21b Enhancer 3* between *D. melanogaster, D. simulans* and *D. mauritiana*. ..163
Supplementary Figure 11: Pairwise alignment of *D. mauritiana* and *D. simulans* Enhancer 3 sequences. ..165
Supplementary Figure 12: Abd-B staining in *D. melanogaster* Enhancer 3 Abd-B binding site mutant. Enhancer 3 Abd-B mutant GAL4 line is driving *UAS-GFP* and terminalia are stained against GFP to visualise enhancer activity. The mutated enhancer has prominent activity in the cerci, and comparatively less activity in the posterior lobes, lateral plates and claspers. ..165

Supplementary Figure 13: Multiple sequence alignment of *Sox21b Enhancer 3* between *D. melanogaster*, *D. mauritiana* and *D. simulans*. 166
Supplementary Figure 14: Multiple sequence alignment of Enhancer 3 between lobed and non-lobed species. 166
Supplementary Figure 15: Schematics of male genital expression patterns of known tarsal factors in larval discs and pupal tissues. 167
Supplementary Figure 16: TF expression in the proximal and distal clasper domains. 168
Supplementary Figure 17: *C15* RNAi knockdown effects on the cercus bristle count and T2 tibia length in *D. melanogaster* males. 169
Supplementary Figure 18: Pupal pretarsal leg expression of C15, Al and Lim1 169
Supplementary Figure 19: Anti-GFP staining against bab2GFP in the larval leg disc and pupal male claspers. 170
Supplementary Figure 20: *C15* RNAi knockdown has no significant effect on the hypoproct size and bristle count, or T2 tibia length in *D. melanogaster* females. (A) 170
Supplementary Figure 21: C15 antibody staining in *D. simulans*, *D. mauritiana* and *D. melanogaster* developing male terminalia. 171
Supplementary Figure 22: *C15* enhancer screening in the imaginal discs. (A,A',D,D') 171
Supplementary Figure 23: *C15* Intron 2 Enhancer 1 screening in female pupa, wing discs and eye/antennal discs. 172
Supplementary Figure 24: PWMs for TFs predicted to regulate *C15* expression during male genital development. 173

List of Tables

Table 1: Summary of the periphallic male genital phenotypes upon RNAi knockdown of selected differentially expressed TFs in *D. melanogaster*. 27

Table 2: Expression levels of Group B Sox Genes in the developing male terminalia of *D. mauritiana* and *D. simulans*. 45

Supplementary Table 1: Antibody staining conditions used for staining. 146

Supplementary Table 2: Oligonucleotide sequences used to generate the *Sox21b-DsRed* reagents and the *Sox21b* Enhancer 3 GAL4 constructs. 151

Supplementary Table 3: Statistical test results for Chapter 3 and 4 RNAi knockdowns phenotypic analysis. 151

Supplementary Table 4: Statistical test results for Chapter 3 *Sox21b* reciprocal hemizygotes phenotypic analysis. 151

Supplementary Table 5: EFA descriptors for *Sox21b* reciprocal hemizygotes posterior lobe shape analysis. 151

Supplementary Table 6: EFA descriptors for *D. mauritiana* and *D. simulans* posterior lobe shape analysis. 152

Supplementary Table 7: Statistical test results for Chapter 5 male RNAi knockdown and UAS-OE phenotypic analysis. 152

Supplementary Table 8: Statistical test results for Chapter 5 female RNAi knockdown and UAS-OE phenotypic analysis. 152

Supplementary Table 9: TFs used to scan *C15* intron 2 Enhancer 1 sequence for binding sites 152

Supplementary Table 10: TF binding sites predicted within *C15* intron 2 enhancer 1 152

Note: Supplementary tables 3 to 10 are available as a separate excel file.

Abbreviations

ATAC-seq: Assay for Transposase-Accessible Chromatin and sequencing
bp: base pair
Cas9: CRISPR associated protein 9
ChIP-seq: Chromatin Immunoprecipitation and sequencing
CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats
ºC: degrees Celsius
DNA: Deoxyribonucleic Acid
EGFR: Epidermal Growth Factor Receptor
EtOH: Ethanol
FIMO: Find Individual Motif Occurrences
GFP: Green Fluorescent Protein
gRNA: guide Ribonucleic Acid
hAEL: hours After Egg Laying
hAPF: hours After Pupal Formation
HMG: High Mobility Group
IL108: Introgression Line 108
Kb: Kilobase
Mb: Megabase
mRNA: messenger Ribonucleic Acid
PBS: Phosphate Buffer Saline
RNAi: Ribonucleic Acid interference
RNA-seq: Ribonucleic Acid sequencing
UAS: Upstream Activating Sequences

Attribution Statement

This thesis used the following datasets and resources for experimental design and data analysis:

- *D. simulans*^{w501} and *D. mauritiana* D1 male pupal terminalia RNA-Seq datasets (Hagen *et al.*, 2019).
- *D. simulans*^{w501} and *D. mauritiana* D1 male pupal terminalia ATAC-Seq dataset (Joanna Hagen, unpublished).
- Janelia *D. melanogaster* GAL4 lines (Jenett *et al.*, 2012).
- *D. melanogaster*^{w1118} T1, T2 and T3 larval leg disc RNA-Seq and ATAC-Seq datasets (Feng *et al.*, 2022).
- *D. melanogaster* Oregon R T2 pupal legs (20-28 hAPF) RNA-Seq dataset (Kittelmann *et al.*, 2018).
- *In situ hybridisation D. melanogaster* male terminalia database for the top 100 expressed transcription factors (Vincent *et al.*, 2019).

I would like to express my appreciation and thanks to the above authors for generating these datasets. Without these resources, a lot of the research presented in this thesis would not have been achievable.

1. Chapter One: General Introduction

1.1 Morphological Diversity

Morphological diversity encompasses complex to subtle differences between and within species (Carroll, Grenier and Weatherbee, 2009; Das Gupta and Tsiantis, 2018; Deline *et al.*, 2018; Gaspar *et al.*, 2020; Rothier *et al.*, 2023). To better understand how these phenotypic differences have arisen required identifying the underlying genes to help decode the molecular and developmental mechanisms (Weatherbee *et al.*, 1999; Wasik, Rose and Moczek, 2010; Hagen *et al.*, 2019). This has not only revealed the genetic basis of many phenotypic differences within and between species but also so called 'hotspot' genes underlying the convergent evolution of traits, and has helped to genetically describe the developmental networks underpinning structural novelties (Tanaka *et al.*, 2011; Arnoult *et al.*, 2013; Martin and Orgogozo, 2013; Glassford *et al.*, 2015; Gibert *et al.*, 2016; Yassin *et al.*, 2016; Hagen *et al.*, 2019).

For example, two subspecies of the oldfield mouse (*P. polionotus*) diverged just 200,000 years ago and exhibit very different pigmentation and patterning across the body (Figure 1) (Steiner, Weber and Hoekstra, 2007). *P. p. subgriseus* has a dark brown colouring across the body, whereas *P. p. leucocephalus* has lighter coloured pigmentation, and a lack of any pigment at the face and tail (Figure 1) (Steiner, Weber and Hoekstra, 2007). Despite this short evolutionary timescale, it was found that coding and regulatory mutations in *Mcr1* and *Agouti,* respectively, were largely responsible for the pigmentation differences (Steiner, Weber and Hoekstra, 2007; Wooldridge *et al.*, 2022). The *Mcr1* coding difference between the two subspecies was resolved to a single nucleotide polymorphism that encoded for an opposingly charged amino acid shown to affect receptor function (Hoekstra *et al.*, 2006).

Wing pigmentation patterns between breeds of rock pigeons (*C. livia*) varies extensively (Derelle *et al.*, 2013; Vickrey *et al.*, 2018). Comparisons between these pigeons led to the identification of *NDP*, a gene required for pigmentation (Figure 1) (Vickrey *et al.*, 2018). This gene has regulatory and coding differences between these pigeon breeds (Vickrey *et al.*, 2018). In those that lacked plumage colouring, *NDP* was homozygous for a nonsense mutation, which truncated the encoded protein by eleven residues (Vickrey *et al.*, 2018). Consequently, this *NDP* coding difference between breeds was speculated to disrupt its ability

to be secreted from the cell, and thus impact upon downstream pigmentation mechanisms (Vickrey *et al.*, 2018).

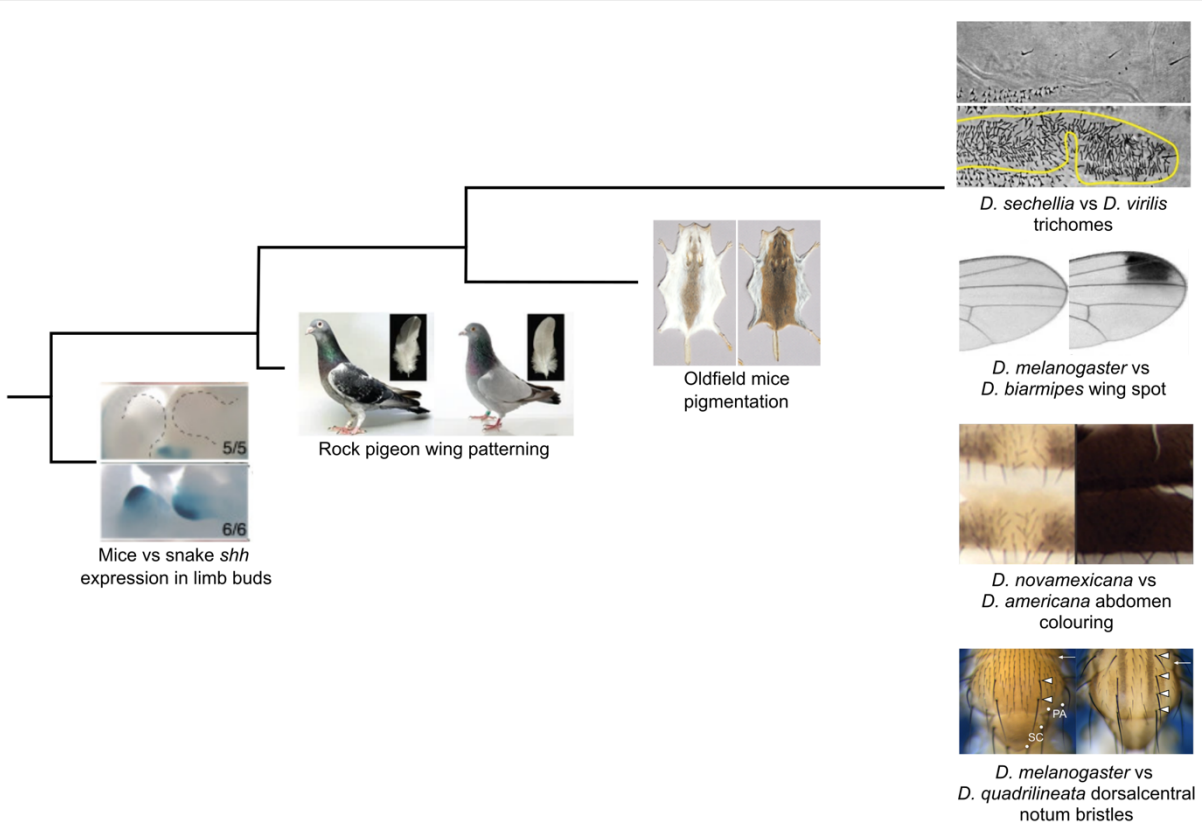


Figure 1: Examples of morphological diversity across the metazoan division of the animal kingdom. Phenotypic variation is shown across phyla, within genus' and between species. Enhancer expression patterns in snake limb buds: (Top) mouse, (Bottom) cobra (Kvon *et al.*, 2016). Rock pigeon differences in plumage pigmentation: (Left) *C. livia* T-check breed, (Right) *C. livia* barless breed (Vickrey *et al.*, 2018). Oldfield mice pigmentation differences within a genus: (Left) *P. p. leucocephalus*, (Right) *P. p. subgriseus* (Steiner, Weber and Hoekstra, 2007). *Drosophila* larval trichomes absence and presence between species: (Top) *D. sechellia*, (Bottom) *D. virilis* (Stern and Frankel, 2013). *Drosophila* male-specific wing spot absence and presence: (Left) *D. melanogaster,* (Right) *D. biarmipes* (Gompel *et al.*, 2005). *Drosophila a*dult abdomen colouring differences: (Left) *D. novamexicana*, (Right) *D. americana* (Wittkopp *et al.*, 2009). *Drosophila* dorsalcentral notum bristles differences between species: (Left) *D.melanogaster,* (Right) *D. quadrileneata* (Marcellini and Simpson, 2006). The above phylogenetic branches connecting these examples are for illustrative purposes only and do not reflect the true ancestral relationships between organisms.

The above examples describe coding changes that contribute to morphological diversity, however, there are many cases where the causative changes have been mapped to enhancer sequences (Wray *et al.*, 2003; Hoekstra *et al.*, 2006; Steiner, Weber and Hoekstra, 2007;

Carroll, 2008; Carroll, Grenier and Weatherbee, 2009; Stern, 2011; Villar *et al.*, 2015; Vickrey *et al.*, 2018; Wooldridge *et al.*, 2022).

1.1.1 Enhancer evolution drives phenotypic variation

Enhancers are orientation-independent DNA sequences that encode binding motifs for sequence-specific transcription factors (TFs) (Banerji, Rusconi and Schaffner, 1981; Berman *et al.*, 2002). Much of our understanding of enhancers describes their gene regulatory activities as modular, ensuring spatiotemporal precision during development (Wray *et al.*, 2003; McGregor *et al.*, 2007; Carroll, Grenier and Weatherbee, 2009; Frankel *et al.*, 2011; Stern, 2011). Typically, these sequences are unoccupied by nucleosomes (classed as accessible chromatin) or are bound to chromatin abundant in 'activating' histone marks, all in which enable the successful recruitment of TFs (Barski *et al.*, 2007; Zhao, Zheng and Cvekl, 2019). Upon TF binding to enhancers, these regulatory proteins can then recruit co-repressors, co-activators, and mediator complexes, that interact with and coordinate the promoter's transcriptional response, and thus determine gene expression levels (Shlyueva, Stampfel and Stark, 2014; Richter *et al.*, 2022).

Wing pigment patterns in *Drosophila* have served as a powerful model to study phenotypic evolution (Gompel *et al.*, 2005; Prud'homme *et al.*, 2006; Massey *et al.*, 2020; Ling *et al.*, 2023). *D. biarmipes* has a male-specific distally located pigmented spot on its wings, whereas *D. melanogaster* does not (Figure 1) (Gompel *et al.*, 2005). It was shown that changes in the cis-regulatory sequence acting on *yellow* were responsible for this phenotypic difference, whereby the *D. biarmipes* version of the sequence contained novel Engrailed binding sites (Gompel *et al.*, 2005). Sequence alignment of the *yellow* 'spot' enhancer between *D. elegans* (wing spots) and *D. gunungcola* (no wing spots) further confirmed the importance of cis-regulatory evolution as a driver of morphological change (Prud'homme *et al.*, 2006). The two species shared 92% identity of the 'spot' enhancer of *yellow* (Prud'homme *et al.*, 2006). It was shown that the substitution of just ten nucleotides from the enhancer sequence of *D. elegans* to *D. gunungcola* restored the spot phenotype lost in the latter species (Prud'homme *et al.*, 2006).

The proneural genes, *achaete* (*ac*) and *scute* (*sc*), control bristle development in dipterans (Calleja *et al.*, 2002). *sc* is known to be regulated by a series of enhancers throughout the fly, and the enhancer annotated as 'dorsocentral' specifically regulates *sc* expression at the notum (Gómez-Skarmeta *et al.*, 1995; García-García *et al.*, 1999). Interestingly, anterior dorsocentral

notum bristles are seen in *D. quadrileneate*, but not in *D. melanogaster,* nor *D. virillis* (Figure 1) (Marcellini and Simpson, 2006). *sc* dorsocentral enhancer activity comparisons between all three species revealed that only the *D. quadrileneata* version showed anteriorly expanded activity, subsequently driving *sc* expression and bristle development (Marcellini and Simpson, 2006).

The genetic basis for the loss of legs in snakes serves as another excellent example of enhancer evolution underlying morphological change (Cohn and Tickle, 1999). The morphogen, Sonic hedgehog (Shh), is a critical determining signal for initiating leg growth (Riddle *et al.*, 1993). It was shown that *Shh* is activated during the early development of python leg buds, but expression is not sustained in contrast to leg-bearing animal counterparts (Figure 1) (Leal and Cohn, 2016). This led to the discovery of snake-specific sequence differences in the enhancer acting upon *Shh* expression in limbs, most notably a 17 base-pair (bp) deletion in comparison to limb-bearing animals (Kvon *et al.*, 2016; Leal and Cohn, 2016). Astonishingly, replacement of the mouse version of *Shh* enhancer with the snake ortholog led to substantial limb loss in mice (Kvon *et al.*, 2016).

The genetic basis of morphological diversity is commonly tracked back to enhancer differences between species (Gompel and Carroll, 2003; Erwin and Davidson, 2006; Carroll, 2008). Changes to enhancer sequences have been shown to alter target gene expression levels and patterns during development (Marcellini and Simpson, 2006; Prud'homme *et al.*, 2006; McGregor *et al.*, 2007; Kvon *et al.*, 2016). These regulatory differences are commonly associated with sequence changes at TF binding sites (TFBS), which subsequently modifies TF occupancy at the enhancer and can coordinate differential expression levels between species (Gompel *et al.*, 2005). Enhancer evolution depends on its modularity, where mutations in just one cis-regulatory sequence can deviate gene expression in a tissue-specific manner, and diminish entire organism developmental effects (Sucena and Stern, 2000; Carroll, 2008; Preger-Ben Noon *et al.*, 2017). However, to better understand the evolution of genes and their regulation that underlie phenotypic changes, we firstly require a better characterisation of the components and their interactions in the gene regulatory network (GRN) coordinating development.

1.1.2 Hotspots for phenotypic evolution

So called 'hotspot genes' are genes that are repeatedly prone to cause convergent phenotypic changes across lineages (Martin and Orgogozo, 2013). Changes in these genes can span

coding and non-coding sequences (Martin and Orgogozo, 2013; Yassin *et al.*, 2016; Buchberger *et al.*, 2021). The identification of hotspot genes has relied on a detailed understanding of the GRNs underlying trait development and evolution.

Adult abdomen colouring varies extensively throughout and within *Drosophila* species (Figure 1) (Wittkopp and Beldade, 2009; Dembeck *et al.*, 2015). To understand the genetic basis of pigmentation diversity, many independent approaches comparing different groups of species repeatedly isolated the same two genes, *tan* and *ebony*, which are required for the catalysis of pigment gene synthesis (Bridges and Morgan, 1923; Rebeiz *et al.*, 2009; Wittkopp and Beldade, 2009; Wittkopp *et al.*, 2009; Signor *et al.*, 2016). *D. melanogaster tan* null mutants are rescued by *D. americana* (dark brown abdomens) and *D. novamexicana* (light yellow abdomens) intact *tan* alleles, whereby the colouring of the rescued *D. melanogaster* flies resembled the species origin of the working *tan* allele (Wittkopp *et al.*, 2009). Between species, *tan* was found to exhibit sequence differences in coding and non-coding regions, as was also the case for *ebony* (Wittkopp *et al.*, 2009; Martin and Orgogozo, 2013; Gibert *et al.*, 2016; Signor *et al.*, 2016). *tan* and *ebony* are evolutionary hotspot loci that are recurrently identified when deciphering the genetic basis of intra- and interspecies abdominal pigmentation variation (Martin and Orgogozo, 2013; Gibert *et al.*, 2016; Lamb *et al.*, 2020).

One of the most detailed studies of morphological diversity is the evolution of trichome (hair like projections from the cuticle) patterns in *Drosophila* (Kittelmann *et al.*, 2021). *D. sechellia* exhibits the evolutionary loss of larval trichomes in specific regions of every segment (Figure 1) (Sucena and Stern, 2000; McGregor *et al.*, 2007; Arif, Kittelmann and McGregor, 2015). Central to trichome specification is the hotspot gene *svb* (Sucena and Stern, 2000; Delon, Chanut-Delalande and Payre, 2003). Mutant versions of *svb* in *D. melanogaster* present a naked cuticle phenotype, absent of trichomes, resembling that of its sister species *D. sechellia* (Delon, Chanut-Delalande and Payre, 2003). In *D. sechellia*, five out of seven mapped *svb* enhancers were shown to have reduced activity compared to trichome-bearing species (McGregor *et al.*, 2007). Sequence differences between these species were subsequently mapped to TFBS and found to underpin reduced *svb* expression in *D. sechellia* (McGregor *et al.*, 2007; Frankel *et al.*, 2010; Tsai, Alves and Crocker, 2019). Following this, introducing the mutations of one *D. melanogaster* enhancer into the *D. sechellia* version generated a visible increase in *svb* expression (Frankel *et al.*, 2011). *svb* evolution was also shown to underlie larval trichome patterning in other *Drosophila* species (Sucena and Stern, 2000; Sucena *et al.*, 2003; Frankel, Wang and Stern, 2012). In the *D. virilis* species subgroup, *svb* expression mimicked that seen for *D. sechellia* embryos in four species lacking larval trichomes (Sucena *et al.*, 2003). Upon closer inspection, *svb* enhancer differences identified in *D. sechellia* were

also seen in *D. ezoana* despite evolving independently (Frankel, Wang and Stern, 2012). These examples of convergent evolution suggest that *svb* is a hotspot gene for the morphological diversity of larval trichome patterning (Stern and Frankel, 2013).

As evident from the examples above, there has been great progress so far to characterise and understand the genetic basis of morphological diversity between species. Despite this, the evolutionary forces driving variation still lacks understanding. It is evident too, that organ shape and size evolution is lacking a comparable depth of knowledge to traits like bristle and pigment phenotypic variation. To address this, the study of rapidly evolving organs between closely related species presents an excellent model to begin to decode shape and size variation.

1.2 Rapid evolution of *Drosophila* external male genital structures

Male external genital structures are considered to be among the fastest evolving body parts (Eberhard, 1985; Arnqvist, 1997; Brennan and Prum, 2015; Frazee and Masly, 2015), and is thought this has been driven by sexual selection (Eberhard, 1985; Hosken and Stockley, 2004).

The external male genital structures in the *D. melanogaster* species group are composed of the anus (cercus), posterior lobes, lateral plates, and claspers (Figure 2A-A””). These structures are periphallic and not required for direct sperm transfer (G. Rice *et al.*, 2019), but nevertheless are important for copulation (Jagadeeshan and Singh, 2006). They all exhibit varying degrees of morphological diversity throughout *Drosophila* (Kopp and True, 2002). Within the *D. simulans* clade, *D. simulans and D. mauritiana* periphallic genitalia and cerci, have extensive size, shape, and bristle differences despite diverging just 240,000 years ago (Figure 2C) (Coyne, 1983; True *et al.*, 1997; Garrigan *et al.*, 2012).

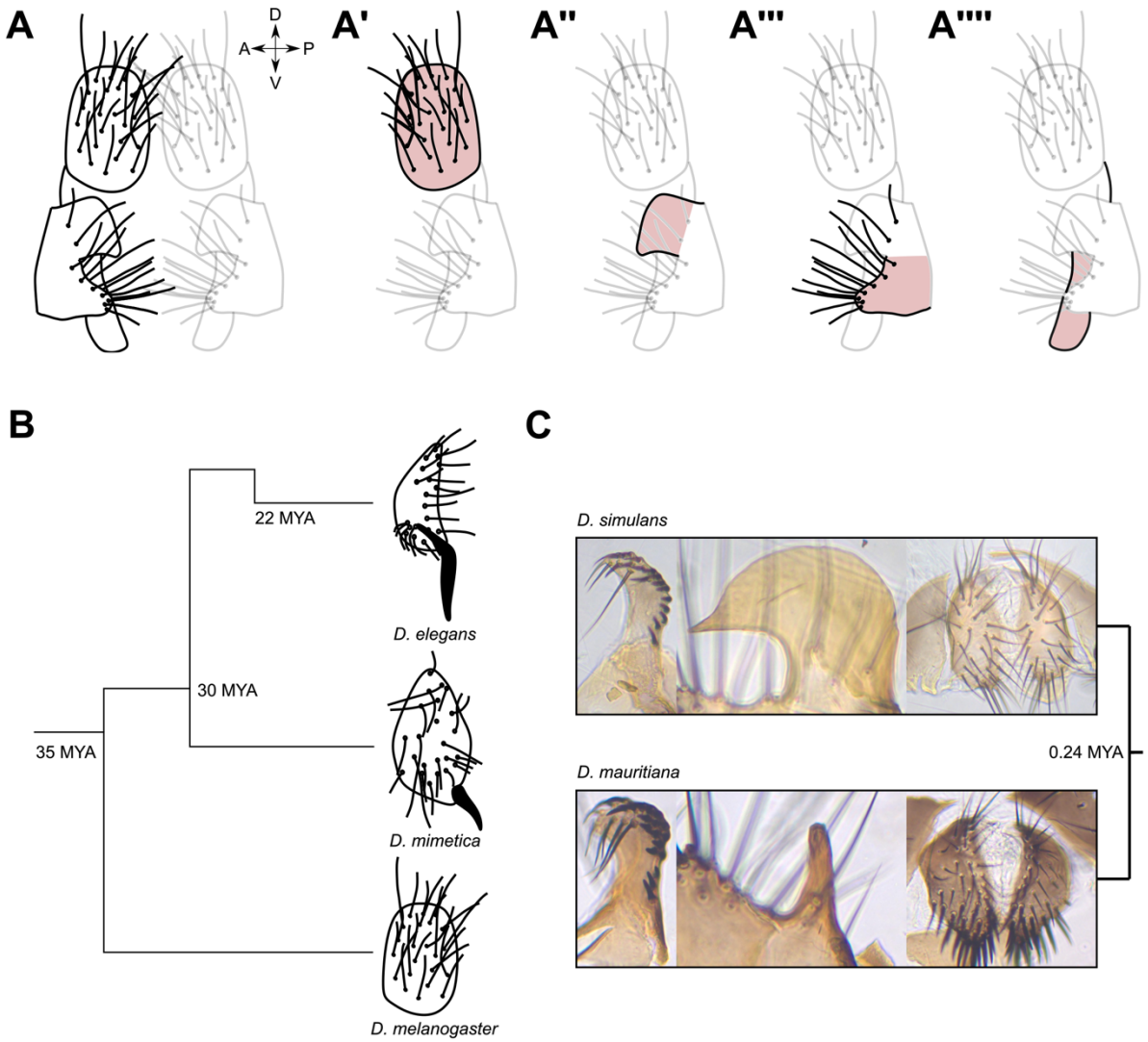


Figure 2: Structural diversity of the periphallic male genitalia among *Drosophila* species. (A) A simplified schematic of *D. melanogaster* external male genital structures, with the dorsal/ventral and anterior/posterior axis orientation provided. The following are highlighted in pink: (A') cercus (anal plates), (A'') posterior lobe, (A''') lateral plate, (A'''') clasper. The male genitalia exhibits bisymmetry along the dorsal-ventral axis. (B) Based on (Kopp and True, 2002), outlines of one half of the bilaterally symmetrical cerci. The prominent melanised bristle is present in *D. elegans* and *D. mimetica* cerci, but not in *D. melanogaster*. Evolutionary timescale based on (Rane *et al.*, 2019). (C) Light microscopy images of *D. simulans* and *D. mauritiana* external male genital structures (Left to Right): clasper, posterior lobe, and cerci (Garrigan *et al.*, 2012). MYA: Million Years Ago.

The cerci (or anal plates) originate from the A10 primordium of the genital disc (Casares *et al.*, 1997; Sánchez and Guerrero, 2001). The cerci are two bilaterally symmetrical ovals, covered in bristles, located at the dorsal position of the male terminalia (Figure 2A') (G. Rice *et al.*, 2019). The *D. mauritiana* cerci are significantly larger and covered in more bristles than in *D. simulans* (Figure 2C). Outside of the *D. simulans* clade, differences in the cerci between

species are even more extreme (Figure 2B). An enlarged individual bristle protruding from the ventral-posterior side of the cerci seen in the *D. elegans* subgroup has convergently evolved in *D. mimetica* (Figure 2B) (Kopp and True, 2002).

The cerci play important roles during copulation. The cerci have been shown to contact the ventral side of the female oviscapts to stabilise male positioning on the female (Kamimura and Mitsumoto, 2011). *D. mauritiana* chromosomal arm 2L introgressions into a *D. simulans* genetic background caused significant enlargement of the cerci, which was described to implicate 59% of the interspecies cercus size differences (Tanaka, Kamimura and Takahashi, 2018). This *D. mauritiana* introgression line showed extended copulation duration with *D. simulans* females compared to *D. simulans* control males, and resulted in ejaculate leakage onto the female outer body surface (Tanaka, Kamimura and Takahashi, 2018). Although the cerci have attracted relatively little attention when studying male genital diversity, there could be great profit in better understanding the genetics underpinning their evolution between species.

The claspers are appendage-like structures in the periphallic male genitalia, patterned with mechanosensory bristles (Figure 2A"") (Jois *et al.*, 2022). The claspers separate from the lateral plates during pupal development at around 30 hours after pupal formation (hAPF) until 10 hours later, by a cellular mechanism not yet understood (Smith, Davidson and Rebeiz, 2020). The GRN coordinating clasper development is not known, however it has been speculated that claspers represent derived walking appendages (See Chapter 5 for more detail). During copulation, the claspers ensure correct copulatory positioning, as well as female oviscapt opening (Robertson, 1988; Jagadeeshan and Singh, 2006; Masly and Kamimura, 2014). In *D. melanogaster*, unilateral ablation of the single long mechanosensory bristle on the distal head of the claspers has been shown to cause an asymmetric mating posture upon copulation with intraspecific females (Acebes, Cobb and Ferveur, 2003).

D. mauritiana claspers are larger and covered in more bristles than *D. simulans* (Figure 2C) (True *et al.*, 1997; Garrigan *et al.*, 2012; Hagen *et al.*, 2019). Genetic mapping between these two species and phenotypic analysis identified *trn* as a gene driving clasper evolution (Hagen *et al.*, 2019). *trn* encodes a leucine-rich repeat membrane protein required for cell-cell interactions (Chang *et al.*, 1993; Paré *et al.*, 2019). To test whether *trn* has evolved between *D. simulans* and *D. mauritiana*, a reciprocal hemizygosity test was performed where phenotyped male progeny only differed in the working species allele of *trn* (*D simulans* or *D. mauritiana*) (Hagen *et al.*, 2019). The *D. mauritiana* working version of *trn* made the claspers

larger in size compared to the reciprocal (Hagen *et al.*, 2019), thought to be mediated by alterations to the cell-cell boundary specification between species (Hagen *et al.*, 2019).

The posterior lobe protrudes from the lateral plate from approximately 36 hAPF until it's fully developed form at 52 hAPF (Figure 2A'' and A''') (Kopp and True, 2002; Smith, Davidson and Rebeiz, 2020). The posterior lobes intercalate between the tergites of the female abdominal VII and VIII segments to aid attachment of the male to the female during copulation (Robertson, 1988; Kamimura and Mitsumoto, 2011; Yassin and Orgogozo, 2013), and the lateral plates assist with achieving the required copulatory posture (Acebes, Cobb and Ferveur, 2003; Jagadeeshan and Singh, 2006). The most striking difference in periphallic structures between *D. mauritiana* and *D. simulans* can be seen at the posterior lobe (Figure 2C). The characteristic 'helmet' shaped lobe in *D. simulans* is much larger and more complex in shape than *D. mauritiana*'s 'finger-like' projection (Figure 2C) (True *et al.*, 1997).

1.2.1 The posterior lobes are a morphological novelty

The posterior lobes are a morphological novelty that has recently evolved in only *D. melanogaster* and the *D. simulans* clade (Kopp and True, 2002). Morphological novelty describes an evolved structure that is not present in the ancestral species (Muller and Wagner, 1991; True *et al.*, 1997; Moczek, 2009), said to originate from modified predisposed GRNs, as seen in butterfly eyespot (Murugesan *et al.*, 2022) and beetle horn development (Moczek and Rose, 2009; Shubin, Tabin and Carroll, 2009).

Characterising the genes and interactions required for posterior lobe development have provided great insights into the evolution of these novel structures. Glassford *et al.,* (2015) observed many similarities between posterior lobe and larval posterior spiracle gene expression patterns during development, which suggested that the GRN underlying posterior lobe development was co-opted from posterior spiracle development. Key evidence was that the enhancers of genes including, *poxn* and *eya*, were active in the spiracles and in the developing posterior lobes (Glassford *et al.*, 2015). However, this study did not pinpoint the upstream gene(s) or enhancer(s) that were mutated to produce the co-opted posterior lobe network in only the *D. melanogaster* subgroup (Glassford *et al.*, 2015). Moreover, little is known about how the GRN for posterior lobe development subsequently evolved to diverse posterior lobe size and shapes in *D. melanogaster* and species of the *D. simulans* clade.

1.2.2 Sexual selection and male posterior lobe diversity

Drosophila male genital structures are strikingly different in size and shape amongst species (Coyne, 1983; True *et al.*, 1997; Kopp and True, 2002). Sexual selection is thought to be the driving force behind this, yet the exact mechanism is unclear (Hosken and Stockley, 2004; Yassin and Orgogozo, 2013; Masly and Kamimura, 2014; Frazee and Masly, 2015). Sexual conflict, sperm competition, good gene female choice (selection based on male genetics to benefit the offspring's fitness) and Fisherian female choice (selection based on arbitrary male traits), are the four mechanisms commonly associated with sexual selection (Hosken and Stockley, 2004). Outside of *Drosophila*, water striders have been shown to elicit varying rates of fertilisation success in response to the male sclerite shape (Arnqvist and Danielsson, 1999; Rowe and Arnqvist, 2012), where the shape variation between species is thought to be driven by a combination of pre and post copulatory sexual selection mechanisms (Rowe and Arnqvist, 2012).

In *D. melanogaster* subgroup females, a 'pouch' located between the oviscapt and tergite VIII varies in size and depth between species (Kamimura and Mitsumoto, 2011). Electron micrographs of the complementary female ovipositor revealed that these pouches are specific to the conspecific male posterior lobe size and shape (Yassin and Orgogozo, 2013). There has been speculation as to whether the complementary pairing of the female pouch with the male posterior lobe is driven by coevolution or antagonistic coevolution (Yassin and Orgogozo, 2013). In addition to sexual selection, natural selection (temperature differences) has also been shown to influence posterior lobe size and shape (House *et al.*, 2013; Brennan and Prum, 2015), and it was proposed that both sexual and natural selection may interact cooperatively to facilitate genital evolution (House *et al.*, 2013).

1.2.3 Importance of posterior lobe shape and size during copulation

As mentioned above, the posterior lobes are an important grasping tool during copulation. Physical ablation and alteration of lobes in *D. simulans* males significantly decreased copulatory success when paired with *D. simulans* females in a competitive environment (LeVasseur-Viens, Polak and Moehring, 2015). A precopulatory effect was also observed in *D. melanogaster poxn* ribonucleic acid interference (RNAi) knockdown males with reduced posterior lobe size, where they failed to copulate with females (Frazee and Masly, 2015). Both cases emphasise the necessity of posterior lobes for successful genital coupling prior to insemination. *D. sechellia* is part of the *D. simulans* clade and has slightly larger and differently shaped posterior lobes compared with *D. mauritiana* (Macdonald and Goldstein, 1999). Subtle

alterations to lobe shape and size were exhibited in *D. mauritiana* introgression lines within a *D. sechellia* genetic background (Frazee *et al.*, 2021). Positioning of these males on *D. sechellia* females generated 45 degrees skew upon normal copulatory alignment, and subsequently experienced shorter durations of copulation (Frazee *et al.*, 2021).

The effect of alterations to the posterior lobe on post-copulation has also been investigated (Frazee and Masly, 2015; LeVasseur-Viens, Polak and Moehring, 2015; Frazee *et al.*, 2021). Although laser ablation of lobes in *D. simulans* did not result in a significant effect on sperm transfer, subtle reduction of the posterior lobe size in *D. melanogaster* males did show a decrease in the number of eggs laid from mated females in comparison to female pairing with males beholding unaltered lobes (Frazee and Masly, 2015; LeVasseur-Viens, Polak and Moehring, 2015). This effect was also seen with *D. sechellia-D. mauritiana* introgression lines (Frazee *et al.*, 2021).

1.3 Thesis Aims

The aim of my thesis was to investigate the GRN underlying male periphallic genital development in *D. melanogaster* and identify how they have diverged to generate morphological variation in the posterior lobes, claspers and cerci. To address this overall aim, my experimental plan was split into the following three chapters:

1. **Chapter Three Results**: *Sox21b* contributes to the rapid diversification of a novel male genital structure between *Drosophila simulans* and *Drosophila mauritiana*

The aim of this chapter was to better understand the genetic basis for differences in posterior lobe shape and size between species.

This was addressed through the following objectives:

- Testing the role of differentially expressed transcription factors between *D. simulans* and *D. mauritiana* during male genital development.
- Characterising the spatial expression profile of the candidate gene *Sox21b.*
- Testing if *Sox21b* underlies posterior lobe evolution using a reciprocal hemizygosity test.

2. **Chapter Four Results**: Characterisation of *Sox21b* enhancers and trans-regulatory landscape between species

The aim of this chapter was to identify any *Sox21b* genital enhancers and understand their evolution and regulation.

This was addressed through the following objectives:

- Screening the *Sox21b* regulatory regions in *D. melanogaster* for male genital enhancer activity.
- Comparing *Sox21b* genital enhancer activity between species.
- Mapping and testing transcription factor binding sites in the *Sox21b* genital enhancer.

3. **Chapter Five Results**: A comparative assessment of *C15*'s role during the development of the terminalia compared to the pretarsal segment of the leg

The aim of this chapter was to characterise the role and regulation of *C15* during genital development compared to previous data from legs, to better understand the GRN underlying clasper development.

This was addressed through the following objectives.

- Characterising the expression and function of *C15* in male and female genital structures.
- Characterising the expression and function of pretarsal leg genes known to interact with *C15* in the developing genital structures.
- Identifying and validating *C15* enhancers and their activity.

2. Chapter Two: Materials and Methods

2.1 GAL4-UAS: RNAi knockdown and overexpression

Selected TFs were assessed for developmental roles in the male and female periphallic genitalia using the GAL4-UAS system to drive RNAi knockdown and Upstream Activating Sequence (UAS) for gene overexpression (UAS-OE) in *D. melanogaster*. *UAS-RNAi* lines for each gene were provided by the Vienna *Drosophila* Resource Centre (VDRC) (Dietzl *et al.*, 2007) and from Bloomington *Drosophila* Stock Centre (BDSC) (NIH P40OD018537). UAS-OE lines were produced in house (see Section 2.7). All *UAS-RNAi* used are detailed in Supplementary Table 3 and Supplementary Table 7. *UAS-RNAi* and *UAS-OE* males were crossed to *NP6333-GAL4 (P(GawB)PenNP6333)* virgin females, (which drives GAL4 expression in all imaginal discs from larval stages and during metamorphosis) also carrying *UAS-Dicer-2 (P[UAS-Dcr-2.D]* (Stieper *et al.*, 2008). *Sox21b* and *E(spl)m3-HLH* RNAi were also carried out using the *POXN-GAL4 (14.1.1)* driver, which specifically drives in the posterior lobe primordia in larval genital discs and subsequently in pupae (Boll and Noll, 2002). *Sox21b* Enhancer 1 (GMR73B05) and Enhancer 2 (GMR73A06) GAL4 lines sourced from BDSC were also used to drive UAS-*Sox21b* RNAi in specific regions of the male periphallic genitalia. All crosses were carried out at 25 ºC, with *E(spl)m3-HLH* repeated at 29 ºC to enhance the effect (Blake *et al.*, 2017). All crosses were performed using a 1:2 male to female ratio. Crosses were transferred to standard cornmeal food every two days and maintained in a 12-hour light/dark cycle incubators. The same conditions with a sample size of n > 14, were phenotyped for each cross. Progeny were aged at least 3 days and collected and stored in 70% Ethanol (EtOH) at -20 ºC for phenotyping.

2.1.1 Phenotyping

The cerci, posterior lobes and claspers of the adult male genitalia, and the epiproct, hypoproct and ovipositor of adult female genitalia, were both dissected in Hoyer's solution (Stern and Sucena, 2012). This was done using 0.14 mm diameter stainless steel pins followed by mounting in Hoyer's solution on slides containing eight individual 6 mm diameter wells. To account for body size, the T2 legs of each fly were also dissected and mounted. A Zeiss Axioplan light microscope with a Jenoptik ProgRes C3 camera was used to image each dissected structure. 250X magnification was used for the genital structures, and 160X magnification for the T2 legs. The area of posterior lobe, cercus, lateral plate, hypoproct,

epiproct and length of T2 tibia leg and posterior lobe base width were measured manually using ImageJ (Schneider, Rasband and Eliceiri, 2012). The bristles were counted using the light microscope and a tap counter. When drawing the outline of the posterior lobe, an artificial baseline was used as previously described (Hackett *et al.*, 2016). This method was also used when drawing the outline for the lateral plate, hypoproct and epiproct too. The area, bristle count and length were recorded for both structures of each pair per individual and the average was then used for the statistical analysis.

2.1.2 Statistical analysis

All statistical analyses were carried out using R version 4.2.0 (R Core Team, 2020). Measurements of each structure described above were first assessed for normality using the Shapiro Wilk test (Shapiro and Wilk, 1965). Normally distributed data were analysed using Dunnett's test and ANOVA, whereas the Kruskal Wallis test was used for non-normally distributed data (Hothorn, Bretz and Westfall, 2008). RNAi knockdowns and UAS-OE lines were compared to both parental controls. If this test was evaluated as statistically significant, the Tukey's Test / Wilcoxon Rank Sum Test (BH p-adjusted method (Benjamini and Hochberg, 1995)) was used to identify if the effect was significantly different to both parental controls (Supplementary Table 3, Supplementary Table 7 and Supplementary Table 8). Results were concluded as non-significant when p =/> 0.05 or the effect detected in the RNAi knockdown or UAS-OE was an intermediate of the two parental controls (Supplementary Table 3, Supplementary Table 7 and Supplementary Table 8).

Where the T2 tibia length of the RNAi knockdown or UAS-OE line was significantly different to both parental controls following the tests described above, Pearson/Spearman correlation analysis was performed on the cercus, posterior lobe, epiproct and hypoproct area of these crosses dependent on Shapiro Wilk test results (Supplementary Figure 17A-B'). If statistically significant, the area of the structure was divided by the tibia length squared, and statistical tests were carried out on the normalised measurements.

2.1.3 Scanning electron microscopy

Scanning electron microscopy (SEM) was carried out on *NP6333-GAL4* and *C15* RNAi knockdown male and female terminalia, as well as *Sox21b* RNAi knockdown in only male terminalia. Flies stored in 70% EtOH were moved to 100% EtOH at least 24 hours prior to imaging. The posterior of the fly was dissected in 100% EtOH. Samples were processed in a

critical point dryer and mounted on SEM stubs, then gold coated for 30 seconds. The genitalia were imaged using SE mode at 5 kV in a Hitachi S-3400N SEM, with a working distance of 13 to 14 mm. Whole genitalia was imaged at a magnification of 250X, and individual periphallic structures at 900X.

2.2 Immunohistochemistry staining

Immunohistochemistry was used to detect protein expression patterns. Sample preparation, antibody staining and mounting protocols differed depending on the developmental timepoint used, as detailed below. For all stains, the primary and secondary antibody concentrations and emission wavelengths are detailed in Supplementary Table 1. Where annotated as preabsorbed primary antibody (Supplementary Table 1), the working concentration of the primary antibody was diluted in 0.3% TRITON-X-100 in PBS (PBTX) for pupae, and 0.1% Tween-20 in PBS (PBT) for larvae and embryos, before being incubated with inverted, fixed larvae for 1 hour at room temperature, then stored at 4 ºC before use. All samples were fixed in 4% formaldehyde in PBS prior to staining. Primary antibody stains for all stages were incubated overnight at 4 ºC.

2.2.1 Collection, fixation and staining of embryos

D. melanogaster w^{1118} was set up in 'cages' consisting of a plastic cylinder with a netted covering at one end, and an apple juice agar plate with yeast paste at the opposing end. Flies were allowed to lay embryos for up to 20 hours. Embryo samples were collected and dechorionated in a 50% bleach solution for 2 minutes, then washed in distilled water. A solution of 4% formaldehyde, and equal volume heptane, was used to immerse embryos and shaken for 20 minutes. Following fixation, embryos were devitellinised using MeOH, washed, and then stored in MeOH at -20 ºC.

Embryos were washed in 0.1% PBT, and blocked in 10% normal goat serum (NGS) in 0.1% PBT for 1 hour at room temperature whilst rocking. Following primary antibody incubation, embryos were washed, followed by addition of the secondary antibody, and incubated at room temperature for 2 hours. Samples were then washed, DAPI stained, and mounted in VECTASHIELD and sealed with a coverslip.

2.2.2 Larval discs

If the timepoint was specified for the larval stain, parental flies were crossed in a vial for 2 hours, then removed. This was denoted as 0 hours after egg laying (hAEL), and larvae were then collected at the timepoint required. If the timepoint was not specified in the staining, late L3 wandering larvae were used. Larvae were collected in ice cold PBS, cut in half and inverted. Larval samples were washed and stained for in 0.1% PBTX and blocked as described for embryos. Samples were fixed at room temperature for 20 minutes then washed. They were then incubated in 0.3% PBTX for 30 minutes and blocked for 20 minutes. Larval samples were then incubated with the primary antibody (Supplementary Table 1).

Samples were washed five times for 5-minute durations. The secondary antibody was added and incubated at room temperature for 2 hours. Samples were washed accordingly, DAPI stained, then dissected on slides to collect just the imaginal discs required. Discs were mounted in VECTASHIELD and sealed with a coverslip for imaging.

2.2.3 Pupal terminalia

Male and female white prepupae were defined as 0 hAPF. Males were identified by two circular transparent cuticles on either side of the posterior abdomen (Bainbridge and Bownes, 1981), and then stored at 25 ºC until the stage required. Using micro-scissors, pupae were cut near the posterior end in ice cold PBS, then flushed with PBS to remove as much fat as possible, followed by fixation for 30 minutes at room temperature. Samples were washed in PBS, then gradually brought to 100% EtOH for storage at -20 ºC.

Before staining, the membranes were removed from the pupal casing, leaving just the exposed genital tissue samples. All steps following this were carried out in 0.3% PBTX, and blocked in 2% NGS. Pupal samples were blocked for 2 hours, then incubated with the primary antibody. Unbound primary antibody was washed in 6 x 20-minute intervals, before addition of the secondary antibody, and then incubated at 4 ºC overnight. Samples were subsequently washed, DAPI stained if need be, then the outer membrane was removed using dissection forceps prior to mounting.

Microscope slides with coverslips used as walls were required to create a well to protect the sample from being crushed during imaging. Poly-L-Lysine solution was added to the well prior to mounting to avoid movement from the tissue. All pupal samples were mounted in Hydromount and a coverslip was placed on top with nail varnish to seal it in place.

2.2.4 Imaging

All stained samples were imaged at Oxford Brookes University Bioimaging Facility on a Zeiss LSM800 upright Confocal Laser Scanning Microscope. Laser excitation values aligned with the secondary antibody's fluorophore (Supplementary Table 1). A 20X air objective was used for embryo and larval disc samples, and a 40X oil objective for pupal samples. Alternatively, some pupal structures were imaged using a 63X oil objective when capturing individual periphallic structures. Z-stacks with 0.7 µm and 1 µm separation were used for pupal and larval/embryo samples, respectively. All images were processed with Fiji (Schindelin *et al.*, 2012).

2.3 Hybridisation chain reaction *in situs* (HCR-IS)

HCR-IS was used to detect mRNA localisation in embryos, genital discs at 96 hAEL and 120 hAEL, and pupal terminalia. All samples were prepared, fixed, mounted and imaged as described above for staining. The HCR-IS protocol was based on established protocols (Choi *et al.*, 2018; Younger *et al.*, 2022). The probe from Molecular Instruments to target *Sox21b* was designed as 20 individual probes spanning the entirety of the gene, ensuring that all isoforms were captured (Schwarzkopf *et al.*, 2021). The *wg* probe is a standard probe available from Molecular Instruments (Schwarzkopf *et al.*, 2021). A 16 nM probe solution was used, and the samples were incubated for 24 hours at 37 °C on an orbital shaker at 60 RPM. Following this, hairpins complementary to the probe initiator sequences were used for signal detection accordingly (Choi *et al.*, 2018; Schwarzkopf *et al.*, 2021). Note, we used half volumes of each solution (Choi *et al.*, 2018; Younger *et al.*, 2022). DAPI was diluted in the probe wash solution prior to staining.

2.4 Generation of *Sox21b* mutants using CRISPR-Cas9

To test if *Sox21b* has evolved between *D. simulans* and *D. mauritiana*, I designed and generated transgenic flies that could be used for reciprocal hemizygosity tests. Clustered Regularly Interspaced Short Palindromic Repeats with associated protein 9 (CRISPR-Cas9) directed and catalysed the insertion a 3XP3-DsRed fluorophore into the coding region of the two species alleles of this gene to disrupt the reading frame.

2.4.1 CRISPR design and cloning

The guide RNA (gRNA) target site was designed 152 bp into *Sox21b* exon 1 to avoid the conserved high mobility group (HMG) box domain shared between Group B Sox genes (Figure 11B) (McKimmie, Woerfel and Russell, 2005). It was also ensured that the target region was flanked by DNA absent of restriction sites that would have been targeted by restriction enzymes in later cloning steps (Gratz *et al.*, 2015). gRNAs were designed using FlyCRISPR (Gratz *et al.*, 2014) whereby the same gRNA was used for both species alleles (Supplementary Table 2). The sense and antisense gRNAs were phosphorylated and inserted into pCFD3 plasmid (Addgene plasmid # 49410) (Gratz *et al.*, 2015) using *BbsI* restriction digest and T4 DNA Ligase (Gratz *et al.*, 2015).

Homology arms were designed as 1 (kilobase) Kb in length and spanned adjacent to either side of the cut site (Supplementary Table 2) (Gratz *et al.*, 2015). Homology arms were amplified from salt extracted *D. simulans*^{w501} and *D. mauritiana D1* genomic DNA using Q5 Polymerase (Miller, Dykes and Polesky, 1988), and subsequently inserted into pHD-DsRed plasmid (Addgene plasmid # 133559) (Gratz *et al.*, 2014) using *AarI* and *SapI* restriction digest and T4 Ligase. Both constructs were transformed into Dh5alpha competent cells for isolation of plasmids. Both plasmids were sequenced to verify correct insertions prior to injection.

2.4.2 Embryo injections

The introgression line 108 (*IL108)* (*D. mauritiana D1* introgression region spanning 3L: 12,277,961-15,075,323 in a *D. simulans*^{w501} genetic background) (Hagen et al., 2021), was crossed to *D. simulans*^{w501} *nanos>cas9;;* to generate a homozygous stock (*nanos>cas9;;IL108*) verified by sequencing and restriction digests. The *nanos* promoter drives expression in the germline (Ren *et al.*, 2013). *D. simulans*^{w501} and *IL108*, both carrying endogenous *nanos>cas9* on the X chromosome, were subsequently injected (Gompel and Ayla Schröder, 2015) with 200 ng/µl of gRNA-pCFD3 and 500 ng/µl of HA-pHD-DsRed (using a Eppendorf FemtoJet 4i and Leica light microscope). 48 hours prior to injection, cages were set up as described previously, which were changed twice per day prior to injections.

2.4.3 Validation of DsRed insertion

Surviving adults from injected embryos were backcrosses to non-injected adults of the same strain. Progeny were then screened for the DsRed eye marker using a Zeiss Axiozoom microscope using the mRFP filter. The *D. simulans* line was established as homozygous viable, whereas the *IL108* line remained heterozygous for the *Sox21b-DsRed* allele. Positive

transgenics were sequenced using primers designed to amplify exon 1 (Supplementary Table 2), where DsRed was inserted and caused a frameshift mutation (Figure 11C).

2.4.4 Reciprocal hemizygosity test and phenotyping

Transgenic *D. simulans*^{w501} and *IL108* males with the disrupted *Sox21b-DsRed* alleles were crossed to non-injected *IL108* and *D. simulans*^{w501} female virgins, respectively, to generate reciprocal hemizygotes (Figure 12A). Reciprocal hemizygote crosses were set up with same male to female ratio, temperature and vial transfer schedule as described for the GAL4-UAS crosses. Adult males, at least three days old, were collected and phenotyped for posterior lobe area, T2 tibia length and posterior lobe base width. Reciprocal hemizygotes and null mutants (homozygous *D. simulans*^{w501} *nos>cas9;;Sox21b-DsRed* males) phenotypic data were analysed using an independent t-test or Wilcoxon Rank Sum Test dependent on the normality of the data (Supplementary Table 4).

2.4.5 Morphometric analysis of posterior lobe shape

The posterior lobe shape was analysed between *Sox21b* reciprocal hemizygotes to characterise morphometric changes induced by species-specific alleles. One randomly chosen posterior lobe from each strain was manually traced on ImageJ (Schindelin *et al.*, 2012) and exported as a region of interest (ROI). Collective ROIs from each reciprocal hemizygote were processed using PAT-GEOM (Chan, Stevens and Todd, 2019) for Elliptical Fourier Analysis (EFA). PAT-GEOM performs morphometric analysis independent of scale, rotation, trace-point start and translation using the readily available Elliptical Fourier descriptors (Chan, Stevens and Todd, 2019). Twenty harmonics were assigned to each ROI for EFA (Supplementary Table 5). This same approach was replicated for pure *D. simulans* and *D. mauritiana* lines (Supplementary Table 6).

Principal component analysis (PCA) was performed in R using factoextra and prcomp (Kassambara and Mundt, 2020), to summarise shape variation between lines. This package standardised the data to have a mean of zero and variance of one before PCA was conducted. Each PC generated were assessed for their Eigenvalues, and those with a value above one were retained (Supplementary Figure 3B). A MANOVA test was used to assess the significance of the PCs between reciprocal hemizygotes.

2.5 Janelia GAL4 enhancer reporter lines

To identify genital enhancers, male Janelia GAL4 reporter lines (Figure 18A) (Jenett *et al.*, 2012) that spanned the *Sox21b* and *C15* loci were crossed to *UAS-GFP* virgins females (Figure 19 and Figure 52). Resultant progeny were screened for Green Fluorescent Protein (GFP) expression following the same sample preparation, immunohistochemistry protocol and imaging technique as described above for genital discs and pupal terminalia. Note anti-GFP primary antibody was used to detect GAL4 driven GFP expression (Supplementary Table 1).

2.5.1 Pupa preparation and live imaging

Activity in live pupae was analysed for all Janelia enhancer reporter lines. Pupae were collected based on hAPF as described previously. Upon collection, the outer cuticle was removed using dissection forceps, and pupae were mounted on a slide in a drop of 80% glycerol for imaging. The entire pupa was imaged using Zeiss Axiozoom microscope GFP filter. Dorsal and ventral images of the pupa were captured (Figure 19 and Figure 52). Z-stacks of 1 µm separation were used to image expression driven by *Sox21b* Enhancer 1, 2 and 3, and *C15* enhancers (Figure 21, Figure 27A and Figure 52).

2.6 *Sox21b* Enhancer 3 GAL4 reporter lines

The 811 bp overlap region between *D. melanogaster Sox21b* Enhancer 1 and 2 was characterised as *D. melanogaster* Enhancer 3. This region was also identified in *D. mauritiana* and *D. simulans* by aligning Enhancer 1 and 2 to these species genomes and extracting the overlap region.

2.6.1 Generation of pBPGUw Enhancer 3 constructs

All Enhancer 3 species-specific alleles and AbdB TFBS mutant sequences were sent to GENEWIZ from Azenta for oligonucleotide generation (Supplementary Table 2). Sequences were generated in the negative orientation consistent with its genomic structure, like the Janelia lines (Jenett *et al.*, 2012). Initially, 5' *AatII* and 3' *KpnI* restriction sites were added to all species versions of Enhancer 3 oligonucleotides accordingly for directional insertion into the pBPGUw plasmid (Addgene plasmid # 17575). However, upon arrival of constructs from GENEWIZ, it was apparent that the *KpnI* restriction site removed the minimal promoter region from the pBPGUw plasmid, preventing downstream GAL4 expression.

Primers were therefore designed to amplify the minimal promoter from original pBPGUw plasmid, whereby the forward primer contained an overhang complementary to each enhancer insert required for latter cloning steps (Supplementary Table 2). The minimal promoter was amplified using HiFi polymerase. pBPGUw plasmids containing the enhancer lines were digested using KpnI. The digested plasmid and amplified minimal promoter were assembled using HiFi Gibson Assembly. The resulting constructs were transformed into DH5alpha competent cells and the extracted plasmids were sequenced to verify insertion of the minimal promoter region (Supplementary Table 2).

2.6.2 Injections and screening

The resultant Enhancer 3-GAL4 sequence was flagged either side by attB sites required to facilitate subsequent attp2 recombination. Constructs containing the pBPGUw plasmid with inserted Enhancer 3 alleles were injected by BestGene into *D. melanogaster* lines that expressed endogenous *nanos>phiC31 integrase* and contained the attp2 landing site. This is the same insertional location used for the Janelia lines (Jenett *et al.*, 2012).

2.7 Generation of UAS lines

UAS-OE lines for *Sox21b*, *C15*, *bowl*, *Dichaete* and *al* were generated.

2.7.1 Cloning strategy

pDONR221 entry vectors containing the coding sequence of each TF were previously established, thanks to Korneel Hens. Isoform A was used for all genes, apart from isoform D for *bowl*.

LR gateway cloning was used to insert the TF coding sequences from the entry vectors into the destination vectors (pGW-HA.attB for all genes) (Bischof *et al.*, 2013). The LR reaction (Untergasser, 2006) was left for 4 hours instead of 1 hour at room temperature. The constructs were transformed into Dh5alpha competent cells, and plasmids were isolated for sequencing using the HSP70 primer for insertion verification (Supplementary Table 2).

2.7.2 Injections

UAS-Dichaete and -bowl were injected in-house (as described above for the reciprocal hemizygote lines), into *nanos>phic31 integrase;;attp2*. *UAS-Sox21b*, *-al* and *-C15* were injected into the same line by BestGene. Transgenic progeny were screened for w+ marker in the eye, and lines were made homozygous.

2.8 Bioinformatic analysis

2.8.1 ATAC-seq processing

Assay for Transposase-Accessible Chromatin and sequencing (ATAC-seq) data was generated by Hagen *et al.,* (unpublished) from *D. simulans*^{w501} and *D. mauritiana D1* pupal terminalia. Bedgraph files were created by Javier Figueras-Jimenez. The R package 'sushi' was used to visualise ATAC-seq data against species-specific genomic coordinates accordingly (Phanstiel *et al.*, 2014). ATAC-seq data for T1 and T2 *D. melanogaster* w^{1118} leg discs were acquired from (Feng *et al.*, 2022).

2.8.1.1 Footprint analysis for *Sox21b*

Transcription Factor Occupancy Prediction by Investigation of ATAC-seq Signals (TOBIAS) V0.8.0 was used to assign footprints at the *Sox21b* locus (Bentsen *et al.*, 2020). ATAC-seq bedgraph files were converted to Bigwig files using the programme, bedGraphToBigWig. Footprint and ATAC-seq files were visualised on IGV V2.8.13 (Robinson *et al.*, 2011) to identify regions predicted to be bound by TFs.

2.8.2 TF binding site search

2.8.2.1 *Sox21b* Enhancer TFBS mapping

JASPAR was used to acquire position weight matrices (PWM) for each target TF (Castro-Mondragon *et al.*, 2022). PWMs are used to represent binding motifs likelihood of a TF from experimental and non-experimental data (Wasserman and Sandelin, 2004). Find Individual Motif Occurrences (FIMO) on MEME suite was used to search significant PWM matches with *Sox21b D. melanogaster* Enhancer 3 input sequence (Grant, Bailey and Noble, 2011).

2.8.2.2 *C15* Enhancer TFBS mapping

The CIS-BP database was used to acquire PWM for each target TF (Weirauch *et al.*, 2014). The following R programmes were used to scan PWMs against the *C15* enhancer sequence: Biostrings (Pagès, Aboyoun and Gentleman, 2023), universalmotif (Jean-Marie Tremblay, 2021), knitr (Xie, 2022), svglite (Wickham *et al.*, 2022) and kableExtra (Zhu, 2021). A significance threshold of 0.001 was used to identify mapped TFBS. The binding site coordinates were generated and subsequently mapped back to the genome (Supplementary Table 10).

2.8.3 Multiple sequence alignment

Sequences used for multiple sequence alignment (MSA) of Enhancer 3 were sourced from: *D. simulans* (Prin_Dsim_3.1), *D. melanogaster* (R6.48), *D. mauritiana* (ASM438214v1), *D. biarmipes* (RU_DBia_V1.1), *D. yakuba* (Prin_Dyak_Tai18E2_2.1), and *D. ananassae* (ASM1763931v2). MSA was performed using Clustal Omega (Supplementary Figure 10 and Supplementary Figure 14) (Madeira *et al.*, 2022). MSA between Abd-B site analysis was performed using MAFFTv7 6-mer distance (Supplementary Figure 13).

2.8.4 Pairwise sequence alignment

Sequences used for pairwise sequence alignment were sourced as outlined for MSA. EMBOSS needle alignment was used for *D. simulans* versus *D. mauritiana Sox21b* Enhancer 3 comparisons (Supplementary Figure 11) (Madeira *et al.*, 2022). EBLOSUM62 alignment algorithm was used to align *Dichaete* and *Sox21b* protein sequences (Supplementary Figure 6) (Madeira *et al.*, 2022).

3. Chapter Three Results

Sox21b contributes to the rapid diversification of a novel male genital structure between *Drosophila simulans* and *Drosophila mauritiana*

3.1 Background: Mapping studies to resolve genetic regions contributing to morphological diversity in *Drosophila* male genital structures

The rapid evolution of male external genital structures has led to striking phenotypic variation between species (Coyne, 1983; Eberhard, 1985; Arnqvist, 1997; Kopp and True, 2002; Brennan and Prum, 2015; Frazee and Masly, 2015). What genetic and developmental changes have occurred between species to generate these striking phenotypic alterations? Quantitative trait locus (QTL) and introgression mapping between two or more *Drosophila* species have paved the foundations to identifying potential candidate genes required for genitalia development and evolution (Coyne, 1983; Liu *et al.*, 1996; Laurie *et al.*, 1997; True *et al.*, 1997; Macdonald and Goldstein, 1999; Masly *et al.*, 2011; McNeil, Bain and Macdonald, 2011; Peluffo *et al.*, 2015; Tanaka *et al.*, 2015; Hackett *et al.*, 2016; Hagen *et al.*, 2019, 2020; Frazee *et al.*, 2021; Macdonald *et al.*, 2022). For example, between *D. simulans* and *D. mauritiana*, previous QTL mapping of posterior lobe differences revealed a minimum of nineteen QTLs throughout the genome (Zeng *et al.*, 2000), and five for the claspers, which contribute 34.59% towards bristle number and 22.99% size variation (Tanaka *et al.*, 2015). In addition, up to seven QTLs associated to cercus size differences, and three QTLs for cercus bristle variation, have also been identified (True *et al.*, 1997).

Subsequently, introgressed regions covering QTLs increased mapping resolution (Tanaka et al., 2015; Tanaka, Kamimura and Takahashi, 2018; Hagen et al., 2019, 2021; Frazee et al., 2021). Hagen *et al.,* (2020) found five regions on 3L, contributing 29.3% of the posterior lobe size difference between *D. simulans* and *D. mauritiana* but these regions still contained hundreds of candidate genes. A 1.1 megabase (Mb) region termed P5 had a significant effect upon posterior lobes, where introgression of *D. mauritiana* DNA reduced posterior lobe area (Figure 11A) (Hagen et al., 2021). This suggested that inter-specific variation in a gene(s)

within this region contributes to part of the difference in lobe size between these two species (Hagen *et al.*, 2020).

3.1.1 Chapter three aims

To better understand how the male genital structures have evolved between *D. simulans* and *D. mauritiana*, I therefore used a candidate gene approach to interrogate Ribonucleic Acid Sequencing (RNA-seq) data and genomic regions identified by previous introgression mapping on chromosomal arm 3L (Hagen et al., 2021). I focussed on genes encoding TFs because in many previous studies they have been shown to occupy key nodes in GRNs and contribute to morphological evolution (Kerber *et al.*, 2001; Williams *et al.*, 2008; Carroll, Grenier and Weatherbee, 2009; Chan *et al.*, 2010; Stern, 2011; Stern and Frankel, 2013; Rizzo and Bejsovec, 2017; Zhang, Mazo-Vargas and Reed, 2017; López-Varea *et al.*, 2021; Quiquand *et al.*, 2021). In this chapter, I describe how this approach led to the identification of *Sox21b* as a new gene involved in genital development and evolution.

- Testing the role of differentially expressed TFs between *D. simulans* and *D. mauritiana* during male genital development.
    - RNAi knockdown and UAS overexpression crosses were generated to identify those required for genital development.

- Characterising the spatial expression profile of the posterior lobe candidate gene *Sox21b.*
    - HCR-IS *Sox21b* stains were conducted during male genital development to characterise its expression pattern and differences between species.
    - Staining was performed during embryogenesis to see whether *Sox21b* is co-opted from the ancestral posterior spiracle GRN.

- Testing if *Sox21b* underlies posterior lobe evolution using a reciprocal hemizygosity test.
    - CRISPR was used to generate mutant *Sox21b* lines to test whether it has evolved between *D. simulans* and *D. mauritiana*.
    - Phenotypic analysis was conducted on the male genitalia of the reciprocal hemizygotes to assess whether working species alleles of *Sox21b* caused posterior lobe shape and size differences.

- Investigating whether other Group B Sox genes are required for posterior lobe development.
    - RNAi knockdown, UAS overexpression and staining of *Dichaete* was performed to identify a role during male genital development.

3.2 Results: Transcription Factors differentially expressed between species contributes to the development of periphallic male genital structures

Previous analysis of RNA-seq data between *D. simulans*^{w501} and *D. mauritiana D1* developing male genitalia revealed a total of 49 differentially expressed TF-encoding genes across all four chromosomes (Hagen et al., 2021). 24 of these candidates were located on chromosome 3, a genomic region shown previously to contribute to clasper and posterior lobe divergence (Hagen *et al.*, 2019, 2020). Five out of the 49 candidates are also characterised as one of the top 100 expressed TFs during *D. melanogaster* male genitalia development, in which three of these five are also located on chromosome 3 (Vincent *et al.*, 2019).

This resulted in a list of 26 candidate genes that I functionally assessed for a role in male genitalia development with RNAi knockdown in *D. melanogaster* using the *NP6333-GAL4* driver. This driver is expressed throughout the developing male terminalia from larvae to late pupae (Stieper *et al.*, 2008) (Table 1). RNAi knockdowns were generated using both VDRC and TRiP *UAS-RNAi* lines, where available, to prevent the identification of false positives and negatives generated from off-target and no-target effects, respectively (Supplementary Table 3). I phenotyped the clasper bristle count, posterior lobe area, cercus bristle count, and cercus area in the RNAi knockdowns, as well as in the parental lines, for statistical comparison. Alongside this, I measured the T2 tibia length as a proxy for body size control. Note, RNAi knockdown flies did not generally differ in T2 tibia length compared to controls, consistent with the hypoallometric nature of the genitalia (Supplementary Table 3) (Dreyer and Shingleton, 2011).

	Differentially Expressed Transcription Factors[1]				
	Periphallic Structure Significantly Altered				No Effect
	Posterior Lobe	Clasper	Cercus		
	Area	Bristle Count	Area	Bristle Count	
Upregulated in *D. mauritiana D1*	*tna* *MED24* *Sox21b* *E(spl)m3-HLH*	*tna* *MED24* *Ssb-C31A* *CG17359*	*Asciz* *E(spl)m3-HLH* *CKIIalpha-il*	*Ssb-C31A* *MED24* *Su(z)12*	*CG10147* *Meics* *hng3* *odj* *pnr* *CG6276* *Myb* *E(var)3-9*
Upregulated in *D. simulans*^{w501}			*knrl* *MED10*		*Taf2* *CG13894* *mirr* *mu2* *CG6843* *nom* *ich*

Table 1: Summary of the periphallic male genital phenotypes upon RNAi knockdown of selected differentially expressed TFs in *D. melanogaster*. Results were identified as significant if statistically different to both parental controls, and not an intermediate phenotype of the two ($p<0.05$) (Supplementary Table 3). Red text indicates a decrease in the phenotype measured upon RNAi knockdown, and blue for an increase in the structure measured, relative to both parental controls. [1] Data generated and identified from (Hagen et al., 2021).

RNAi knockdown of eleven out of 26 TF-encoding genes were found to significantly alter the size and/or bristle count of the posterior lobe, claspers and/or cercus compared to parental controls (Table 1). Nine of these genes disrupting male genital structures are expressed at a significantly higher level in *D. mauritiana D1* developing genitalia than *D. simulans*^{w501} (Table 1) (Hagen *et al.*, 2020). Seven genes tested affected just one structure, while *MED24, tna, E(spl)m3-HLH* and *Ssb-C31A* knockdown affected multiple structures (Table 1 and Supplementary Table 3). Only RNAi against *CKIIalpha-i1, knrl, MED10* (all cercus area) and *Sox21b* (posterior lobe area) altered structures consistent with the direction of the difference in their expression and phenotype in *D. mauritiana D1* compared to *D. simulans*^{w501} (Table 1).

3.2.1 Posterior lobe RNAi knockdown effect genes

From the 26 genes tested, only four produced a significant alteration to posterior lobe size when compared to both parental controls (Figure 3, Table 1 and Supplementary Table 3). The most extreme increase in size was demonstrated by *Sox21b*, further analysis of which will be presented below. Whereas the most significant posterior lobe size decrease resulted from *tna* and *MED24* RNAi knockdown (Figure 3A). *tna* encodes a TF that associates with chromatin remodelling complexes (Gutiérrez *et al.*, 2003). Interestingly, it has been previously noted that it plays a role in modulating Hox gene expression in the leg and the wing (Rosales-Vega *et al.*, 2018). My data reveals a positive regulatory role for *tna* upon posterior lobe growth in *D. melanogaster*, possibly coordinating expression of the terminalia-specific Hox gene *Abd-B* (Figure 3B).

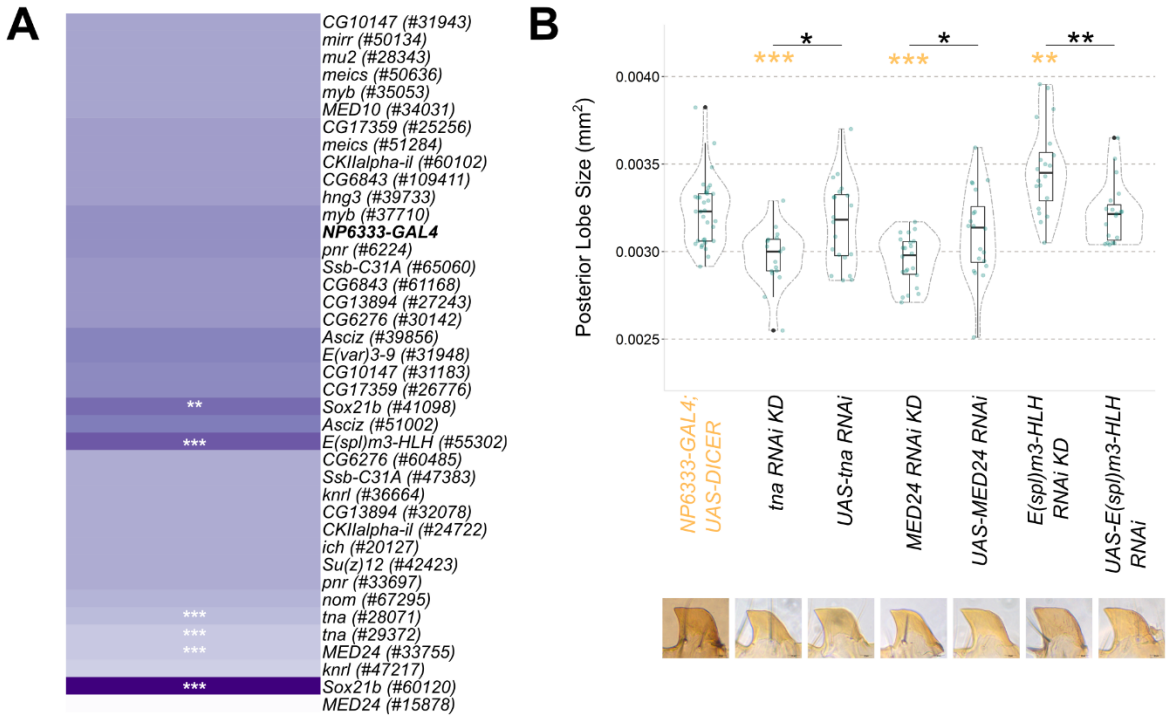


Figure 3: RNAi knockdown posterior lobe effects. (A) Heatmap of the posterior lobe size RNAi knockdown effects relative to *NP6333-GAL4, UAS-DICER*. Those significantly different to both parental controls in the same direction are annotated with asterisks. Darker shading indicates an increase in lobe size, and lighter shading indicates a decrease in lobe size, both relative to *NP6333-GAL4, UAS-DICER*. (#XXXXX indicates stock centre number (Supplementary Table 3)). (B) RNAi knockdowns of TFs where the posterior lobe size is significantly different to both parental controls. Knockdown (KD). n

> 17 for each genotype assayed. Asterisks represent the level of significance (* p<0.05, ** p<0.01, *** p<0.001). All RNAi knockdowns were driven by *NP6333-GAL4, UAS-DICER*.

E(spl)m3-HLH increased the size of the posterior lobe upon RNAi knockdown, revealing a repressive regulatory function during development in *D. melanogaster* (Figure 3). This was also validated by a significant increase in lobe size at a higher temperature, as well as with a different GAL4 driver which also increased lobe size (Supplementary Figure 1A). *E(spl)m3-HLH* is one of the top 100 expressed TFs in *D. melanogaster* developing male genitalia and is expressed at the base of the posterior lobe during late pupal development (Vincent *et al.*, 2019). It is also more highly expressed in *D. mauritiana* than *D. simulans* consistent with the species difference in phenotype exhibited (Table 1 and Supplementary Table 3).

3.2.2 Clasper RNAi knockdown effect genes

From the 26 genes tested, only three resulted in a significant change in clasper bristles (Figure 4 and Supplementary Table 3), again including *tna* and *MED24* (Figure 3B, Figure 4B and Table 1). Interestingly, RNAi knockdown of *MED24* removed the longest mechanosensory bristle on both claspers (Figure 4B).

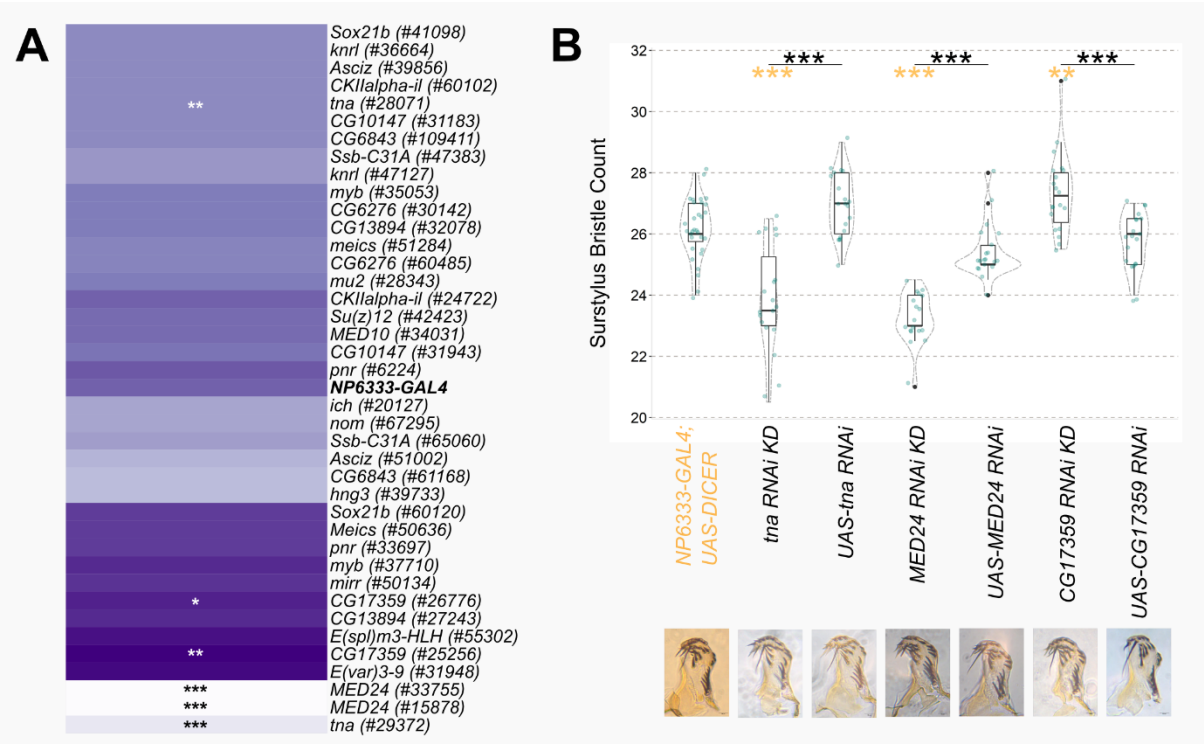


Figure 4: RNAi knockdown clasper effects. (A) Heatmap of the clasper bristle number RNAi knockdown effects compared to *NP6333-GAL4, UAS-DICER*. Those significantly different to both parental controls in the same direction are annotated with asterisks. Darker shading indicates an

increase in clasper bristle number, lighter shading indicates a decrease in clasper bristle number, both relative to *NP6333-GAL4, UAS-DICER*. (#XXXXX indicates stock centre number (Supplementary Table 3)). (B) RNAi knockdown of TFs where the clasper bristle number is significantly affected. Knockdown (KD). n > 13 for each genotype assayed. Asterisks represent the level of significance (* $p<0.05$, ** $p<0.01$, *** $p<0.001$). All RNAi knockdowns were driven by *NP6333-GAL4, UAS-DICER*.

CG17359 increased the number of clasper bristles upon RNAi knockdown, approximately by 1 (Figure 4B and Supplementary Table 3). This gene is more highly expressed in *D. mauritiana,* yet it's reduced expression in *D. melanogaster* increased the bristle count, revealing that it's species expression difference between *D. mauritiana* and *D. simulans* does not reflect the phenotypic difference in the same direction (Table 1 and Figure 4B).

3.2.3 Cercus RNAi knockdown effect genes

Five out of the 26 differentially expressed TFs tested significantly altered the cercus size (Figure 5A, Table 1 and Supplementary Table 3). *E(spl)m3-HLH* was one of these genes, as already shown to also increase posterior lobe size upon RNAi knockdown (Figure 3B and Figure 5B). Again, it's enlargement of the cercus was replicated with a different GAL4 driver, *POXN-GAL4*, and increased temperature (Supplementary Figure 1B). This suggests a general growth regulatory role for *E(spl)m3-HLH*. The RNAi knockdown of *knrl* caused the most significant increase in cercus size (Figure 5). *knrl* is known to be involved in hindgut endoreplication domains required for DNA ploidy and in turn, an increase in cell size (Fuß *et al.*, 2001). Anatomically, the hindgut is associated with the male terminalia, also evident by the need for both to rotate during metamorphosis (Gorfinkiel, Sánchez and Guerrero, 1999; Inaki, Liu and Matsuno, 2016). Although the driver used to generate this RNAi knockdown is imaginal disc specific in expression, previous studies have shown a contrasting relationship between the hindgut and cercus size (Gorfinkiel, Sánchez and Guerrero, 1999). For example, a hypomorphic allele of *Dll* is known to completely ablate cerci in males, in response to an enlargement of the hindgut (Gorfinkiel, Sánchez and Guerrero, 1999). With this in mind, *knrl* could be causing a size specific effect on the hindgut, indirectly from its effect on the cerci.

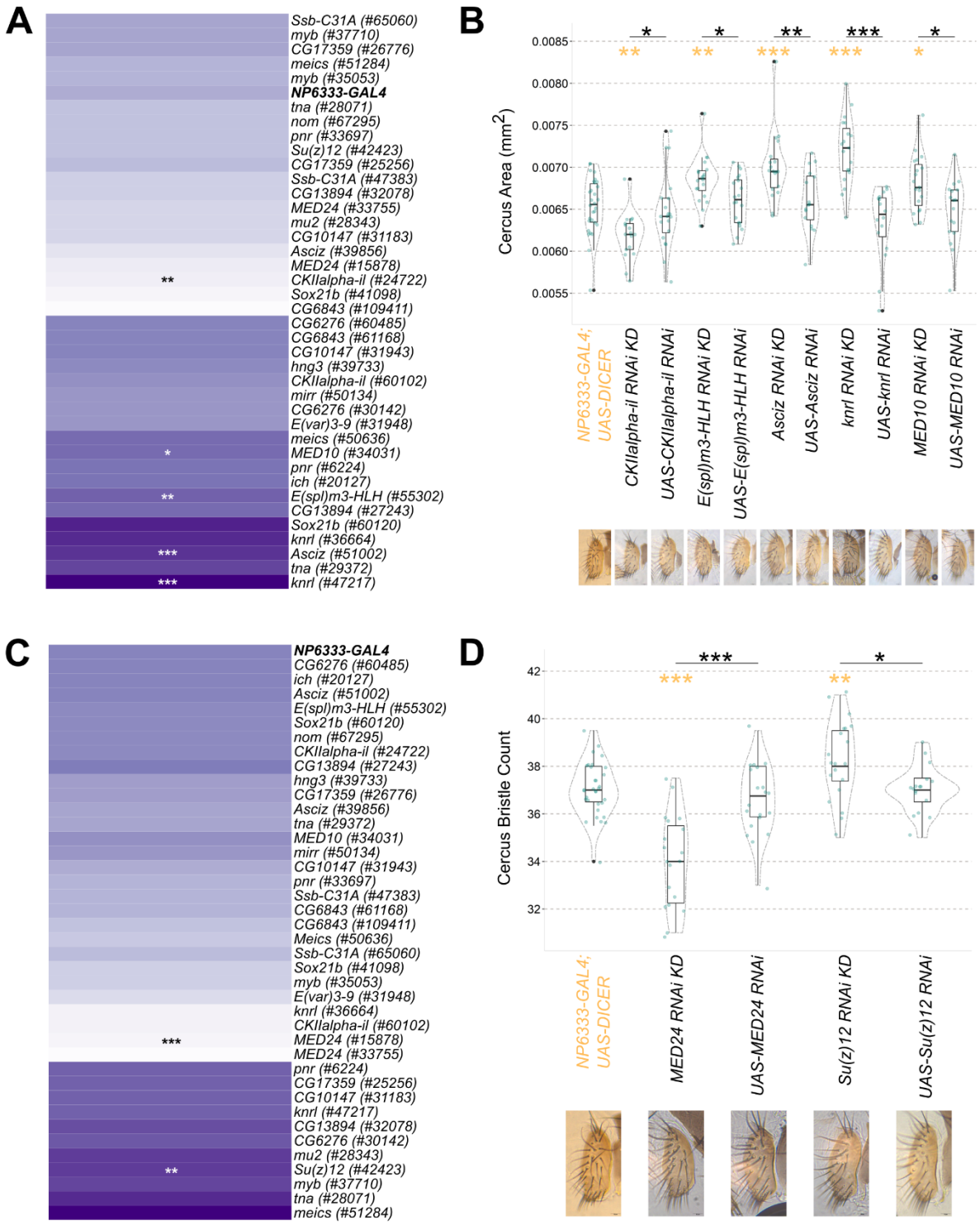


Figure 5: RNAi knockdown cercus effects. (A) Heatmap of the cercus size RNAi knockdown effects, (C) and cercus bristle RNAi knockdown effects compared to *NP6333-GAL4, UAS-DICER*. Those significantly to different to both parental controls in the same direction are annotated with asterisks. (#XXXXX indicates stock centre number (Supplementary Table 3)). (B) RNAi knockdown of TFs where the cercus size is significantly affected. (D) RNAi knockdown of TFs where the cercus bristle number is significantly affected. Knockdown (KD). (B,D) n > 13 for each genotype assayed. Asterisks represent

the level of significance (* p<0.05, ** p<0.01, *** p<0.001). All RNAi knockdowns were driven by *NP6333-GAL4, UAS-DICER*.

The bristle count of the cercus is significantly altered by just RNAi knockdown of two genes; *MED24* and *Su(z)12* (Figure 5C and D). *MED24* caused an 8% decrease in the cercus bristle count, and *Su(z)12* a 2% increase (Figure 5D). In both scenarios, the patterning of the cercus bristles does not observably differ in the knockdown compared to both parental controls (Figure 5D).

Overall, my functional analysis in *D. melanogaster* identified eleven TF-encoding genes differentially expressed between *D. simulans* and *D. mauritiana* that have a role in the development of the periphallic genitalia.

3.3 *Sox21b*-coordinated posterior lobe development and evolution

From the initial analysis, *Sox21b* was shown to cause the most significant increase in posterior lobe size upon RNAi knockdown in *D. melanogaster* (Figure 3A). In addition, it is differentially expressed between *D. mauritiana D1* and *D. simulans*^{w501} developing male genitalia, where higher expression is present in the former (Hagen et al., 2021). Previously, introgression lines have been generated in *D. simulans* from *D. mauritiana* to map evolved regions targeting periphallic structure of the male genitalia (Hagen et al., 2021). A region denoted as 'P5' (3L: 13,393,862 – 14,532,063), on chromosomal arm 3L contained 93 genes, thirteen of which are differentially expressed between the two species, and accounts for up to 9% of the posterior lobe phenotypic effect size (Hagen *et al.*, 2020). Prior to my study, none of the thirteen differentially expressed genes in P5 had been functionally tested for during male genital development (Hagen et al., 2021). From the three I tested for above, two (*meics* and *CG17359*), showed no significant phenotype upon RNAi knockdown (Table 1). *Sox21b* therefore was the only differentially expressed TF mapped to this evolved region to cause a significant phenotype to the posterior lobe upon RNAi knockdown (Table 1 and Figure 3A). In this section, I set out to understand the developmental role this TF plays in *D. melanogaster* male genital development, as well as investigating whether it has evolved between *D. simulans* and *D. mauritiana*.

3.3.1 RNAi knockdown of *Sox21b* in *D. melanogaster*

The posterior lobes were significantly larger when *Sox21b* expression was reduced by RNAi knockdown in *D. melanogaster*, with and without body size normalisation (Figure 6A and Supplementary Table 3). To assess this effect further, I carried out additional RNAi knockdowns using a different *Sox21b UAS-RNAi* line that is designed to target a different region of the *Sox21b* mRNA, as well as using the *POXN-GAL4*, a posterior lobe specific GAL4 driver (Figure 6A) (Boll and Noll, 2002). All the RNAi knockdowns resulted in an increase in posterior lobe size relative to controls (Figure 6A). The quantitative data is complemented by the SEM images captured of the *NP6333-GAL4* parental control versus the *Sox21b* RNAi knockdown (TRiP #60120), to emphasise the difference in posterior lobe size observed (Figure 6B-C'). I also overexpressed *Sox21b* using the same GAL4 line, but resultant progenies were not able to metamorphosise, thus unable to be phenotyped.

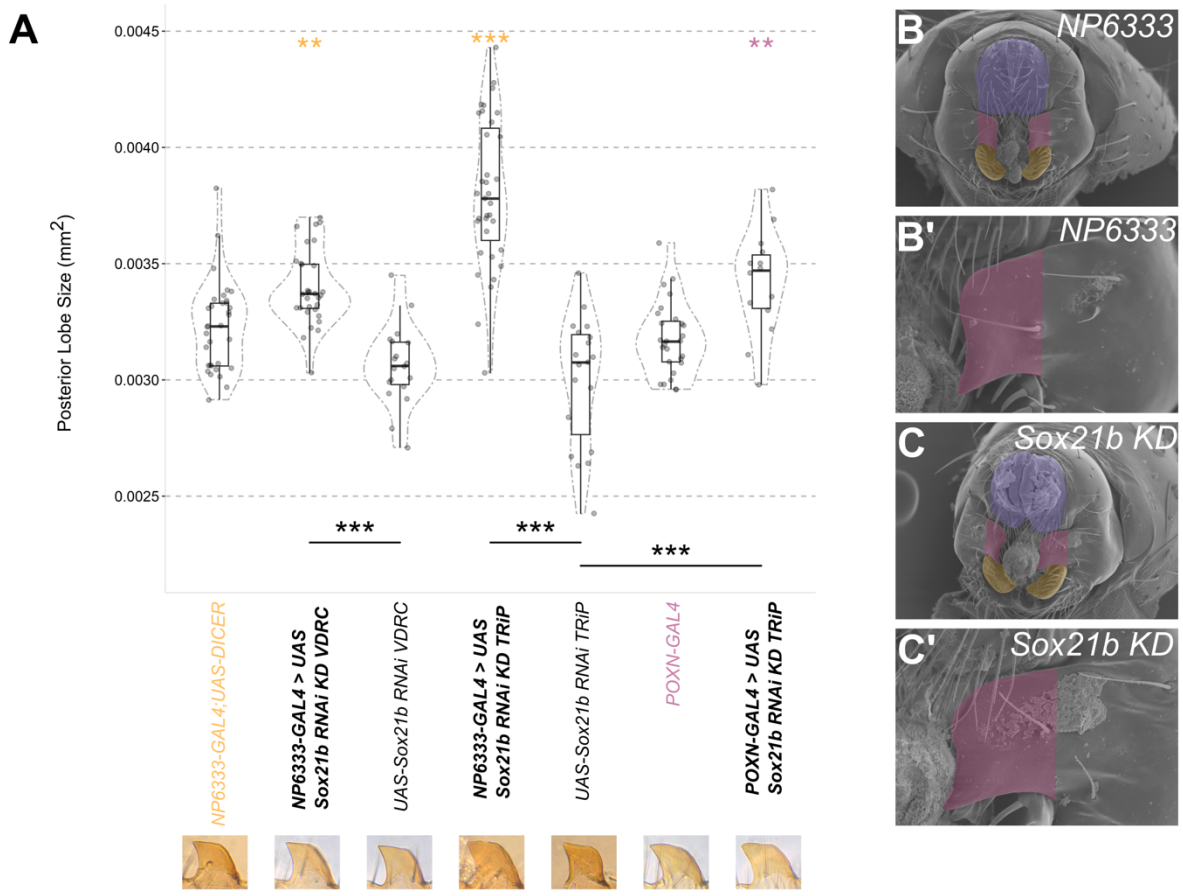


Figure 6: Posterior lobe size effects of *Sox21b* RNAi knockdown. (A) RNAi knockdown of *Sox21b* generated a significant enlargement of the posterior lobe area. Genotypes in bold represent *Sox21b* RNAi knockdown lines. KD: knockdown. *UAS-RNAi* lines represent the parental control lines carrying just the *UAS-RNAi* allele. Asterisks above the knockdown line indicates the level of statistical significance between the RNAi knockdown and associated parental GAL4 driver: orange for *NP6333-GAL4, UAS-DICER* and pink for *POXN-GAL4*. *** $p < 0.001$, and ** $p < 0.01$. $n > 13$ for each line (Supplementary Table 3). (B, B', C, C') SEM images of *NP6333-GAL4, UAS-DICER* control male genitalia (B, B') and *NP6333-GAL4>UAS-Sox21b-RNAi* 25 degrees (TRiP #60120) male genitalia (C, C'). Purple marks the cerci, pink marks the posterior lobes and yellow marks the claspers.

3.3.2 *Sox21b* RNAi knockdown posterior lobe base width and lateral plate effects

Following this, the base width of the posterior lobe was also compared between *Sox21b* RNAi knockdowns and parental controls (Figure 7A). This revealed that as well as increasing the overall posterior lobe area, reduction of *Sox21b* expression also increased the base width significantly (Figure 7A and Supplementary Table 3). As the posterior lobe protrudes from the lateral plate, I next assessed whether a change in the posterior lobe size reflects an opposing change in the size of the lateral plates. *Sox21b* RNAi led to a subtle decrease in the size of the lateral plates (Figure 7B). The reduction of the lateral plate and reciprocal enlargement of the lobe therefore reveals a potential trade-off mechanism in the proportion of cells assigned to the posterior lobe versus the lateral plate fate.

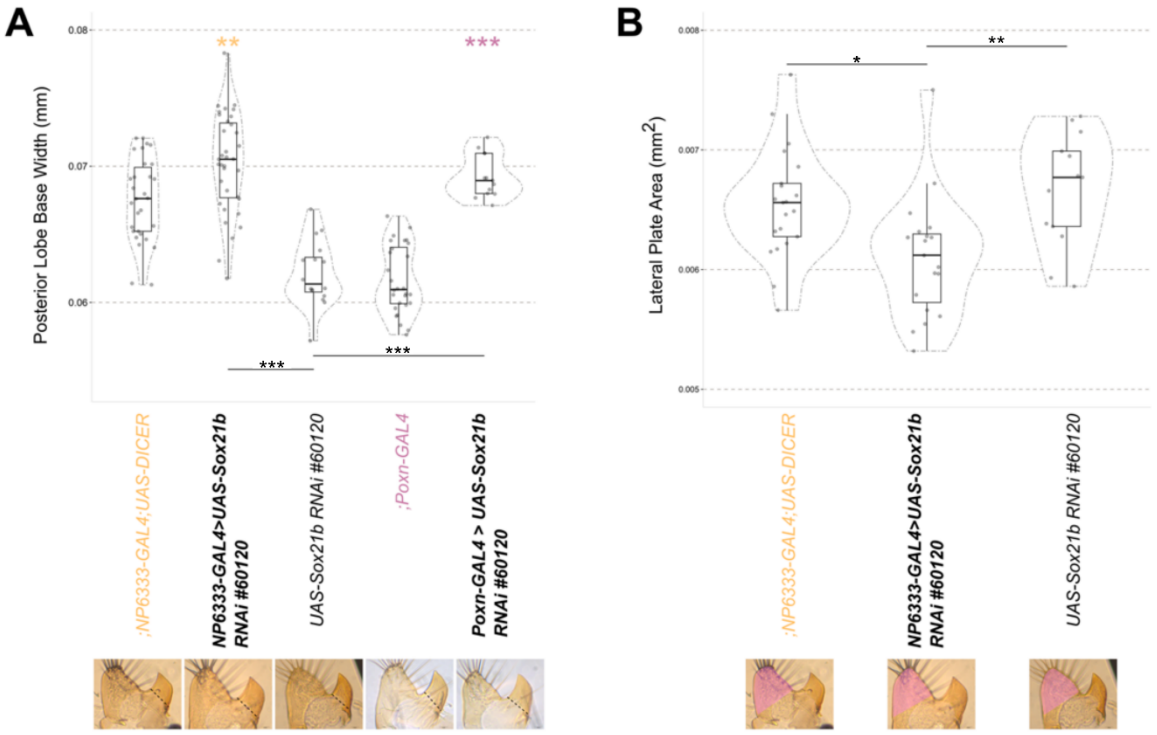


Figure 7: Effects of *Sox21b* RNAi knockdown on the posterior lobe base width and lateral plate area. (A) RNAi knockdown of *Sox21b* generated a significant increase in the posterior lobe base width. Base width is marked by a dashed line across the posterior lobe. (B) RNAi knockdown of *Sox21b* generated a subtle decrease in lateral plate area. Lateral plate area is marked by the shaded region. (A,B) Asterisks above the knockdown line marks the levels of statistical significance between the RNAi knockdown and associated parental GAL4 driver: orange for *NP6333-GAL4, UAS-DICER* and (A) pink for *POXN-GAL4*. *** $p < 0.001$, and ** $p < 0.01$. $n > 13$ for each genotype assessed (Supplementary Table 3).

3.3.3 *Sox21b* HCR-IS in *D. melanogaster*

I next sought to characterise *Sox21b*'s expression pattern in the developing male genital structures of *D. melanogaster*. Sox21b has previously been shown to act repressively during posterior lobe growth. By assessing for *Sox21b*'s expression pattern, this would provide insight to its spatiotemporal specificity required to regulate posterior lobe development. I did this by using Hybridisation chain reaction *in situs* (HCR-IS) designed to target all *Sox21b* isoform transcripts (Schwarzkopf *et al.*, 2021).

At 96 hAEL, *Sox21b* expression stretches across the medial and lateral regions of the male genital disc in *D. melanogaster* (Figure 8A). In L3 larvae at 120 hAEL, medial disc expression is constricted to the central portion with expansion in the dorsal regions still prominent (Figure 8A'). This medial spatial expression pattern overlaps with where the posterior lobe/lateral plate/clasper primordium originates from (Figure 8B). The dorsal expression pattern then overlaps with cells allocated to phallic development (Figure 8A and A'). I also analysed *Sox21b* expression in the female genital discs and found that this gene is expressed in the ovipositor and cercus primordia (Supplementary Figure 2A) (Estrada, Casares and Sánchez-Herrero, 2003).

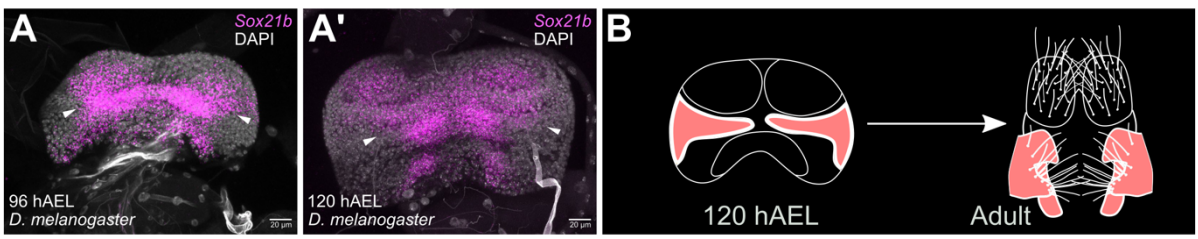


Figure 8: HCR-IS of *Sox21b* in the male genital disc of *D. melanogaster*. Expression is localised to the peripheral primordium specifying the clasper, posterior lobe, and lateral plate, and the phallic region during larval development. (A,A') White arrows mark the posterior lobe primordia in the male genital

disc. Both discs have been DAPI stained. (A) n=5, (A') n=7. (B) Schematic showing the origin of lateral plate, posterior lobe, and clasper primordium in the male genital disc at 120 hAEL in *D. melanogaster*.

Previous analysis of *Sox21b* in the developing male terminalia of *D. melanogaster* revealed expression in the developing posterior lobes and lateral plates during pupal stages (Vincent *et al.*, 2019). However, upon analysis of HCR-IS in pupal *D. melanogaster*, I did not observe expression in periphallic structures, although there was signal captured below the hypandrium (Supplementary Figure 2B and B'). I have shown that *Sox21b* is only expressed in the posterior lobe primordium of *D. melanogaster* genital discs and not during later pupal stages, suggesting an earlier than expected developmental role for this TF.

3.3.3.1 Drop in the posterior lobe

To understand *Sox21b*'s posterior lobe developmental role in the wider context, I then compared its spatial expression pattern with *Drop*, a TF previously shown to also regulate posterior lobe development (Chatterjee *et al.*, 2011). RNAi knockdown of *Drop* has been shown to severely reduce posterior lobe and lateral plate size (Chatterjee *et al.*, 2011). Anti-Drop staining in the late L3 larval male genital discs revealed expression localised to the medial region of the tissue, like that seen for *Sox21b* (Figure 8A' and Figure 9A). This expression pattern for Drop, unlike *Sox21b,* persists subsequently until pupal stages prior to the lateral plate–clasper split, where it localises to the boundary region (Figure 9B and B'; Supplementary Figure 2B and B').

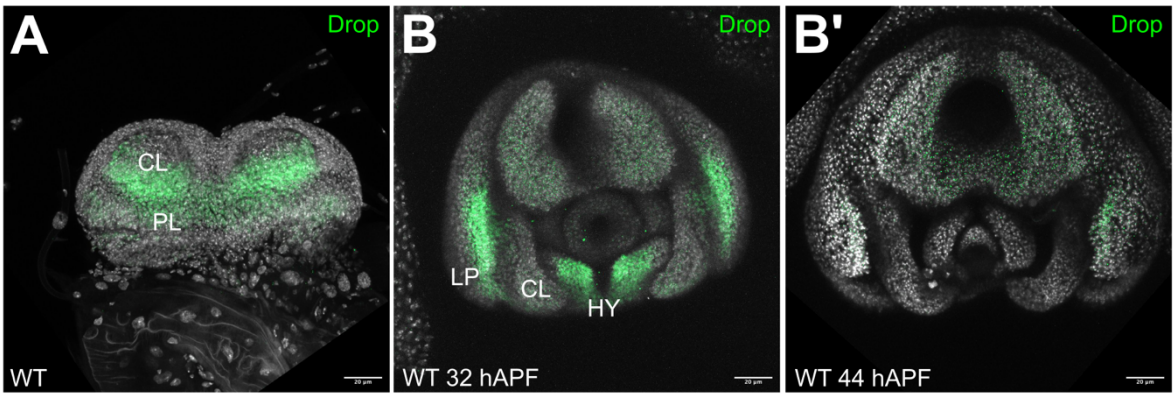


Figure 9: Drop staining in *D. melanogaster* developing male genital structures. (A) Anti-Drop staining in late larval *D. melanogaster* w^{1118} male genital disc shows that Drop is localised to phallic region and posterior lobe/clasper/lateral plate primordia. (B,B') Anti-Drop staining in developing male pupal terminalia of *D. melanogaster*. (B) Early Drop expression localised to the lateral plate–clasper boundary, and the hypandrium. (B') Drop expression is no longer observed during late pupal

development. (A) n=5, (B) n=2, (B') n=2. All samples were stained with DAPI. PL: Posterior lobe, CL: Clasper, LP: Lateral plate and HY: Hypandrium.

Drop and *Sox21b* overlap in male genital disc expression (Figure 8A' and Figure 9A). Collectively, this data suggests that *Sox21b* may play a role in regulating *Drop* expression, or vice versa, in male genital discs in the posterior lobe primordia.

3.3.4 Co-option of posterior spiracle GRN in posterior lobes

Previous work has suggested that co-option of the larval posterior spiracle GRN underlies the evolution of male genital posterior lobes (Glassford *et al.*, 2015). To explore this further with respect to *Sox21b*, I performed HCR-IS against *Sox21b* in *D. melanogaster* w^{1118} embryos to investigate whether it is expressed in the posterior spiracle-specifying cells.

eya is known to be expressed around the periphery of the posterior spiracles (Bonini, Leiserson and Benzer, 1998; Glassford *et al.*, 2015). I stained for Eya using immunohistochemistry as a method to mark for where the posterior spiracles are located on the dorsal orientation of the embryo (Figure 10A and A'). *Sox21b* staining showed segmental striped patterning on the ventral side, as well as abundant expression in the hindgut (Figure 10B and B'). In comparison to Eya staining, *Sox21b* revealed no expression at the posterior spiracles (Figure 10B').

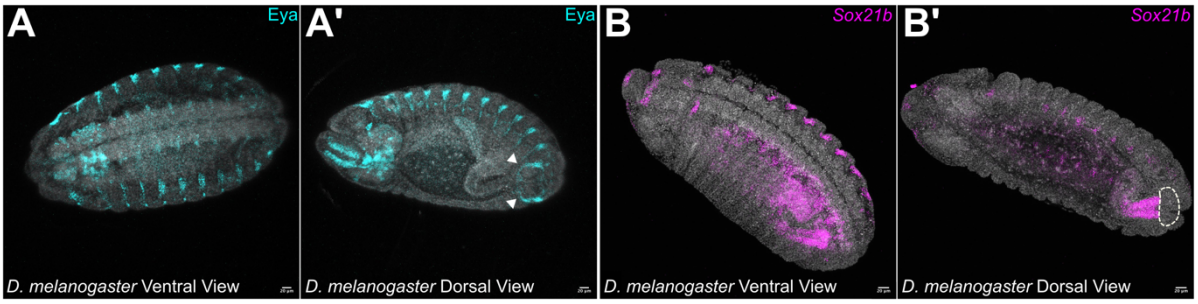


Figure 10: *Sox21b* HCR-IS and Anti-Eya staining in *D. melanogaster* embryos. *Sox21b* is not expressed in the posterior spiracles of embryos. (A,A') Eya protein is localised around the posterior spiracles (white arrows), within the head region and epidermal strips. (B,B') *Sox21b* is expressed in abdominal epidermal stripes on the ventral view, and within the mid ventral line on the dorsal view, as well as the hindgut. Embryos are orientated anterior (left) to posterior (right). All embryos are at stage 16 of development. All samples were stained with DAPI. (A,A') n=15, (B,B') n=8.

3.4 Using CRISPR to test if *Sox21b* has evolved between *D. simulans* and *D. mauritiana*

Next, I aimed to test whether *Sox21b* has evolved between *D. simulans* and *D. mauritiana* to contribute towards posterior lobe shape and size divergence. To do this, I used the introgression line previously described, that contains an introgressed *D. mauritiana* 3L region spanning *Sox21b*, within a *D. simulans* genetic background (*IL108*), and a pure *D. simulans* genetic line (Figure 11A). When compared to one another, these two lines only differ in the species identity of the 2.8 Mb region spanning *Sox21b* on chromosome 3L (Figure 11A).

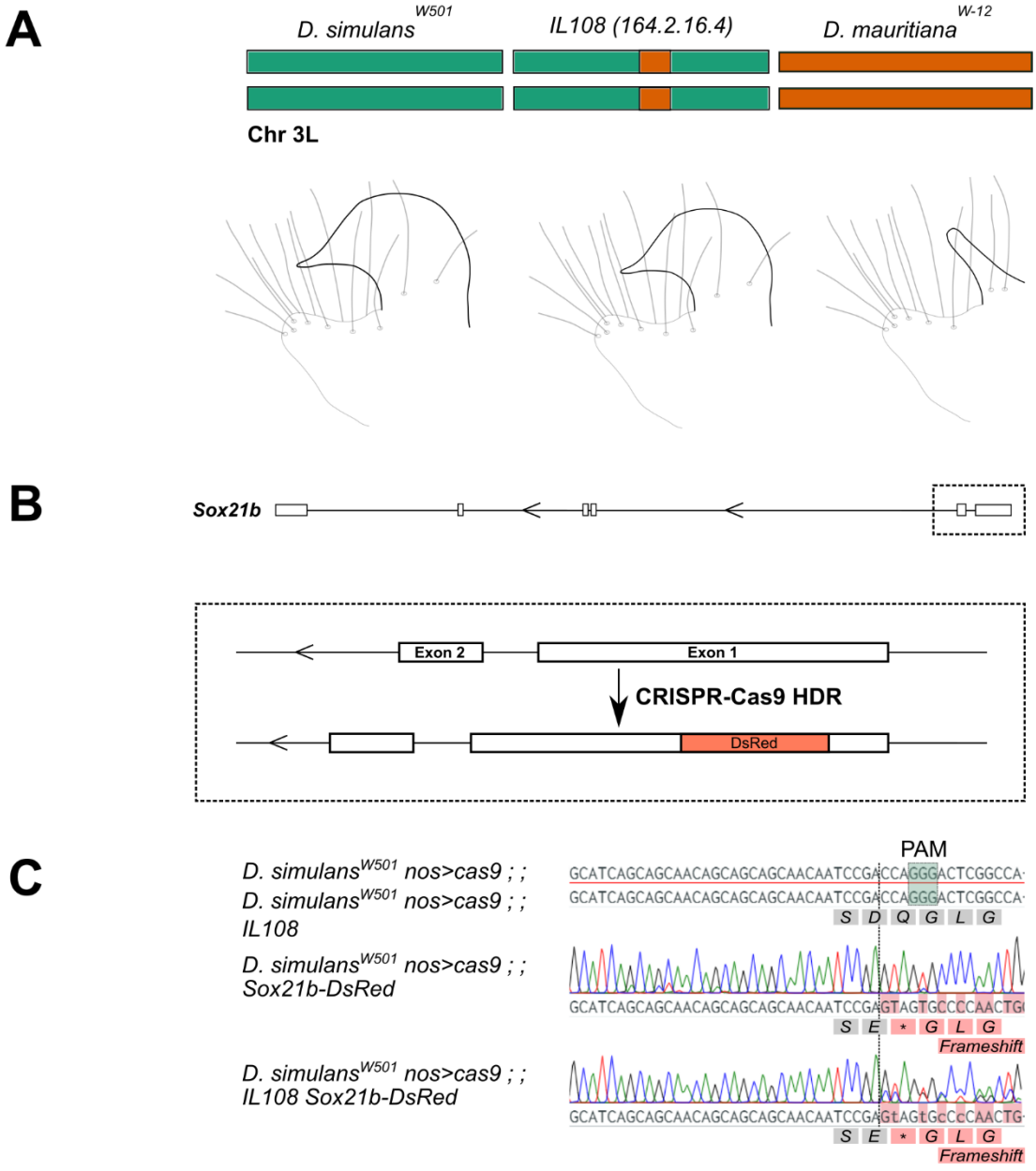


Figure 11: Generation of *; ; Sox21b-DsRed* transgenic flies in a pure *D. simulans*^{w501} *nos>cas9;;* line and in an *nos>cas9;;IL108 D. mauritiana* introgression line. (A) Schematic summarising the genetic structure of the *IL108* introgression line on chromosomal arm 3L. Posterior lobes below each genotype are simplified for size and shape comparisons. Note, *IL108* lobe size difference has been exaggerated to emphasise the reduction in size compared to pure *D. simulans* lines. The orange box represents *D. mauritiana* DNA, the green represents *D. simulans* DNA. (B) A schematic of the exon-intron structure of *Sox21b*. *3XP3-DsRed* was inserted 152 bps into exon 1, executed by homology directed repair. (C) Chromatograms illustrating the *3XP3-DsRed* insertion compared to the original parental sequence. The highlighted green box marks the PAM sequence. The dashed line indicates the CRISPR cut site, three nucleotides upstream of the PAM site. Insertion of the pHD-DsRed plasmid induced a frameshift to the open reading frame. *Dichaete* and *Sox21a* were sequenced too, to confirm no off target CRISPR effects.

These two lines underwent CRISPR-mediated homology directed repair for insertion of *3XP3-DsRed* fluorophore into exon 1 of *Sox21b* (Figure 11B) (Stern, 2014; Gratz *et al.*, 2015). This successfully disrupted the reading frame upon fluorophore insertion, creating two *Sox21b* mutant lines (Figure 11C). The *IL108* line contains a mutant form of *D. mauritiana Sox21b* allele, and the *D. simulans*^{w501} contains a mutant form of *D. simulans Sox21b* allele (Figure 11C). The reagents created here were then able to be used in the following reciprocal hemizygosity crosses, to see whether *Sox21b* has evolved between these two species.

3.4.1 Reciprocal hemizygote posterior lobe size differences

Reciprocal hemizygotes were then generated by crossing the two *Sox21b* mutant lines to generate progeny with either a working *D. mauritiana Sox21b* allele (*Dsim*$^{1.1/1.2}$) or a working *D. simulans Sox21b* allele (*IL108*$^{1.1/1.2)}$, in otherwise genetically identical backgrounds (Figure 12A). I measured the posterior lobe area and base width in the reciprocal hemizygotes to see whether the species-working allele of *Sox21b* has a significant phenotypic effect.

Remarkably, I found that posterior lobe size was significantly different depending on the species-origin of the working *Sox21b* allele (Figure 12A). The *D. mauritiana Sox21b* allele resulted in smaller lobes compared to flies with a working *D. simulans Sox21b* allele (Figure 12A). This is consistent with the direction of the difference in posterior lobe size between these two species (Figure 11A and Figure 12A).

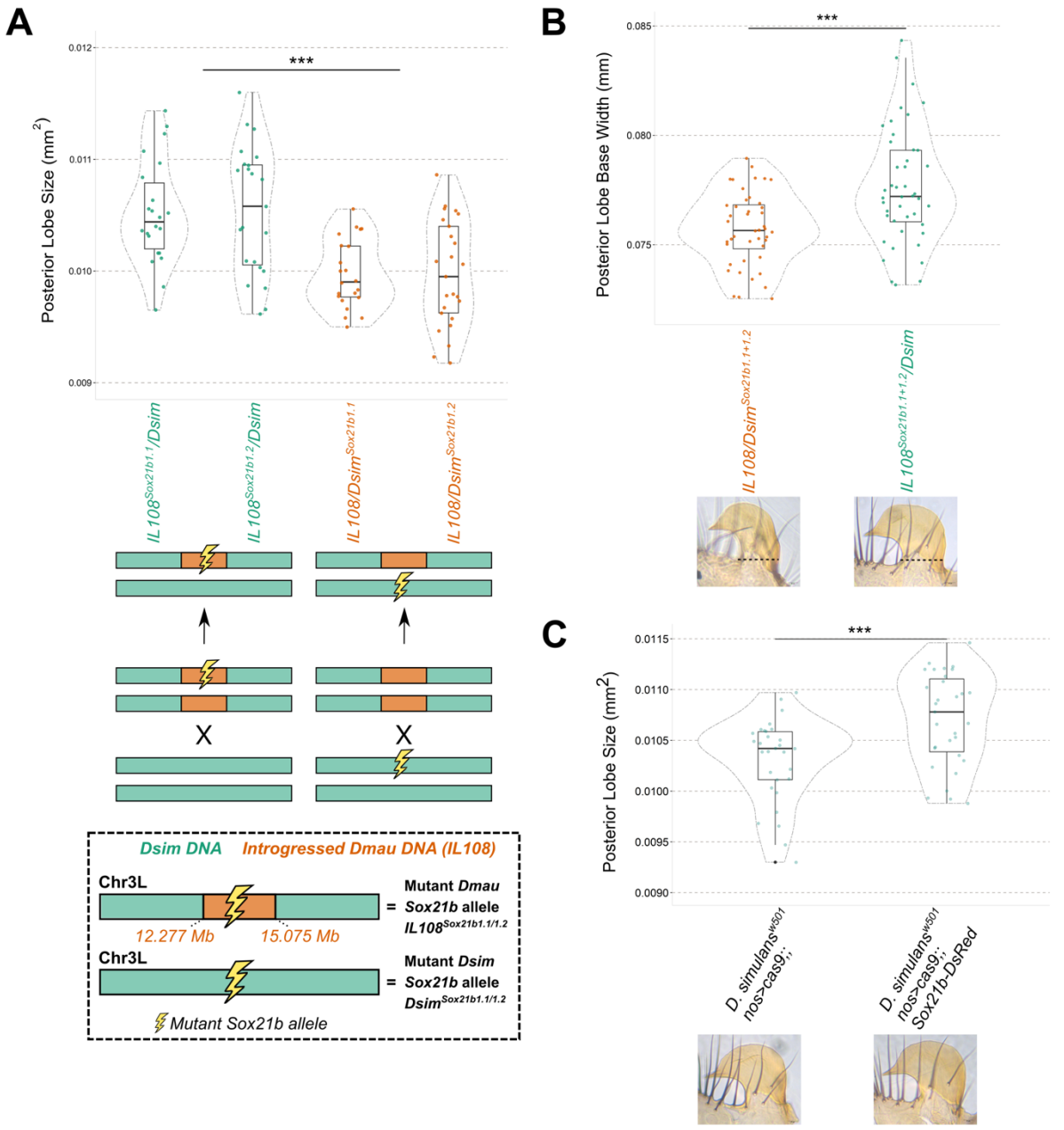


Figure 12: Effects of species-specific *Sox21b* alleles on posterior lobe size. (A) Posterior lobes are smaller in size when only the *D. mauritiana* allele of *Sox21b* is working ($Dsim^{1.1/1.2}$) (n > 22). The *D. simulans* working allele ($IL108^{1.1/1.2}$) generates comparably larger posterior lobes. The lower schematic illustrates the generation of reciprocal hemizygotes containing either a *D. simulans* working allele or a *D. mauritiana* working allele of *Sox21b*. The lightning bolt represents the disrupted *Sox21b* locus, the orange rectangle represents the *D. mauritiana* introgressed region. (B) Posterior lobe base width is significantly reduced when the working *D. mauritiana Sox21b* allele is present compared to the working *D. simulans Sox21b* allele (n > 44). The dashed line marks the posterior lobe base width measurements. (C) Homozygous *D. simulans Sox21b-DsRed* flies compared to the non-modified parental line exhibited a significant increase in posterior lobe area (n > 29). (Supplementary Table 4). *** p < 0.001.

Following this, I found that the base width of the posterior lobe was also reduced when the working *Sox21b* allele was of a *D. mauritiana* origin (Figure 12B). Clasper bristle count, cercus bristle count and cercus area did not differ significantly between the reciprocal hemizygotes (Supplementary Table 4). These results show that variation in *Sox21b* underlies the evolution of posterior lobe size between *D. mauritiana* and *D. simulans*. Interestingly, *D. simulans Sox21b* mutants were homozygous viable, and these flies had significantly larger posterior lobes than the control, consistent with *Sox21b* RNAi knockdown in *D. melanogaster* (Figure 6A and Figure 12C).

3.4.2 Reciprocal hemizygote posterior lobe shape differences

I next sought to determine how the allelic identity of *Sox21b* impacts posterior lobe shape. To do this, I used Elliptical Fourier Analysis (EFA) as a method to evaluate morphometric changes in the reciprocal hemizygote posterior lobes (Caple, Byrd and Stephan, 2017; Chan, Stevens and Todd, 2019). Twenty descriptors were assigned to each lobe and analysed using Principal Components Analysis (PCA) (Supplementary Table 5). Supplementary Figure 3A highlights those descriptors representing a significant proportion of each PC, indicated by the cos2 value. For example, descriptors 7 and 10 showcase the best representation of the shape variation captured by PC2 (Supplementary Figure 3A). Only the first five principal components had an Eigenvalue of above one and were hence selected for analysis to explore any shape alterations arising from the working *Sox21b* allelic origin (Supplementary Figure 3B). The two principal components summarising the largest proportion of shape variation, PC1 and PC2 contributing 27.1% and 12.5% respectively, were modelled against one another to compare distribution of shape differences between the two reciprocal hemizygotes (Supplementary Figure 3C). Shape variation captured by PC1 revealed a negative shift in lobes expressing the *D. mauritiana* working allele of *Sox21b*, with respect to the *D. simulans Sox21b* working version (Supplementary Figure 3C). PC2 captured a more pronounced negative shift in accordance with the variation of shape produced by the *D. mauritiana* working allele of *Sox21b*, respectively (Supplementary Figure 3C).

Figure 13A shows the distribution of reciprocal hemizygote lobe shapes with respect to PC1 and PC2, where the most extreme lobes have been used to mark the minimum and maximum axis values for each working *Sox21b* allele. When assessing the variation captured by PC2, a notable shape alteration in the ‘beak’ region of the lobe (extending from the main body of the lobe, parallel to the artificial baseline) was identified (Figure 13A). PC1 summarised 27.1% of significant shape variation between the reciprocal hemizygotes and captured shape

alterations from the artificial baseline to the beginning of the beak extension (Figure 13A). The *D. mauritiana Sox21b* allele negatively influenced the height of the 'neck' leading up to the 'beak' extension, in comparison to *D. simulans* working allele (Figure 13A).

EFA and PCA analysis of the two reciprocal hemizygotes alongside *D. simulans*^{w501} *nanos>cas9* and *D. mauritiana* D1 flies showed a distinct bimodal clustering, where, as expected, *D. mauritiana* occupies an exclusive domain of the graph (Figure 13B). Although subtle, the posterior lobe outlines derived from the flies expressing the *D. mauritiana* working *Sox21b* allele are shifted towards the pure *D. mauritiana* line along PC1 (Figure 13B). Together, this emphasises the developmental impact of *Sox21b* working species origin on the overall shape of the posterior lobe.

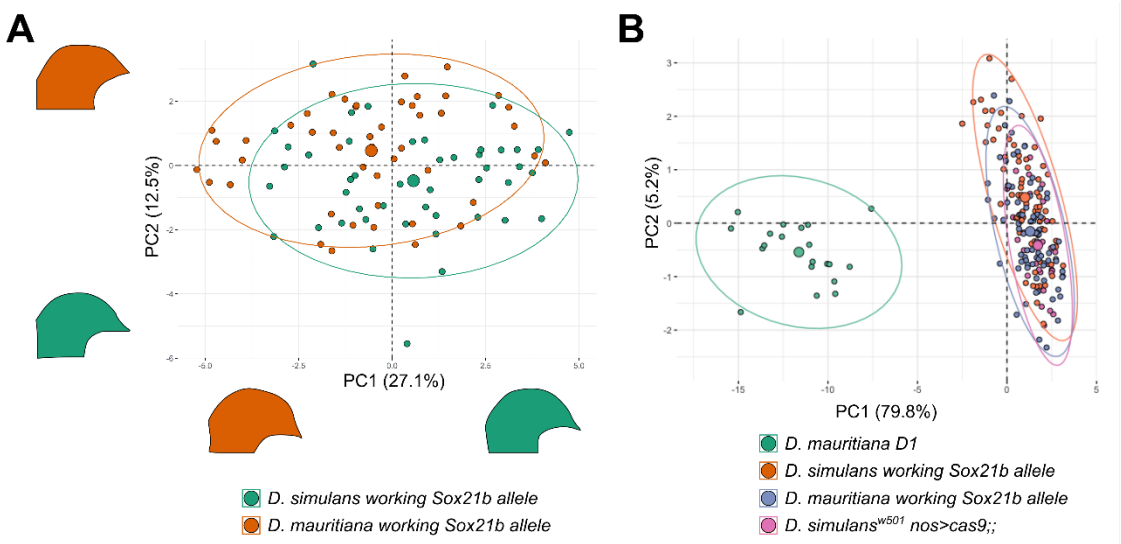


Figure 13: Effects of species-specific *Sox21b* alleles on posterior lobe shape. (A) PC1 versus PC2, whereby both statistical identifiers summarise 37.6% of total posterior lobe shape variation between the two reciprocal hemizygotes. Outlines of posterior lobes show the minimal and maximal data points for each principal component, (Supplementary Table 5). (B) PC1 versus PC2, whereby both statistical identifiers summarise 85% of total posterior lobe shape variation between the two reciprocal hemizygotes, pure *D. simulans* and *D. mauritiana* species, (Supplementary Table 6).

3.4.3 *Sox21b* HCR-IS in *D. simulans* and *D. mauritiana* male genital discs

Using a reciprocal hemizygosity test, I have shown that *Sox21b* has evolved between *D. simulans* and *D. mauritiana*, and contributes to their difference in posterior lobe shape and size. Next, I aimed to assess whether the expression of *Sox21b* differs between these two species to try to understand if changes in the regulation of this gene are responsible for its

role in posterior lobe evolution. To do this, I assayed for *Sox21b*'s expression pattern using HCR-IS in *D. simulans* and *D. mauritiana* male genital discs.

I observed that in *D. mauritiana* at 96 hAEL, *Sox21b* is expressed across the medial region of the genital disc, whereas *D. simulans* lacks *Sox21b* expression in the most lateral parts (Figure 14A and B). At 120 hAEL, the stretch of medial *Sox21b* expression contracts to varying extents between species (Figure 14C and D). *D. simulans* shows the most extreme constriction in expression, with *Sox21b* mostly absent from the lobe primordium (Figure 14C), whereas in *D. mauritiana* the broader *Sox21b* expression persists (Figure 14D). Like *D. melanogaster*, pupal terminalia of *D. simulans* and *D. mauritiana* lack *Sox21b* expression in periphallic structures (Supplementary Figure 4). To summarise, *D. mauritiana*, the species with the smaller sized posterior lobes, has sustained and more widespread *Sox21b* expression during larval development, in comparison to *D. simulans* (Figure 14).

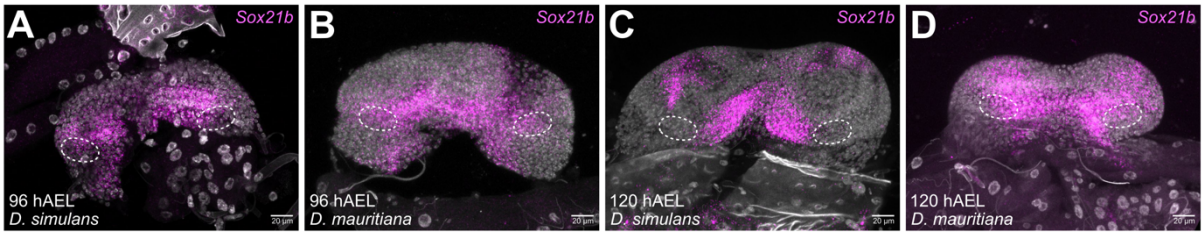


Figure 14: Dynamic developmental and evolutionary expression of *Sox21b* in *D. simulans* and *D. mauritiana*. (A,B) In both male genital discs, *Sox21b* expression is expressed laterally across the genital disc spanning the posterior lobe primordia at 96 hAEL (dashed white ovals), (C,D) and then retracts more in *D. simulans* than *D. mauritiana* at 120 hAEL, to cover the just central region. *Sox21b* HCR-IS of male genital discs at (A,B) 96 hAEL and (C,D) 120 hAEL. (A, A') *D. melanogaster* w^{1118}. (B, B') *D. simulans*^{w501}. (C, C') *D. mauritiana* D1. (A) n=3, (B) n=4, (C) n=4, (D) n=4. All samples were stained with DAPI.

3.5 Group B Sox Genes

In insects, *Sox21b* is one of four Group B Sox family genes, whereby *Sox21b* and *Dichaete* share the highest percentage similarity in protein sequence (Figure 15A and Supplementary Figure 5) (McKimmie, Woerfel and Russell, 2005). McKimmie *et al.,* previously revealed that the clustering of *Sox21a*, *Sox21b* and *Dichaete* on the same chromosome is conserved amongst insects, hypothesising that this may be in response to functional constraints (Figure 15B-D). I explored this in some species of the *D. melanogaster* subgroup and found identical arrangement of these Group B Sox Genes (Figure 15B-D). I also analysed ATAC-seq data

from the developing male terminalia of *D. simulans* and *D. mauritiana* and found similar profiles of accessible chromatin (Figure 15C and D).

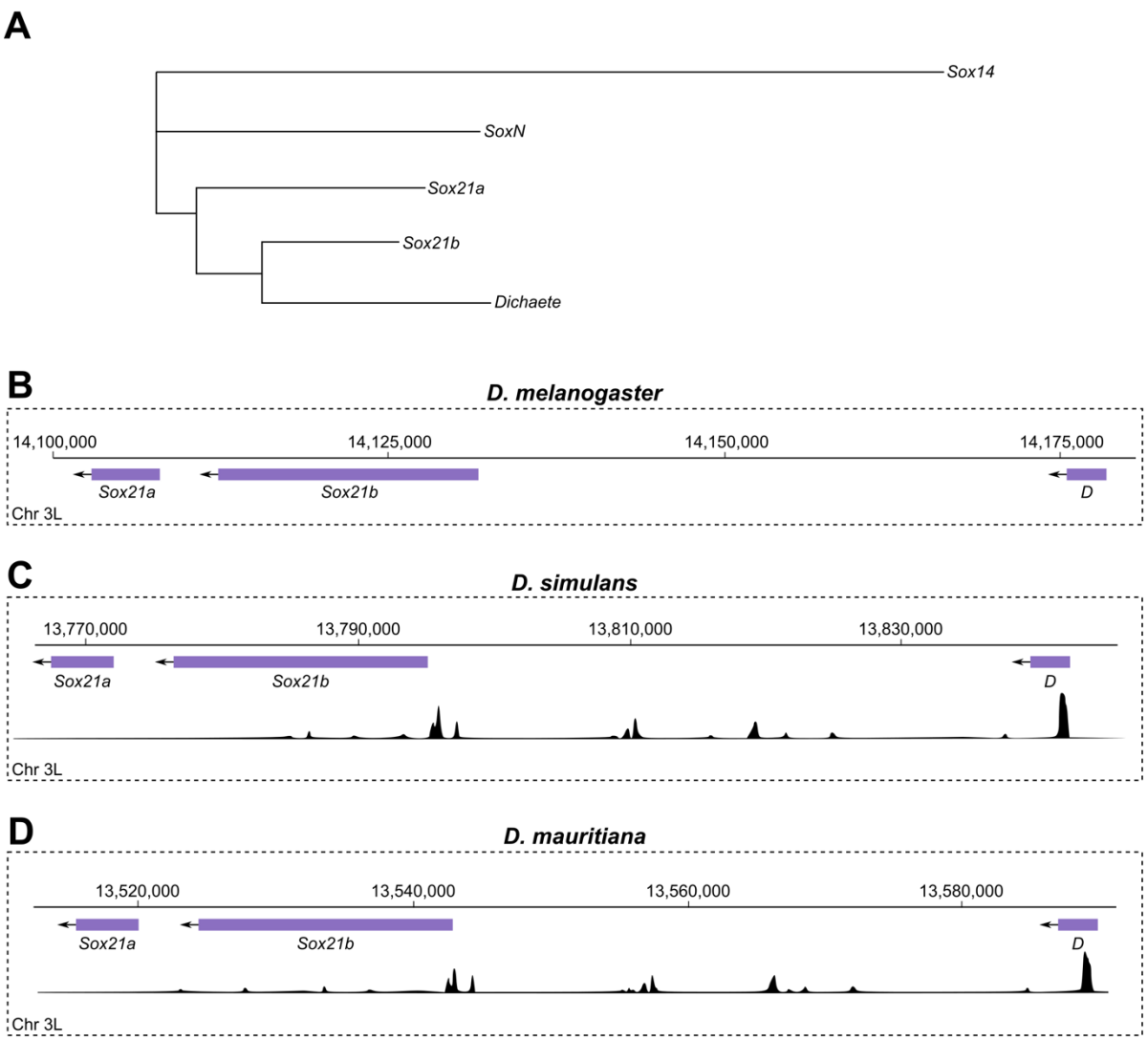


Figure 15: Species comparison of the Group B Sox locus on chromosomal arm 3L. (A) Unrooted Newick tree of the Group B Sox genes in *D. melanogaster*, with *Sox14* as the outgroup, to show ancestral similarities of genes (Supplementary Figure 5). (B-D) Gene structure schematics of the three Group B Sox genes (except for *SoxN* situated on Chr 2L). (C,D) ATAC-seq tracks generated from developing male terminalia below each gene structure diagram Hagen *et al.,* (unpublished).

3.5.1 Group B Sox genes expression levels in *D. mauritiana* and *D. simulans* developing male genitalia

I next wanted to see whether other Group B Sox genes, apart from *Sox21b*, are required for male genital development. To do so, I firstly compared expression levels of these genes using

RNA-seq data from developing *D. simulans*^{w501} and *D. mauritiana D1* male terminalia (Table 2) (Hagen et al., 2021).

Gene	*D. mauritiana D1* TPM	*D. simulans*^{w501} TPM	padj
SoxN	11.53	9.13	0.343
Sox21a	0.00	0.07	N/A
Sox21b	37.20	24.87	0.019
Dichaete	72.97	53.30	0.056

Table 2: Expression levels of Group B Sox Genes in the developing male terminalia of *D. mauritiana* and *D. simulans*. Data taken from (Hagen et al., 2021). Transcript per million (TPM) calculated as the average of three independent replicates. *Sox21b* is the only gene in this group differentially expressed between the two species. Padj: Adjusted p-value.

SoxN was expressed at low levels in both species (Table 2). *Sox21a* was shown to not be expressed in this tissue at this timepoint in both species (Table 2). Both *Dichaete* and *Sox21b* showed high levels of expression in both species developing male genitalia (Table 2). Although *Dichaete* is not differentially expressed between species, it does have higher expression in *D. mauritiana* compared to *D. simulans*, as also seen for *Sox21b* (Table 2). In addition, *Dichaete*, like *Sox21b*, is annotated as one of the top 100 expressed TFs in the developing male genitalia of *D. melanogaster* (Vincent *et al.*, 2019). Therefore, I next decided to further explore the expression and function of *Dichaete* during male genital development.

3.5.2 Dichaete protein localisation across development of the male terminalia compared to *Sox21b*

I used an anti-Dichaete antibody to visualise Dichaete protein localisation in larval male genital discs and pupal samples, and compared this to the HCR-IS expression pattern of *Sox21b*, both in *D. melanogaster* (Figure 16A and B) (Soriano and Russell, 1998). To assess the likelihood of cross-detection using these two reagents, I evaluated the target similarities of the antibody and the HCR probe. The antibody used was previously designed to target a region of Dichaete that shares just 57.4% similarity with Sox21b, considerably lower than the predicted 85% similarity threshold leading to cross-reactivity (Supplementary Figure 6). The HCR probe was designed to not target the shared HMG DNA binding domain sequence shared between *Dichaete* and *Sox21b.*

In the male genital disc, I found that Dichaete protein localisation is strikingly similar to *Sox21b*'s gene expression captured by HCR-IS (Figure 16A and B). Both Group B Sox genes

reveal expression at the posterior lobe primordium, however, Dichaete exhibits a more restricted coverage than *Sox21b* (Figure 16A and B). In pupa samples, prior to clasper and lateral plate splitting, Dichaete is expressed in the bristle precursor cells within the cerci, lateral plates, and claspers (Figure 16C and C'). At first glance, the stain looks characteristic of neuronal cell identities (Veling *et al.*, 2019).

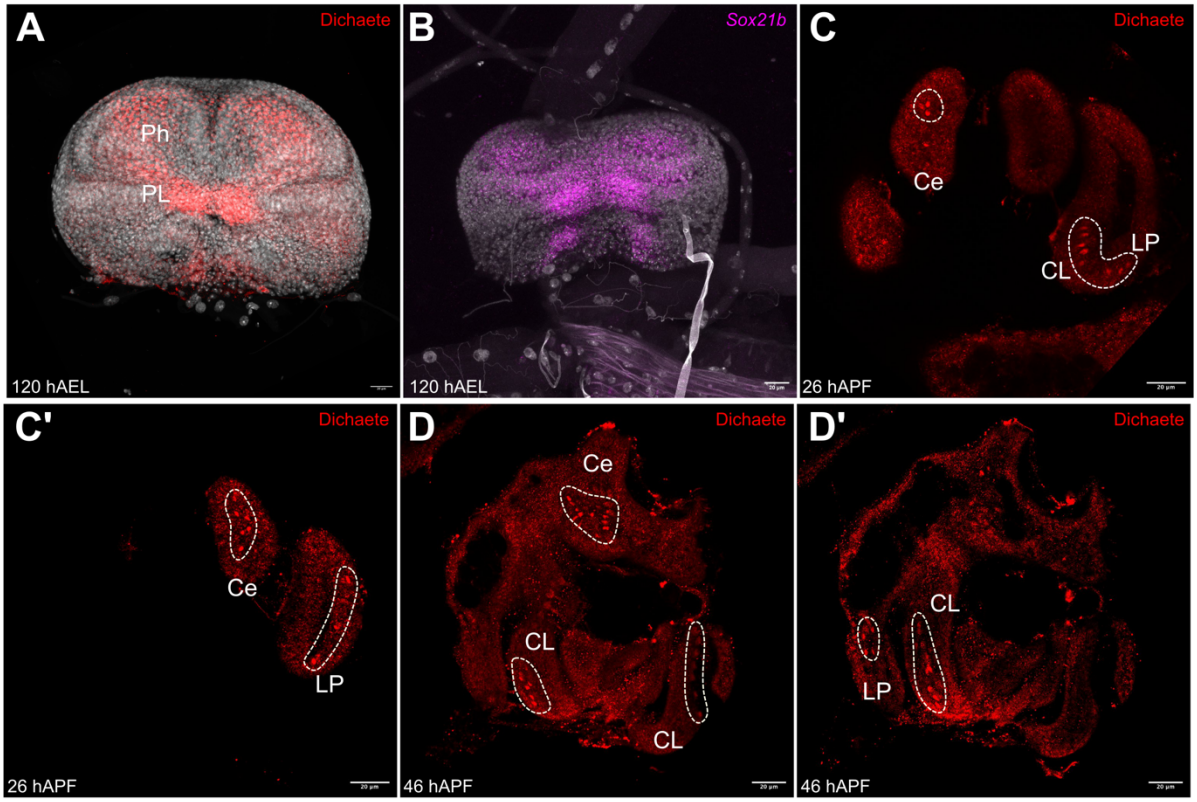


Figure 16: Dichaete antibody staining throughout male genital development. Anti-Dichaete staining compared to *Sox21b* HCR-IS in *D. melanogaster* male genital discs and developing pupal terminalia. (A) Dichaete protein localises to the same region of the genital disc as defined by (B) *Sox21b*. (C-D') Individual z-stack slices of Anti-Dichaete stains in the developing terminalia reveals individual cellular clusters of expression throughout the cerci, claspers and lateral plates. Dashed white outlines mark where Dichaete protein is localised at bristle precursor cells in the cerci, claspers and lateral plates. (A,B) All samples were stained with DAPI. (A) n=3, (C,C') n=2, (D,D') n=2. Ph: Phallus, PL: Posterior lobe, Ce: Cercus, CL: Clasper, LP: Lateral plate.

In summary, at the genital disc developmental timepoint, both *Sox21b* and Dichaete are expressed in similar spatial expression patterns (Figure 16A and B).

3.5.3 *Dichaete* RNAi knockdown and overexpression

To further understand *Dichaete*'s role in the male genitalia, RNAi and overexpression assays were then carried out. RNAi knockdown of *Dichaete* was driven using the same *NP6333-GAL4* driver as used for *Sox21b* (Figure 6A and Figure 17A). Upon reduction in *Dichaete* expression, the posterior lobe significantly increased in size compared to both parental controls (Figure 17A).

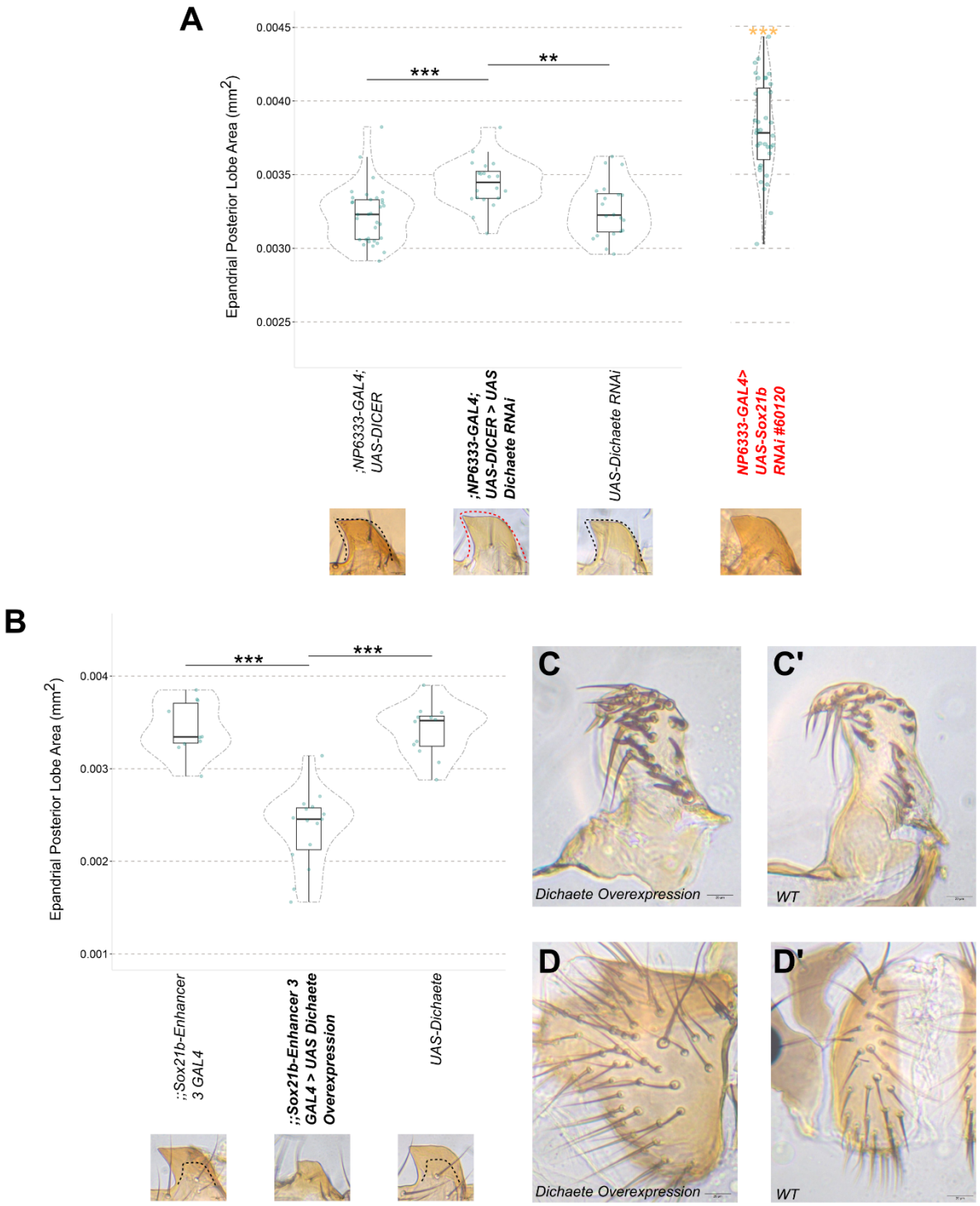


Figure 17: Changes to *Dichaete* expression levels generates significant posterior lobe size phenotypes. (A) RNAi knockdown of *Dichaete* significantly increased posterior lobe size. This was

compared to RNAi knockdown of *Sox21b*, using the same GAL4 driver. Dashed black outline represents *Dichaete* RNAi posterior lobe, dashed red outline represents *Sox21b* RNAi posterior lobe. (n > 19). (B) Overexpression of *Dichaete* during male terminalia development generates a significant decrease in posterior lobe size. Dashed black outlines represents the size of the posterior lobes produced from *Dichaete* overexpression, driven by a GAL4 driver specific to where *Sox21b* is expressed (Enhancer 3) (Figure 27), (n > 10). (C,C') Overexpression of *Dichaete* disrupts the clasper bristle patterning compared to parental *Sox21b-Enhancer 3 GAL4* line, (D,D') and cercus size and shape, (Supplementary Table 3). ** $p < 0.01$, *** $p < 0.001$.

To complement this strategy, I then used the same driver to overexpress *Dichaete* throughout the entire male terminalia, as well as all other imaginal discs in larval development. All progeny experienced early developmental arrest during pupation, and thus never eclosed to adults. This suggests that overexpression of *Dichaete* throughout all imaginal discs is lethal during pupal development, as well as being sex independent.

To combat the lethality generated by widespread *Dichaete* overexpression, a different GAL4 driver was used. Instead of *NP6333-GAL4* that is expressed in all imaginal discs from larval to pupal stages, *Sox21b-Enhancer3-GAL4* was used as it is expressed in just the genital disc (See Section 2.6 for details) (Figure 17B and Figure 27). Overexpression of *Dichaete* produced a significant decrease in the posterior lobe size compared to both parental controls (Figure 17B). In addition, it also led to ectopic bristles on the clasper and cercus, as well as severe enlargement and uncontrolled growth adjoining the two bilateral cercus structures (Figure 17C-D'). An additional observation made in these progeny showed disruption during rotation of the terminalia in early pupa. This was evident by the terminalia being not correctly positioned relative to the fly body axis.

This data has identified a repressive posterior lobe developmental role for *Dichaete*, like *Sox21b*, in the male terminalia, as well as having a developmental effect upon the claspers and cerci too. Future studies would aim to investigate whether *Dichaete* has also evolved between *D. simulans* and *D. mauritiana*, also contributing to the divergence in posterior lobes.

3.6 Chapter Three Discussion

By using a candidate gene approach and GAL4-UAS RNAi knockdowns, I identified *Sox21b* as a gene encoding a TF that represses posterior lobe growth during *D. melanogaster* male genital development. HCR staining of *Sox21b* expression patterns, reciprocal hemizygote

tests and morphometric analysis revealed that *Sox21b* has evolved between *D. simulans* and *D. mauritiana*, contributing to lobe differences observed between these two species.

3.6.1 Group B Sox genes and the regulation of posterior lobe development

In insects, *Sox21b* is one of four genes of the Group B Sox family, alongside *Dichaete, Sox21a* and *SoxN* (McKimmie, Woerfel and Russell, 2005). It is thought that *Sox21b* arose from duplication of *Dichaete* (McKimmie, Woerfel and Russell, 2005; Zhong *et al.*, 2011). These genes all contain a HMG-DNA binding domain with over 60% similarity to the mammalian *Sry* gene, required for male sex development in humans (Sinclair *et al.*, 1990). As group B Sox genes share a conserved HMG DNA binding domain, binding site selection could be redundant between these TFs, giving them similar regulatory outputs and thus phenotypic effects (Overton *et al.*, 2002; Carl and Russell, 2015). I found that both *Sox21b* and *Dichaete* repress posterior lobe growth (Figure 6 and Figure 17). *Sox21b* RNAi knockdown caused a 22% increase in lobe size, whereas *Dichaete* RNAi knockdown resulted in just a 6% increase, indicating that *Sox21b* has perhaps a more important regulatory role during lobe development. However, additional functional experiments would be needed to verify this. This suggests that these genes have some level of functional redundancy. Although *Dichaete* looks like it is required for posterior lobe development, my data suggests that only *Sox21b* has evolved. Further investigation would be required to see whether they could be working together during posterior lobe development (Bernard and Harley, 2010).

3.6.2 *Sox21b* was probably not co-opted from the ancestral posterior spiracle GRN

Previous work has proposed that the posterior spiracle GRN has been co-opted, leading to the development of the novel posterior lobe structure in male genitalia (Glassford *et al.*, 2015). I stained for *Sox21b* expression in embryos and found that it was not expressed in the developing posterior spiracles (Figure 10). Therefore, it does not look like *Sox21b* has been co-opted to posterior lobe development as part of this GRN. Partial co-option of the posterior spiracle GRN is required during male testis development for the spermiation region (Molina-Gil *et al.*, 2023). Here, *Abd-B, ct* and *ems* are not expressed, yet the part of the spiracle network required for external respiratory organ formation is (Molina-Gil *et al.*, 2023). RNA-seq data from *D. melanogaster* male testis revealed *Sox21b* expression throughout this structure, including the region where spermiation takes place (Vedelek *et al.*, 2018). This implies that integration of *Sox21b* into the co-opted spiracle network may have derived from

developmental drift (True and Haag, 2001), where absence of some ancestral genes in the posterior lobes and testes may have then enabled new genetic interactions to take place. Going forward, it would be beneficial to test the genetic similarities between the male testis and posterior lobe GRNs to assess whether *Sox21b* is required for similar interactions in the two structures.

3.6.3 Evolution of *Sox21b* between *D. simulans* and *D. mauritiana*

To understand how *Sox21b* has evolved between *D. simulans* and *D. mauritiana,* we also compared the coding sequences between the two species and found just one fixed difference (Ridgway *et al.*, 2024). Although this difference was conservative, it cannot be completely disregarded when evaluating its evolutionary impact upon Sox21b structure and function. To investigate this further, the coding change could be swapped between species using CRISPR, to assess whether it affects posterior lobe size and shape. However, it is likely that *Sox21b* has evolved from regulatory changes as inferred from expression pattern species differences during genital disc development (Figure 14). In the next chapter I describe the characterisation of a *Sox21b* enhancer.

4. Chapter Four Results

Characterisation of *Sox21b* enhancers and trans-regulatory landscape between species

4.1 Background: Enhancers and their evo-devo importance

Enhancers dictate the spatial and temporal gene expression constraints essential for body patterning, axis specification and metamorphosis (Blackman *et al.*, 1991; Affolter and Basler, 2007; Cheung *et al.*, 2013). Enhancers have been shown to act in a modular fashion, whereby multiple enhancers can act simultaneously upon one gene to drive different aspects of expression (McGregor *et al.*, 2007; Williams *et al.*, 2008; Kalay *et al.*, 2019; Kittelmann *et al.*, 2021). For example, the cis-regulatory landscape of *eve* during embryogenesis is one of the best understood cases of transcriptional regulation (Goto, Macdonald and Maniatis, 1989). The number of enhancers regulating *eve*'s 7-striped pair-rule expression pattern in embryos deviates from early to late development (Goto, Macdonald and Maniatis, 1989; Arnosti *et al.*, 1996). At the resolution of a single *eve* enhancer, activity levels have been shown to also fluctuate dynamically (Bothma *et al.*, 2014). Live imaging of the *eve* stripe 2 enhancer during embryogenesis has revealed dynamic regulatory control, termed as 'transcriptional bursts', during stripe determination (Bothma *et al.*, 2014). As well as fine tuning the rate of transcription, this study also highlighted the importance of temporal control. Interestingly, transcriptional bursts have been described as an early contributor potentially driving phenotypic variation, resulting from unequal levels of allelic transcripts (Reinius and Sandberg, 2015).

eve enhancers have been annotated at the immediate vicinity of the gene (Goto, Macdonald and Maniatis, 1989; Arnosti *et al.*, 1996; Bothma *et al.*, 2014). However, enhancers located more remotely to the target gene that show overlapping activity with proximal enhancers, have been demonstrated and defined as shadow enhancers (Hong, Hendrix and Levine, 2008). It has been shown that when proximal enhancers are deleted, phenotypic effects are suppressed in accordance with the shadow enhancer retaining endogenous expression of the target gene (Hong, Hendrix and Levine, 2008). It is important to consider the enhancer-gene distance, modularity and TF binding environment when dissecting the regulatory information of a gene during development.

Enhancers have attracted a lot of attention as the most likely regions able to underlie morphological diversity amongst species because of their modularity reducing or circumventing the pleiotropy of mutations (Rebeiz and Tsiantis, 2017). For example, *dsx* expression is coordinated by a multitude of enhancer combinations in a tissue and sex specific manner (Goldman and Arbeitman, 2007; Chatterjee *et al.*, 2011; G. R. Rice *et al.*, 2019; Gao *et al.*, 2022; Luecke, Rice and Kopp, 2022). Three of these enhancers regulate *dsx* expression at the male-specific sex combs of *D. melanogaster* (Kopp, 2011), however, *D. willistoni* versions of these enhancers showed no activity in the forelegs, thus suppressing sex comb development within this species (G. R. Rice *et al.*, 2019). Morphological diversity is therefore commonly underpinned by enhancer differences and target gene spatiotemporal changes (Gompel *et al.*, 2005; Prud'homme *et al.*, 2006; McGregor *et al.*, 2007; Carroll, 2008; Frankel *et al.*, 2011).

4.1.1 Chapter four aims

I have previously shown that the gene, *Sox21b*, has contributed to posterior lobe morphological divergence between *D. simulans*^{w501} and *D. mauritiana D1*. There is only one fixed difference in the *Sox21b* coding region between these two species and since they differ in the spatial expression pattern of this gene, this strongly suggests changes in a genital enhancer of *Sox21b* underlies the difference in posterior lobe morphology caused by this gene (Ridgway *et al.*, 2024). Therefore, I next aimed to identify the *Sox21b* genital enhancer(s) in *D. melanogaster* as outlined below:

- Screening the *Sox21b* regulatory regions in *D. melanogaster* for male genital enhancer activity.
    - Enhancer reporter lines were screened to characterise activity patterns during larval and pupal male genital development.
    - A minimal enhancer reporter line, Enhancer 3, was made and screened for its activity during male genital disc development.

- Comparing *Sox21b* genital enhancer activity between species.
    - *D. simulans* and *D. mauritiana* versions of Enhancer 3 were made as reporter lines for subsequent screening in the male genital disc.

- Mapping and testing TFBS in the *Sox21b* genital enhancer.

- TF binding sites were mapped to *Sox21b* Enhancer 3 to understand its upstream regulation.
- Abd-B regulation of *Sox21b* was tested for by mutating mapped TF binding sites and screening for enhancer activity changes.

4.2 Results: *Sox21b* enhancer screening in *D. melanogaster*

The FlyLight project generated thousands of enhancer-GAL4 reporter lines in *D. melanogaster* (Jenett *et al.*, 2012). *Sox21b* was one of their candidates, and nine of these enhancer-GAL4 driver lines were available to test (Figure 18). I defined the 69 Kb locus of focus as the sequence bordered by the two neighbouring genes, *Dichaete* and *Sox21a* (Figure 18A).

ATAC-seq is a high throughput technique used to survey the relative accessibility of chromatin throughout the genome (Yan, Powell and Curtis, 2020). At the time of writing, *D. melanogaster* ATAC-seq data for the developing male terminalia did not exist. However, *D. simulans* and *D. mauritiana* ATAC-seq data was available for analysis (Hagen *et al.,* unpublished). These data were generated from the developing terminalia at 30 – 36 hAPF. Although this is later than when *Sox21b* is expressed in the posterior lobe primordia, these data can be used to visualise regions of open chromatin that may correspond to enhancers (Figure 18B). The *D. simulans* and *D. mauritiana* accessibility profiles at the *Sox21b* locus are remarkably similar (Figure 18B). GMR73A06 and GMR73B12 spans sequences aligned with the most accessible chromatin regions in both species' pupal tissues, disregarding the transcriptional start site (Figure 18B). I would assume that both these enhancers would show genital specific expression upon screening in accordance with its chromatin accessibility.

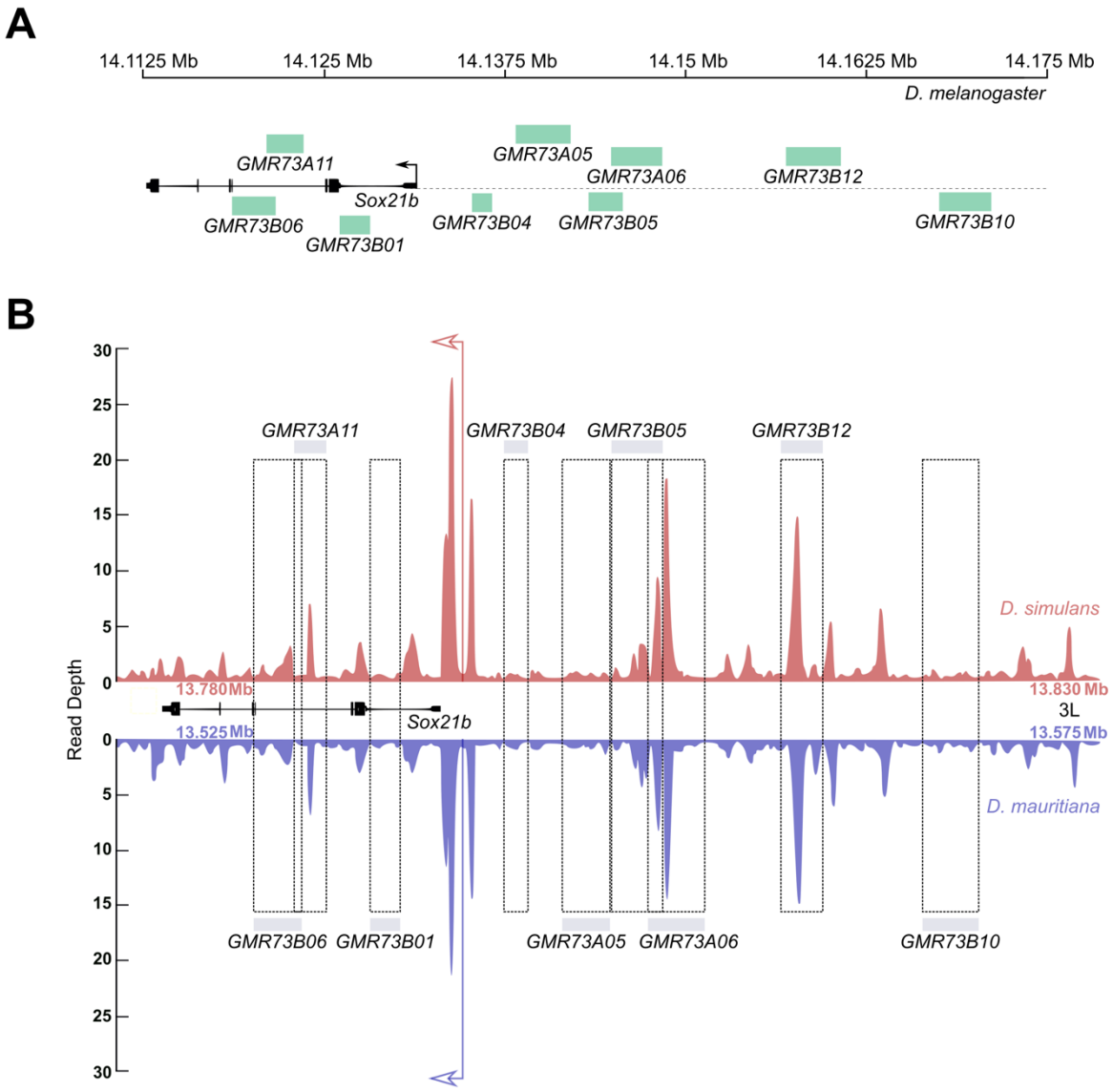


Figure 18: Janelia GAL4 enhancer driver lines tested at the *Sox21b* locus. (A) Readily available Janelia GAL4 lines spanning within and upstream of *Sox21b* mapped to *D. melanogaster* chromosomal arm 3L (R6.54 (Thurmond *et al.*, 2018)). 'GMR73XXX' is the Janelia stock name available from Bloomington *Drosophila* Stock Centre. (B) GAL4 lines mapped to *D. simulans* (top) and *D. mauritiana* (bottom) developing male genitalia ATAC-seq profiles (Hagen *et al.,* unpublished). *Sox21b* is orientated in the negative direction on chromosomal arm 3L.

4.2.1 Live pupa screening of putative *Sox21b* enhancers

To test the genital activity of the nine reporter lines, I crossed them to *UAS-GFP* and screened for expression in live pupae at 24 (early) and 48 (late) hAPF (Figure 19). From the nine possible *Sox21b* enhancers tested for, seven showed no genital expression in the pupae (Figure 19). Multiple lines revealed varying levels of abdominal expression in early pupae

(Figure 19A, B, D and F), where GMR73B06 and GMR73A11 and GMR73B04 all lacked additional enhancer activity elsewhere in the pupae (Figure 19A, B and D). GMR73B01 had activity in the thorax (Figure 19C and C') and GMR73A05 was ubiquitously expressed in dense cellular clusters throughout the pupa (Figure 19E and E'). GMR73B12 activity localised to the wing hinge during late pupal development (Figure 19F'). GMR73B10 was the only line to show activity in the ocelli of live pupae (Figure 19G and G').

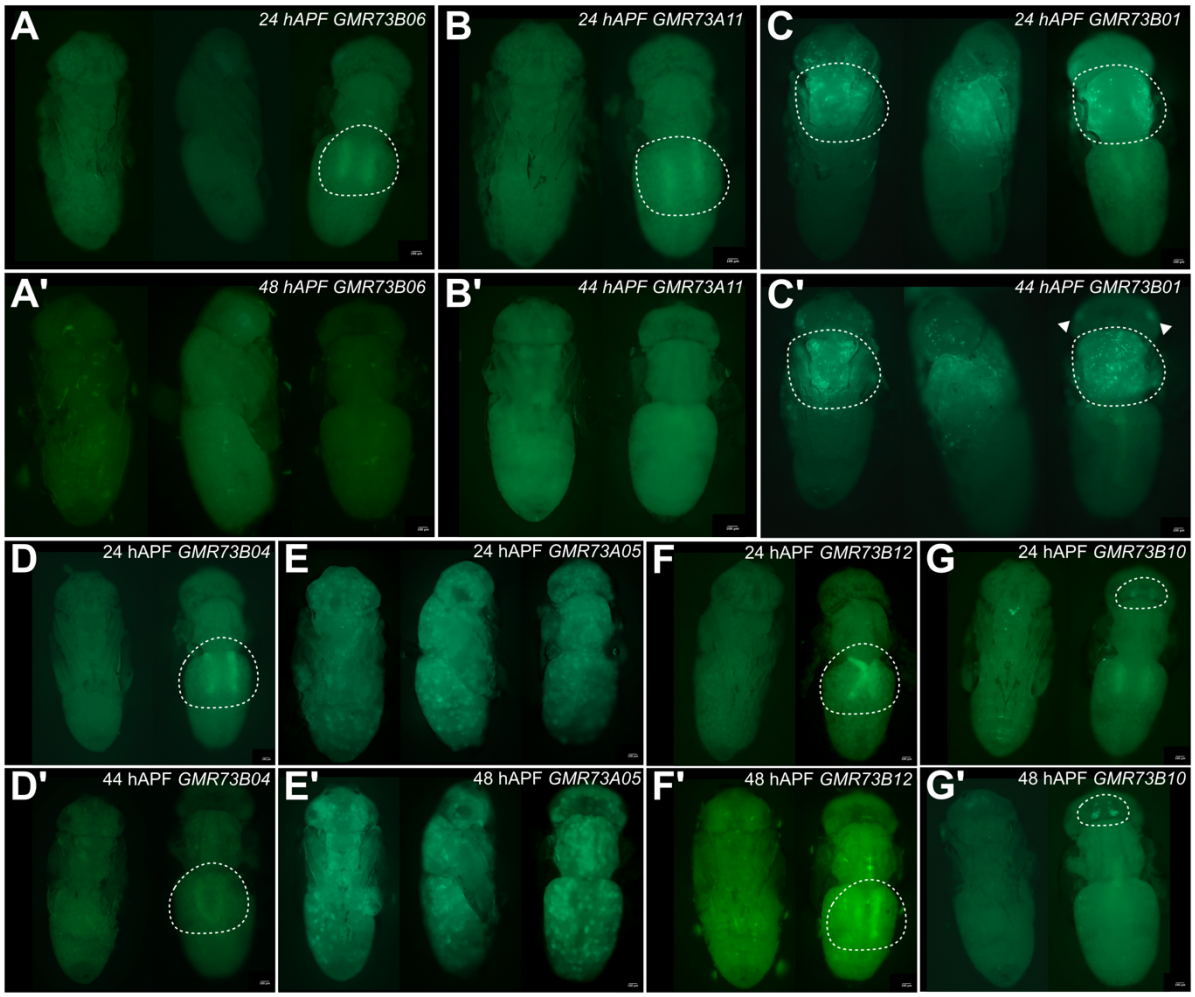


Figure 19: *Sox21b* GAL4-Enhancer driver lines that showed no visible genital activity in live pupae. GAL4 enhancer lines were crossed to *UAS-GFP* and screened for fluorescence in *D. melanogaster* live pupae. Dashed white circles outline regions of enhancer activity. All lines were imaged at early (24 hAPF) and late (48 hAPF) pupal timepoints. Scale bar is 100 µm. (C') White arrows mark enhancer activity in the eyes. (A) n=2, (A') n=1, (B) n=2, (B') n=2, (C) n=2, (C') n=3, (D) n=2, (D') n=2, (E) n=2, (E') n=1, (F) n=1, (F') n=1, (G) n=2, (G') n=2.

4.2.1.1 Pupal genital screening of putative *Sox21b* enhancers

Although no genital activity was visible for the seven regions screened in live pupae above, this approach has limited sensitivity and may include false negatives. Therefore, I next stained for anti-GFP at early and late pupal timepoints.

GMR73B04 was the only region to drive genital expression of the seven screened previously (Figure 20 and Supplementary Figure 7). This enhancer drove expression in isolated clusters at the distal head of the claspers and lateral plates (Figure 20B and C). However, *Sox21b* is not expressed in this region of pupal tissue, and one would expect the activity of this region to be repressed in its endogenous context (Supplementary Figure 2). This is supported by the lack of chromatin accessibility at this region (Figure 20A). This expression pattern is characteristic of Dichaete localisation in pupae (Figure 16C-D'). As *Dichaete* is directly upstream of *Sox21b*, GMR73B04 may coordinate expression of *Dichaete* rather than *Sox21b*.

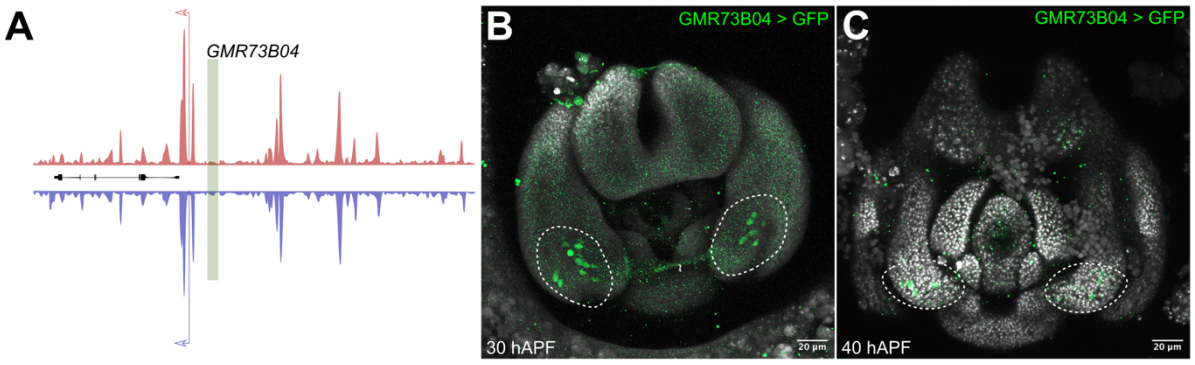


Figure 20: The GMR73B04 GAL4 line is active in isolated cells of the claspers and lateral plates. (A) DNA region directly upstream of *Sox21b* used to drive GAL4 expression. It has minimal chromatin accessibility as assayed by ATAC-seq in *D. simulans* and *D. mauritiana* (Hagen *et al.,* unpublished). (B,C) Anti-GFP staining of *GMR73B04-GAL4* driving *UAS-GFP* expression in *D. melanogaster*. White dashed ovals encircle regions of enhancer activity in the claspers and lateral plates. (B) n=2, (C) n=3. Both samples were DAPI stained.

4.2.2 Live pupa screening of GMR73B05 (Enhancer 1) and GMR73A06 (Enhancer 2)

Two (GMR73B05 and GMR73A06) of the nine showed enhancer activity in the male terminalia of live pupae (Figure 21). Both regions also mapped to accessible chromatin peaks in the ATAC-seq data (Figure 21A and C). From this point onwards, GMR73B05 will be referred to as 'Enhancer 1' and GMR73A06 as 'Enhancer 2'. Enhancer 1 revealed unique activity in the medial region of the dorsal thorax of live pupae (Figure 21B and B'), whereas Enhancer 2 was specific to the developing eyes (Figure 21D and D'). The two enhancers overlap by 811 bp,

and drive similar expression patterns in the terminalia, notum of the wings and ventral thorax (Figure 21). The region of overlap between the two enhancers may therefore delimit a minimal enhancer region for these structures.

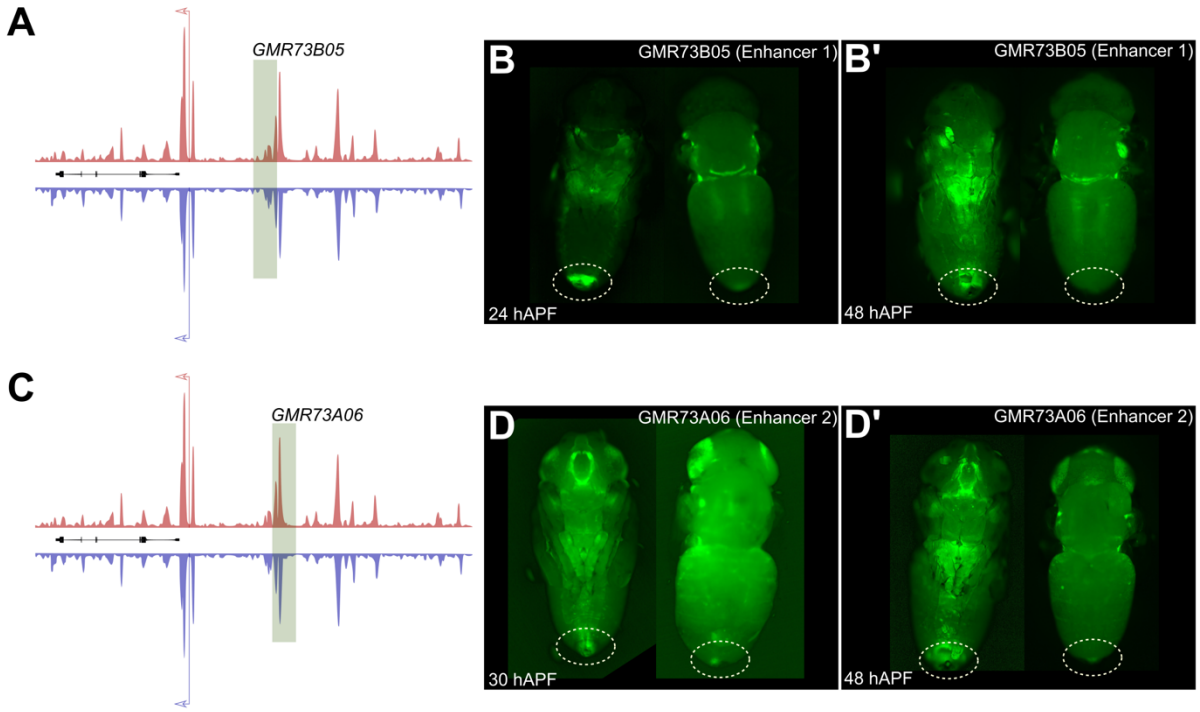


Figure 21: *Sox21b* GAL4-Enhancer driver lines with genital-specific activity in live pupa. (A,C) DNA regions upstream of *Sox21b* driving GAL4 activity is accessible for both lines in both species. (B,B',D,D') Both lines show male genital activity (dashed white ovals) in early (< 32 hAPF) and late (> 32 hAPF) *D. melanogaster* live pupae, as well as in the thorax, wings, eyes and mouthparts. (A-B') Enhancer 1 (GMR73B05) and (C-D') Enhancer 2 (GMR73A06) activity. (B) n=2, (B') n=2, (D) n=2, (D') n=2.

4.2.3 *Sox21b* Enhancer 1

Enhancer 1 spans 2.4 Kb of sequence upstream of *Sox21b* and corresponds to four peaks above a read depth of two in the ATAC-seq data with no obvious difference in the profile between *D. simulans* and *D. mauritiana* (Figure 22A). To understand the activity of this region further and in relation to endogenous *Sox21b* expression, I first stained for Enhancer 1 activity using anti-GFP in the male genital disc of late L3 larvae (Figure 22B and B'). Expression driven by this enhancer was seen to stretch across the entire medial region of the disc, overlapping with the posterior lobe primordia where I showed *Sox21b* is also expressed (Figure 8 and Figure 22B and B'). However, the more lateral Enhancer 1 activity is absent in *Sox21b* HCR-

IS images, indicating that a repressor site is potentially missing (Figure 8 and Figure 22B and B').

I then stained for Enhancer 1 driven GFP expression in fixed pupal terminalia (Figure 22C-E). Enhancer 1 activity was captured throughout the distal head of the clasper, phallic region, and peripheral to ventral regions of the cerci (Figure 22C-E). Interestingly, Enhancer 1 also drove expression in the lateral plate and extending into the adjoining posterior lobe (Figure 22C-E). This activity appears to be ectopic with respect to endogenous *Sox21b* because this gene is not expressed at this stage of development in the lateral plate or posterior lobe.

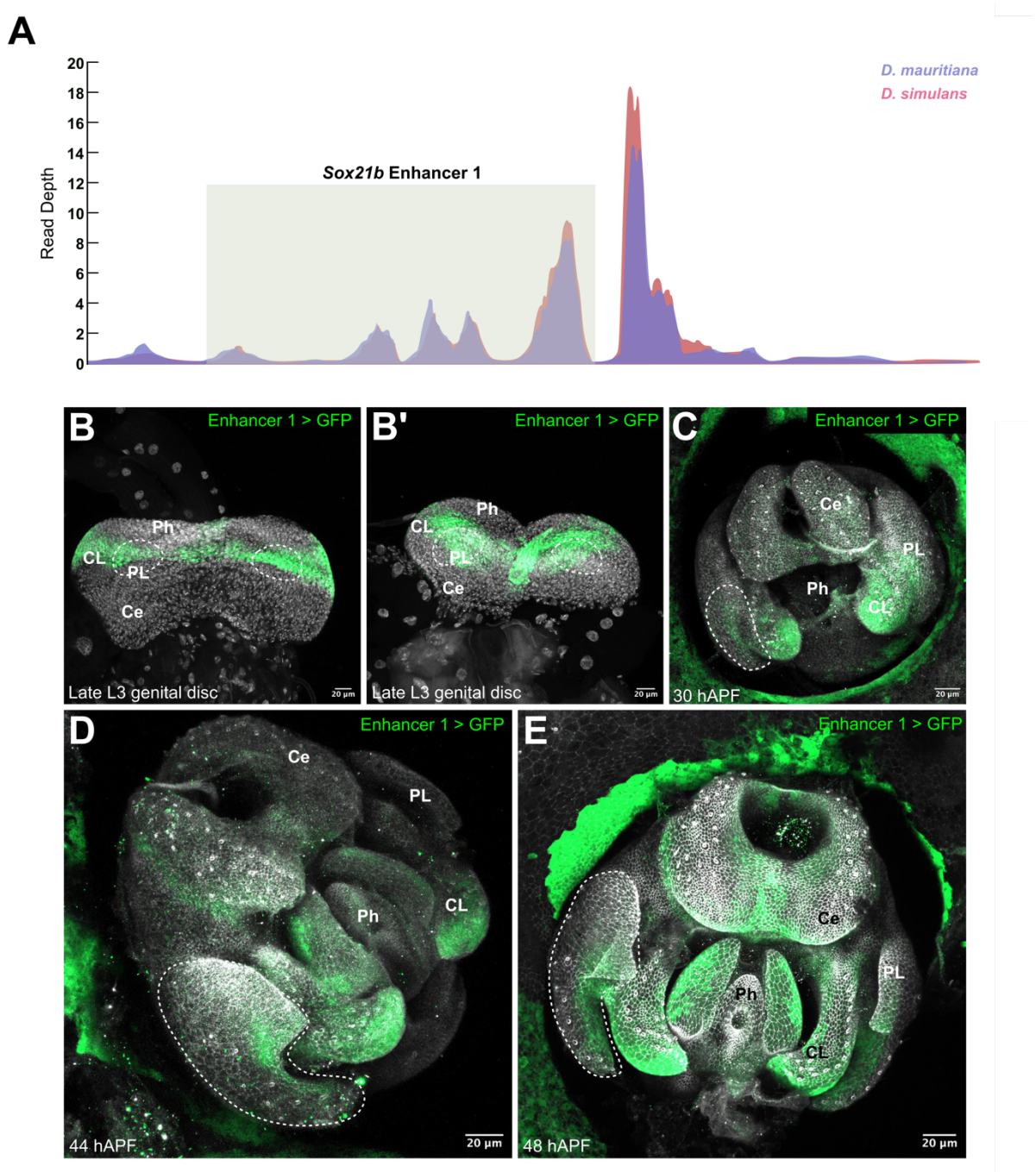


Figure 22: GMR73B05 (Enhancer 1) is active in posterior lobe specifying cells throughout *D. melanogaster* male genital development. (A) Zoomed in ATAC-seq profile highlighting Enhancer 1 accessible peaks within *D. simulans* (red) and *D. mauritiana* (blue) developing male genitalia. (B,B') Enhancer 1 activity is present at the posterior lobe primordia in late L3 male genital discs (dashed white ovals). (C-E) Enhancer 1 activity is sustained in the posterior lobe and lateral plate (dashed white outline) during pupal development. Activity can also be seen in the distal regions of the claspers, phallus and ventral cerci. (B,B') Discs are of a ventral-posterior view and have been DAPI stained, (C-E) pupal tissues have been stained against DCAD2. (B,B') n=7, (C) n=3, (D) ventrolateral view n=4, (E) n=3. CL: Clasper, Ph: Phallus, Ce: Cercus and PL: Posterior lobe.

4.2.3.1 RNAi knockdown of *Sox21b* driven by Enhancer 1 GAL4

To validate the overlap in expression between Enhancer 1 activity and *Sox21b*, I next drove RNAi knockdown of *Sox21b* using this enhancer as the GAL4 driver (Figure 23). Phenotypic analysis showed that the posterior lobe was significantly larger compared to both parental controls, consistent with previous *Sox21b* RNAi knockdown results (Figure 6 and Figure 23). This result confirms the overlap of Enhancer 1 activity with endogenous *Sox21b*.

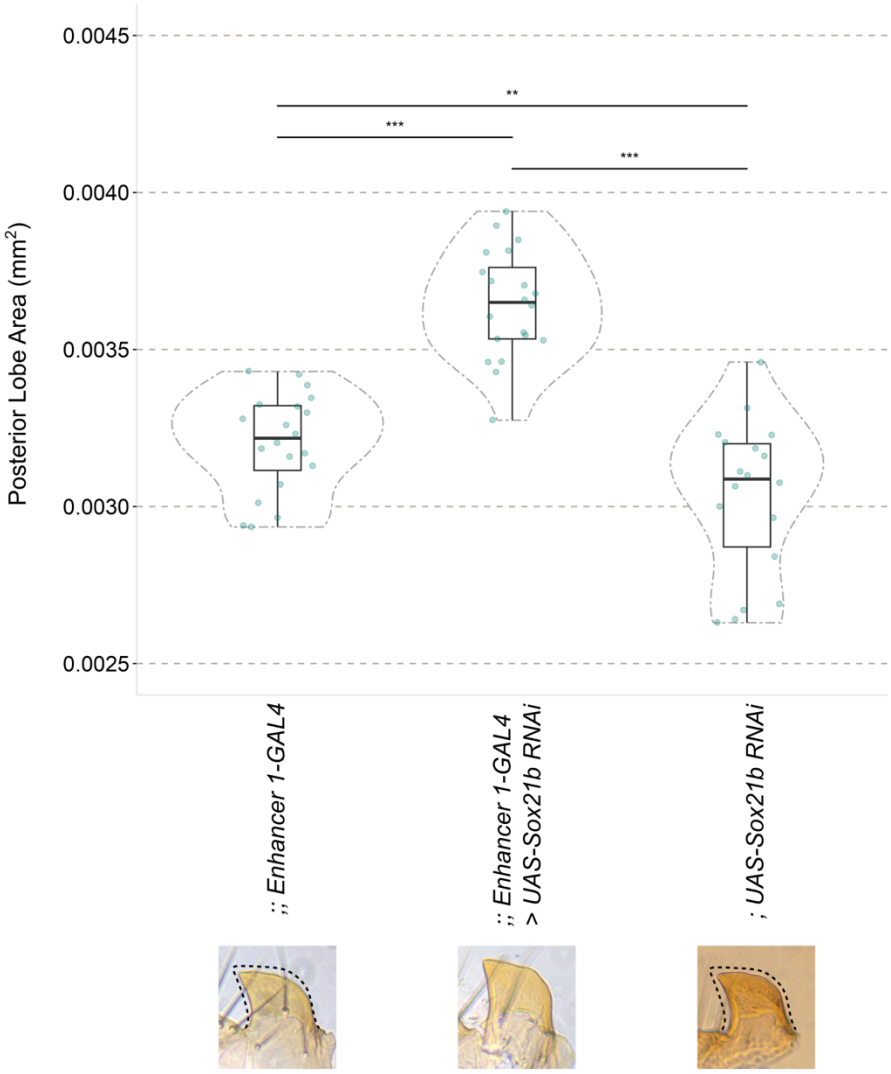


Figure 23: *Sox21b* Enhancer 1 driven *Sox21b* RNAi knockdown posterior lobe size effect. RNAi knockdown of *Sox21b* caused significant enlargement of the posterior lobe compared to both parental controls. Dashed black outline represents *Enhancer 1 GAL4 > UAS Sox21b RNAi* knockdown posterior lobes (n > 18). ** p< 0.01, *** p<0.001. (Supplementary Table 3).

4.2.4 *Sox21b* Enhancer 2

I next characterised *Sox21b* Enhancer 2 (Figure 24). This enhancer spans 3.5 kb upstream of *Sox21b* and corresponds to two peaks above a read depth of two in the ATAC-seq data, whereby the *D. simulans* profile appears more accessible than *D. mauritiana* (Figure 24A).

I stained for Enhancer 2 driven GFP in male genital discs to compare to endogenous *Sox21b* expression (Figure 24B and B'). Enhancer 2, like Enhancer 1, also showed activity in the posterior lobe primordia (Figure 22B' and Figure 24B'). Enhancer 2 activity in male pupal

terminalia appears identical to Enhancer 1, apart from the more abundant cercus activity seen in the former (Figure 22E and Figure 24E). As mentioned previously, *Sox21b* is not expressed at this stage of development in the lateral plate or posterior lobe, thus Enhancer 2 activity also appears to be ectopic with respect to endogenous *Sox21b* expression.

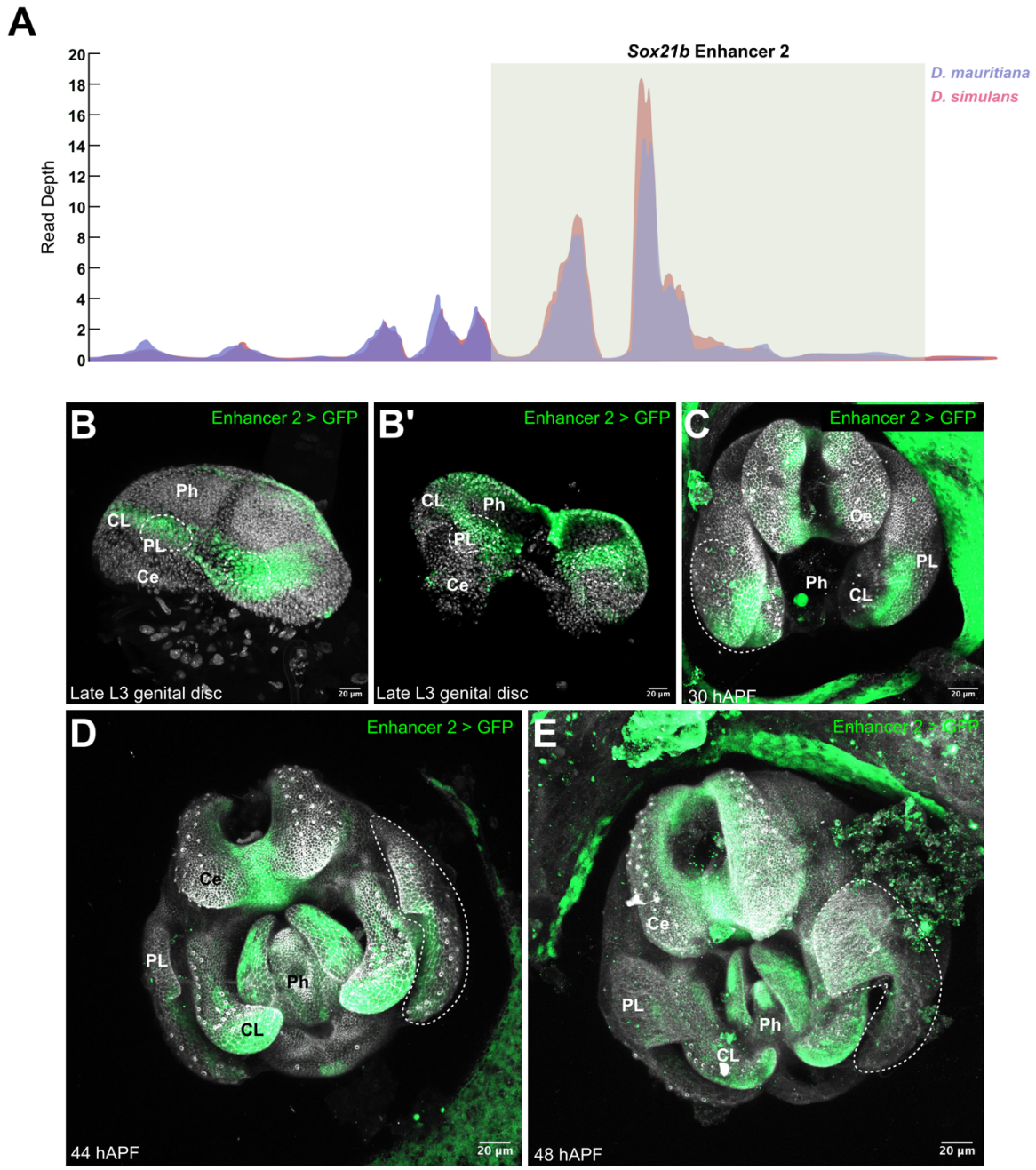


Figure 24: GMR73A06 (Enhancer 2) is active in posterior lobe primordia throughout *D. melanogaster* male genital development. (A) Zoomed in ATAC-seq profile highlighting Enhancer 2 accessible peaks within *D. simulans* (red) and *D. mauritiana* (blue) developing male genitalia. (B,B') Enhancer 2 activity is at the posterior lobe primordia in late L3 male genital discs (dashed white ovals),

as well as active at phallic and clasper primordia. (C-E) Enhancer 2 activity is also sustained in the posterior lobe and lateral plate (dashed white outline) during pupal development. Additional Enhancer 2 activity can be seen in the distal region of the clasper, phallus and along the dorsal-ventral axis of the cerci. (B,B') Discs are of a ventral view and have been DAPI stained, (C-E) and pupal terminalia has been stained against DCAD2. (B,B') n=5, (C) n=2, (D) ventral view n=2, (E) n=5. CL: Clasper, Ph: Phallus, Ce: Cercus and PL: Posterior lobe.

4.2.4.1 RNAi knockdown of *Sox21b* driven by Enhancer 2 GAL4

The Enhancer 2 GAL4 line was also then used to drive *Sox21b* RNAi knockdown (Figure 25). In this cross, the RNAi knockdown progeny showed a significant increase in T2 tibia length compared to both parentals (Supplementary Table 3). However, Pearson correlation revealed no significant relationship between lobe size and tibia length, and therefore I did not correct the posterior lobe area for leg length (Supplementary Table 3). The posterior lobes were significantly larger when *Sox21b* was knocked down using Enhancer 2 driven RNAi (Figure 23 and Figure 25).

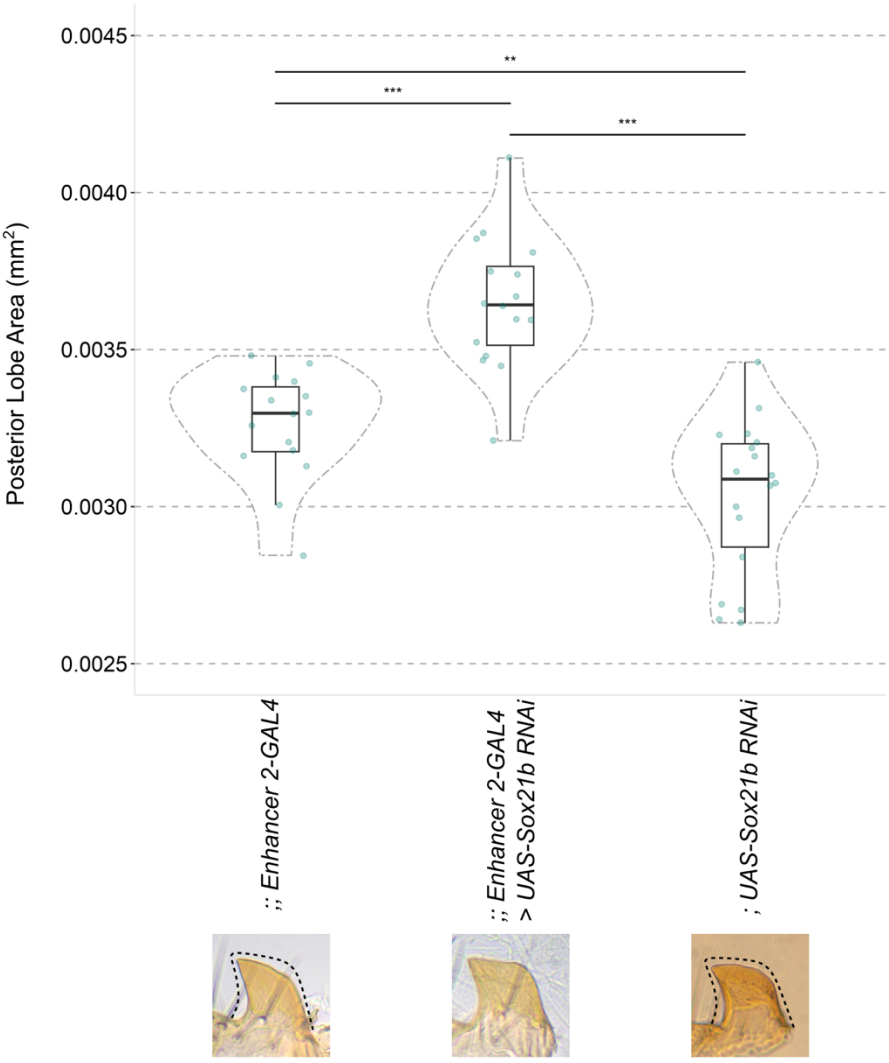


Figure 25: *Sox21b* Enhancer 2 driven *Sox21b* RNAi knockdown posterior lobe size effect. RNAi knockdown of *Sox21b* caused significant enlargement of the posterior lobe compared to both parental controls. Dashed black line represents *Sox21b* RNAi knockdown posterior lobes (n > 15). ** p< 0.01, *** p<0.001. (Supplementary Table 3).

4.3 *Sox21b* Enhancer 3: cis-regulatory sequence shared between Enhancer 1 and Enhancer 2

Enhancer 1 and 2 have been shown to drive similar spatial and temporal expression patterns of substantial similarity in the *D. melanogaster* male genital structures (Figure 22B-B' and Figure 24B-B'). Common to both enhancers is the 811 bp sequence (Figure 26). I next aimed to assess whether this 811 bp region of DNA (Enhancer 3) can capture the shared GFP driven expression patterns seen for Enhancer 1 and Enhancer 2.

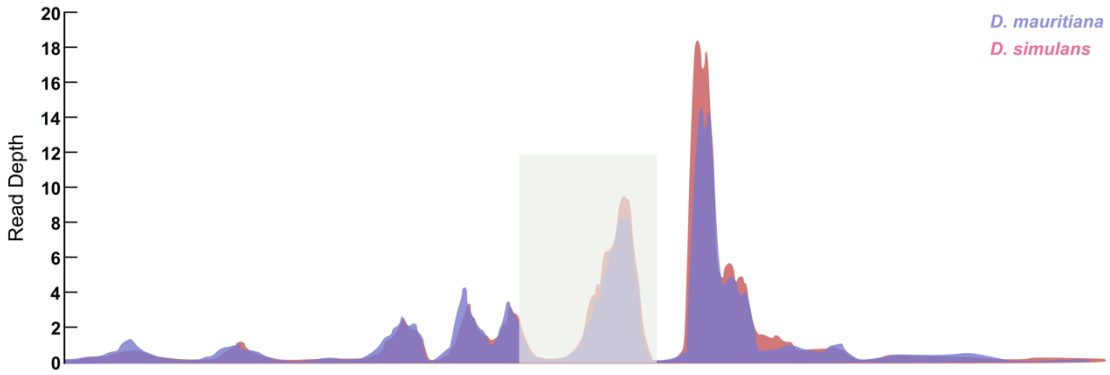


Figure 26: New *Sox21b* Enhancer (Enhancer 3) consisting of the 811 bp region shared between Enhancer 1 and Enhancer 2. Highlighted box marks region of accessibility at Enhancer 3 in *D. simulans* (red) and *D. mauritiana* (blue) developing male genitalia.

To generate the Enhancer 3 GAL4 line, the 811 bp sequence from *D. melanogaster* was inserted into the GAL4-carrying destination vector using a restriction digest approach (see Chapter 2.6). I used the same recombination site that was used for Enhancer 1 and Enhancer 2 GAL4 reporter lines to avoid positional insertion effects (Pfeiffer *et al.*, 2010).

I have shown that *Sox21b* has contributed to the evolution of posterior lobe size and shape differences between *D. simulans* and *D. mauritiana* (Figure 12 and Figure 13). Therefore, I also generated a *D. simulans* and *D. mauritiana* version of Enhancer 3 using the same approach. I defined the putative Enhancer 3 sequence by aligning and mapping Enhancer 1 and Enhancer 2 to both species and selecting the shared sequence between them. The *D. simulans* and *D. mauritiana* Enhancer 3 reporter constructs were inserted into *D. melanogaster*. Multiple sequence alignment using Clustal Omega between the three species' versions of Enhancer 3 revealed 88.9% identity between *D. simulans* and *D. melanogaster*, and 89.8% between *D. mauritiana* and *D. melanogaster* (Supplementary Figure 10). EMBOSS Needle pairwise alignment between *D. simulans* and *D. mauritiana* Enhancer 3 sequence exhibited 95.9% similarity to one another (Supplementary Figure 11).

4.3.1 *D. melanogaster Sox21b* Enhancer 3

I began by screening the *D. melanogaster* Enhancer 3 GAL4 line crossed with *UAS-GFP*, to characterise its activity in live pupae (Figure 27A). When compared to Enhancer 1 and 2 activity, Enhancer 3 recapitulated thorax, notum and peripheral abdomen driven expression

(Figure 21B and D; Figure 27A). The thorax expression was less pronounced than that seen with Enhancer 1 and 2 (Figure 21B and D; Figure 27A). Enhancer 3 also appears to be ectopically active in the pupae with respect to endogenous *Sox21b* expression (Supplementary Figure 2B-B' and Supplementary Figure 8A-A').

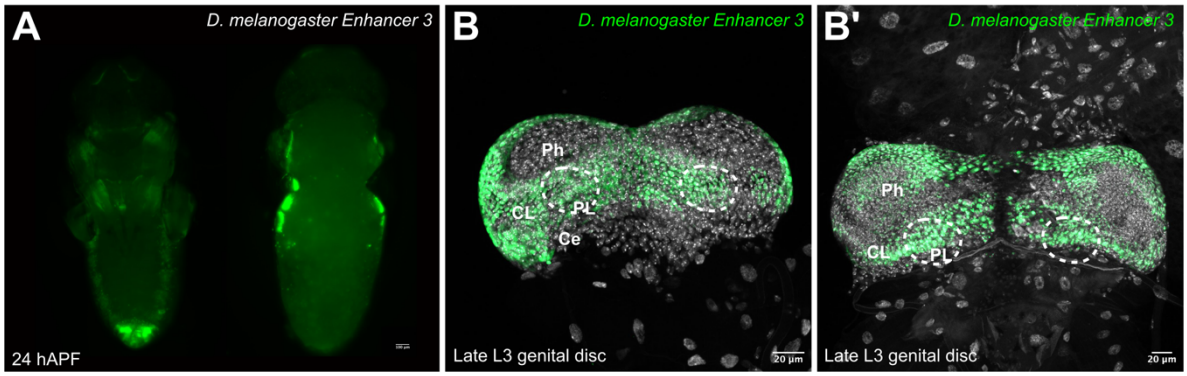


Figure 27: *D. melanogaster* Enhancer 3 activity in live pupa and fixed larval genital discs. Images generated from male samples. (A) Live pupal screening of *D. melanogaster Enhancer 3 GAL4* driving *UAS-GFP* during early pupal development (24 hAPF). (B,B') Dashed white ovals mark the posterior lobe primordia in male genital discs. Both discs are ventral genital disc views, (B') where the cercus primordium is out of view in. Discs have been DAPI stained. Expression can be seen at the posterior lobe, clasper, cercus and phallic primordia. (A) n=2, (B,B') n=5. CL: Clasper, Ph: Phallus, Ce: Cercus and PL: Posterior lobe.

I next stained for anti-GFP in *D. melanogaster* Enhancer 3 GAL4 > UAS GFP late L3 larvae (Figure 27B and B'). *D. melanogaster* Enhancer 3 driven expression resembled that of Enhancer 1 and Enhancer 2 (Figure 22B-B', Figure 24B-B' and Figure 27B-B'). Enhancer 3 also overlapped with posterior lobe specifying cells consistent with endogenous *Sox21b* expression (Figure 8A' and Figure 27B-B'). However, the lateral enhancer expression absent from endogenous *Sox21b* HCR-IS stains is still retained by Enhancer 3, like Enhancer 1 and 2 (Figure 22B-B', Figure 24B-B' and Figure 27B-B').

The ability of Enhancer 3 to drive the expression also seen by Enhancer 1 and 2 suggests that the shared 811 bp sequence is the minimal enhancer region required for male terminalia-specific *Sox21b* expression.

4.3.2 Comparison of *D. simulans* and *D. mauritiana* Enhancer 3 activity

I next assessed the spatial similarities and differences between *D. simulans* and *D. mauritiana* Enhancer 3 driven expression (Figure 28). During pupal development, *D. simulans* and *D.*

mauritiana drove expression in various tissues of pupae, including part of the thorax, the notum of the wing, and periphery of the abdomen consistent with *D. melanogaster* Enhancer 3 activity (Figure 28A and B). *D. mauritiana* Enhancer 3 produced expanded GFP-driven expression in the notum compared to *D. melanogaster* and *D. simulans* (Figure 28B). Notably, both species versions of Enhancer 3 are active in the male pupal terminalia (Figure 28A and B; Supplementary Figure 8B-C').

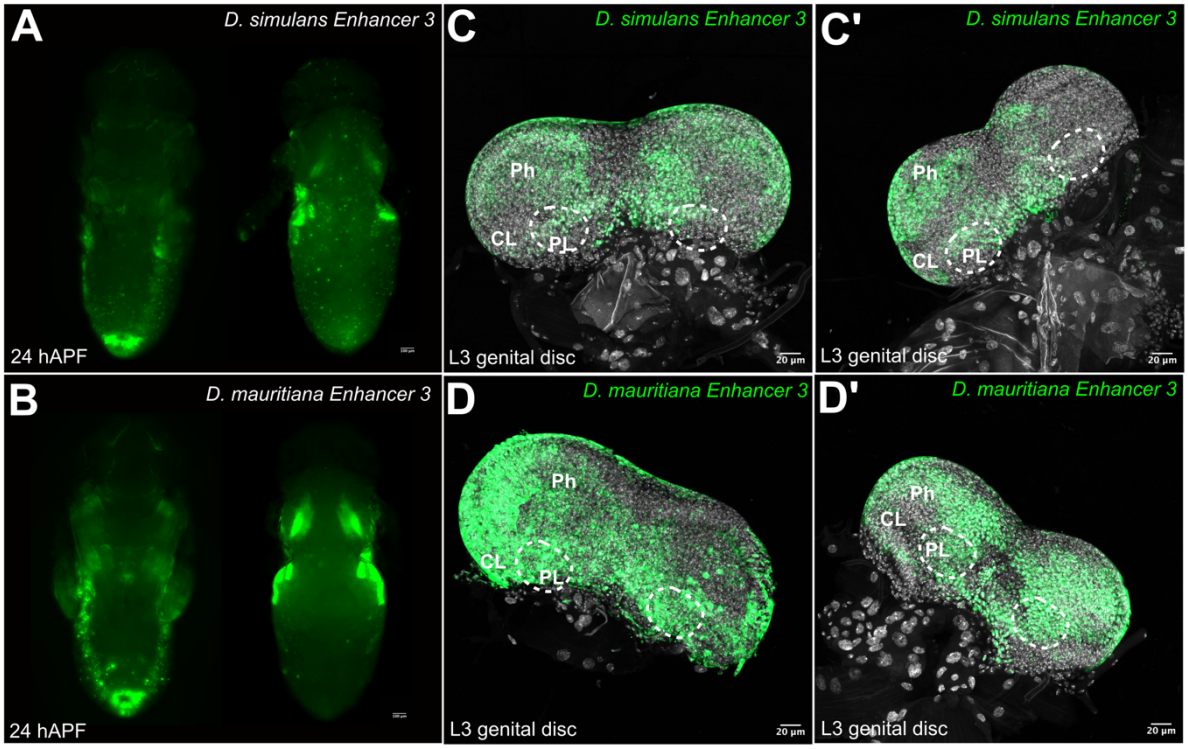


Figure 28: Comparison of *D. simulans* and *D. mauritiana* Enhancer 3 sequences driving GFP expression in live pupae and fixed larval genital discs. Both species-specific alleles for Enhancer 3 are screened in *D. melanogaster*. (A,B) Live pupae imaging of species specific versions of *Enhancer 3* driving *UAS-GFP* expression. (C-D') Male genital discs were stained against species-specific Enhancer 3 driven GFP expression. All samples were stained with DAPI. Dashed white ovals marks posterior lobe primordia in male genital discs. (C,C') *D. simulans* Enhancer 3 is shown to be active in the phallic, clasper and posterior lobe primordia. (D,D') *D. mauritiana* Enhancer 3 is shown to be active in the same primordia as (C,C') yet exhibits a stronger, more widespread signal. (C,C',D') are all ventral views and (D) is a dorsal view of the disc. (A) n=2, (B) n=4, (C,C') n=7, (D,D') n=5. CL: Clasper, Ph: Phallus and PL: Posterior lobe.

I then compared *D. simulans* and *D. mauritiana* Enhancer 3 activity in the male genital discs by again staining for anti-GFP (Figure 28C-D'). Both species versions of Enhancer 3 shows activity in the posterior lobe/lateral plate, clasper, and phallic primordia (Figure 28C-D'). In the posterior lobe primordia, *D. simulans* Enhancer 3 activity sparsely covers this region, whereas

the *D. mauritiana* version is abundant throughout (Figure 28C-D'). This difference in enhancer activity between species is consistent with the species difference in *Sox21b* expression (Figure 14 and Figure 28C-D'). Although Enhancer 3 only differs by 4.1% between *D. simulans* and *D. mauritiana*, some of these changes must result in functional divergence that could underlie the evolution of *Sox21b* expression and difference in posterior lobe size (Supplementary Figure 11).

4.4 Transcription factor binding sites in *Sox21b* Enhancer 3

I have identified Enhancer 3 as a minimal enhancer sequence capable of recapitulating *Sox21b* expression in the posterior lobe primordia of male genital discs (Figure 27B and B'). Coordination of enhancer activity depends on the binding of sequence-specific TFs (Lemon and Tjian, 2000). I therefore next sought to identify which TF(s) bind Enhancer 3 and thus regulate its activity.

4.4.1 Footprint mapping of Enhancer 3

Although the peaks presented in the ATAC-seq data can be presumed to show nucleosome-free DNA, it cannot be assumed that TFs are then subsequently bound. I used a programme called TOBIAS, which assigns 'footprints' to processed ATAC-seq data as a method of predicting TF occupancy (Bentsen *et al.*, 2020). This approach also corrects for Tn5 sequence bias that may have occurred during accessible DNA cleavage (Bentsen *et al.*, 2020).

This pipeline generated predicted footprints for *D. simulans* and *D. mauritiana* from the pupal genital tissue ATAC-seq profiles for the *Sox21b* enhancer sequences from both species (Figure 29). Interestingly, Enhancer 1 contained a peak in the ATAC-seq data, that was absent in the footprint track suggesting no TFs were predicted to be bound following Tn5 sequence bias correction (Figure 29). For both species, the Enhancer 3 accessible peak was predicted to be occupied by TFs (Figure 29).

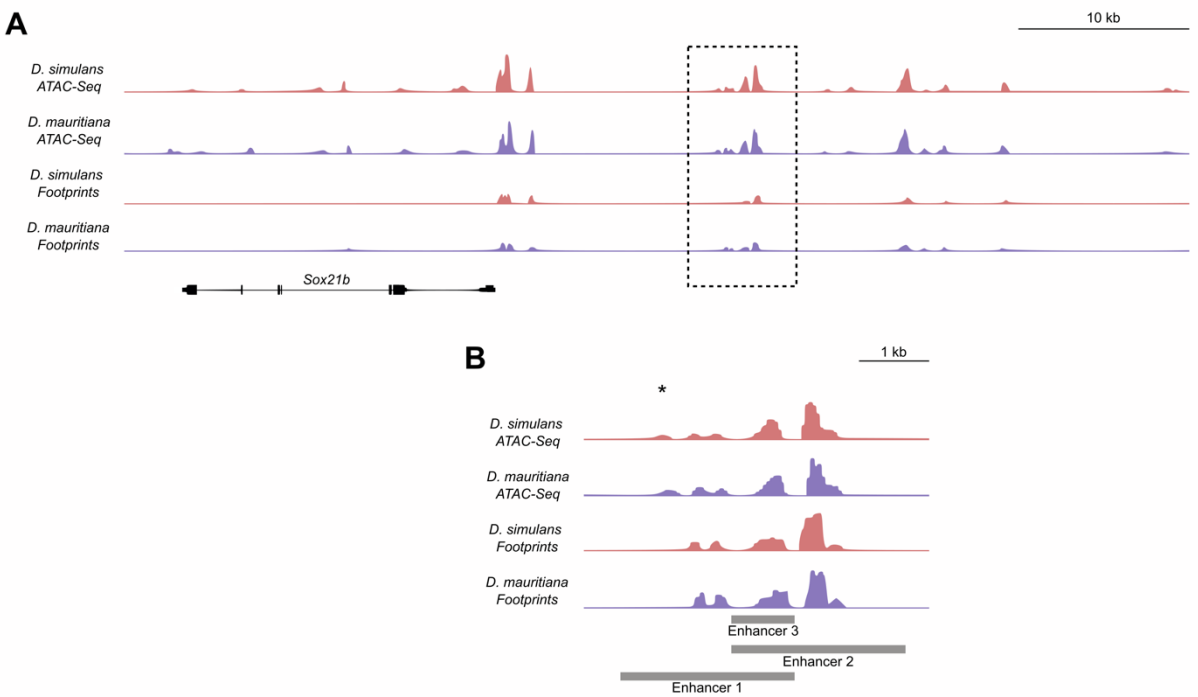


Figure 29: Footprint analysis of *Sox21b* enhancers in *D. simulans* and *D. mauritiana* ATAC-seq data. Footprint tracks represent likelihood of TF binding occupancy generated from ATAC-seq data. ATAC-seq data generated by Hagen *et al.,* (unpublished). Footprint data generated using the pipeline outlined in (Bentsen *et al.*, 2020). (A) Dashed box marks *Sox21b* Enhancers 1, 2 and 3. (B) Zoomed in profile of Enhancers 1, 2 and 3, indicating enhancer TF occupancy in relation to ATAC-seq (Footprints). * region of ATAC-seq in both species void of TF occupancy in footprint data.

4.4.2 TF motif mapping in *D. melanogaster* Enhancer 3

To identify TFs that may regulate Enhancer 3 and their binding motifs, I first filtered for TFs known to be expressed in *D. simulans* and *D. mauritiana* genital RNA-seq datasets (Hagen et al., 2021), and cross referenced this list with known posterior lobe GRN TFs expressed in a similar spatial pattern to the *Sox21b* enhancers (Glassford *et al.*, 2015). This approach identified three candidates that I subsequently used to search for binding motifs within *D. melanogaster* Enhancer 3: Abd-B, STAT92E and So.

Abd-B is expressed throughout the male genital structures, apart from the cerci (Supplementary Figure 9A). Glassford *et al.,* (2015) showed that this TF occupies a hierarchical position in the predicted posterior lobe GRN. I used the FIMO algorithm available on MEME suite to scan the Enhancer 3 sequence for possible Abd-B binding sites (Grant, Bailey and Noble, 2011). I found three Abd-B binding sites of high probability ($p < 0.001$) in Enhancer 3, two of which mapped to the predicted footprint (Figure 30).

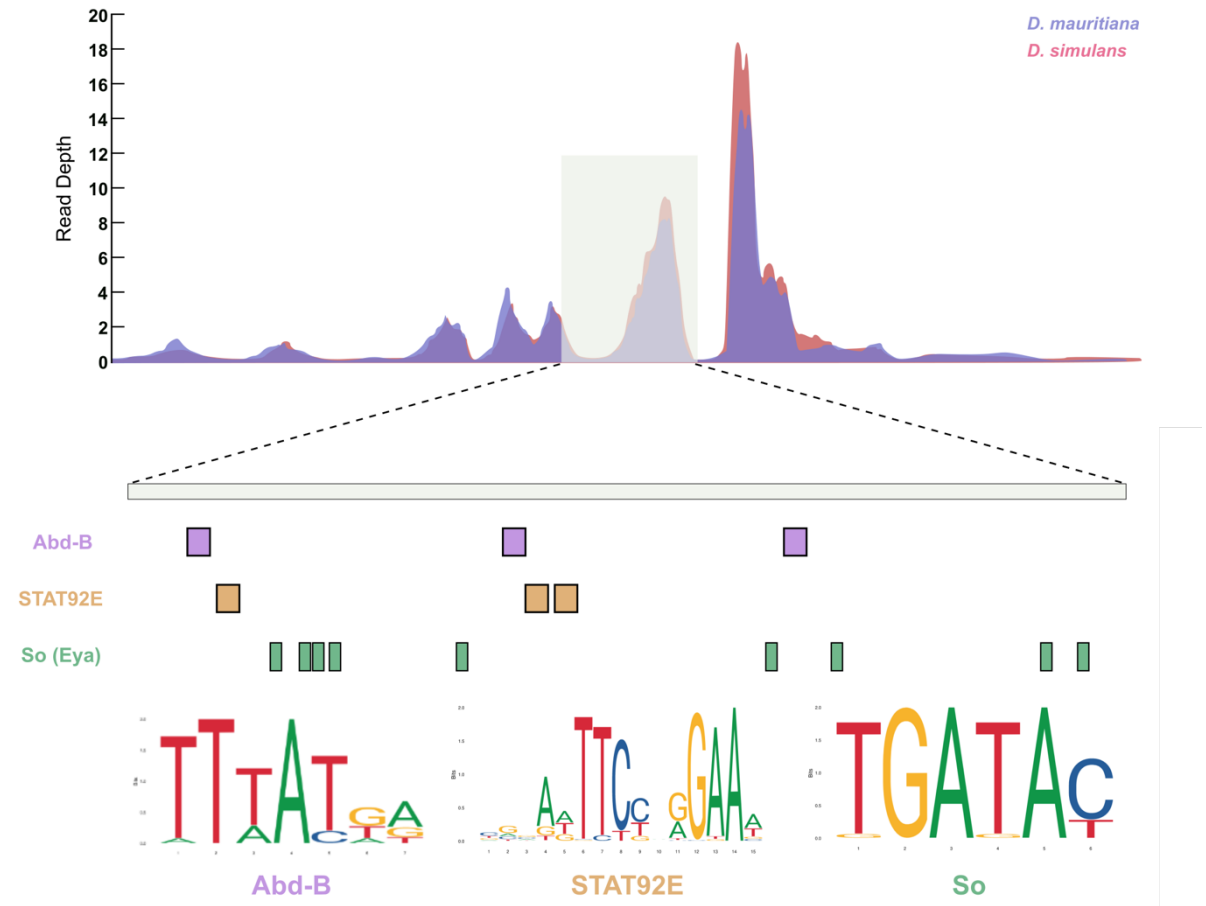


Figure 30: Mapped TF binding sites in *D. melanogaster* Enhancer 3. Abd-B, STAT92E and So are TFs present in the co-opted posterior spiracle network (Glassford *et al.*, 2015). Coloured blocks mark the location of predicted mapped binding sites in Enhancer 3. Binding motifs generated from JASPAR sourced PWMs (Castro-Mondragon *et al.*, 2022). The following p values represent the likelihood of predicted TF binding sites in the enhancer generated from FIMO on MEME suite: Abd-B (0.0000014 - 0.0001068), STAT92E (0.000583 – 0.00632), So (0.00172 – 0.00946).

STAT92E has also been shown to play an essential regulatory role in posterior lobe development (Glassford *et al.*, 2015). Three STAT92E motifs were predicted within Enhancer 3 (Figure 30). Interestingly, two of these predicted binding sites were adjacent to Abd-B motifs, consistent with the arrangement of mapped motifs of these two TFs in the *poxn* enhancer (Figure 30) (Glassford *et al.*, 2015).

eya is expressed in a similar pattern to Enhancer 3 driven expression in the pupal lateral plate and posterior lobe structures (Supplementary Figure 9B). Eya doesn't bind DNA directly but is a co-factor that interacts with multiple other TFs (Pignoni *et al.*, 1997; Ohto *et al.*, 1999),

including Sine oculis (So) (Pignoni *et al.*, 1997). I found nine predicted So binding motifs in Enhancer 3 (Figure 30).

4.4.3 Functional testing of Abd-B binding sites in *D. melanogaster* Enhancer 3

I next aimed to test whether the three Abd-B binding sites predicted in Enhancer 3 were functionally important. The binding site mutation strategy adopted was based on previous manipulation of Ubx binding sites (Figure 31A) (Delker *et al.*, 2019). Hox TFs show preference for binding 'AT' rich sequences (Affolter, Slattery and Mann, 2008), and replacing the flanking sequences of the motif with Gs, decreased the relative binding affinity of Ubx (Delker *et al.*, 2019). A version of the *D. melanogaster* Enhancer 3 GAL4 reporter was generated with all three Abd-B binding sites mutated in this way (Figure 31A).

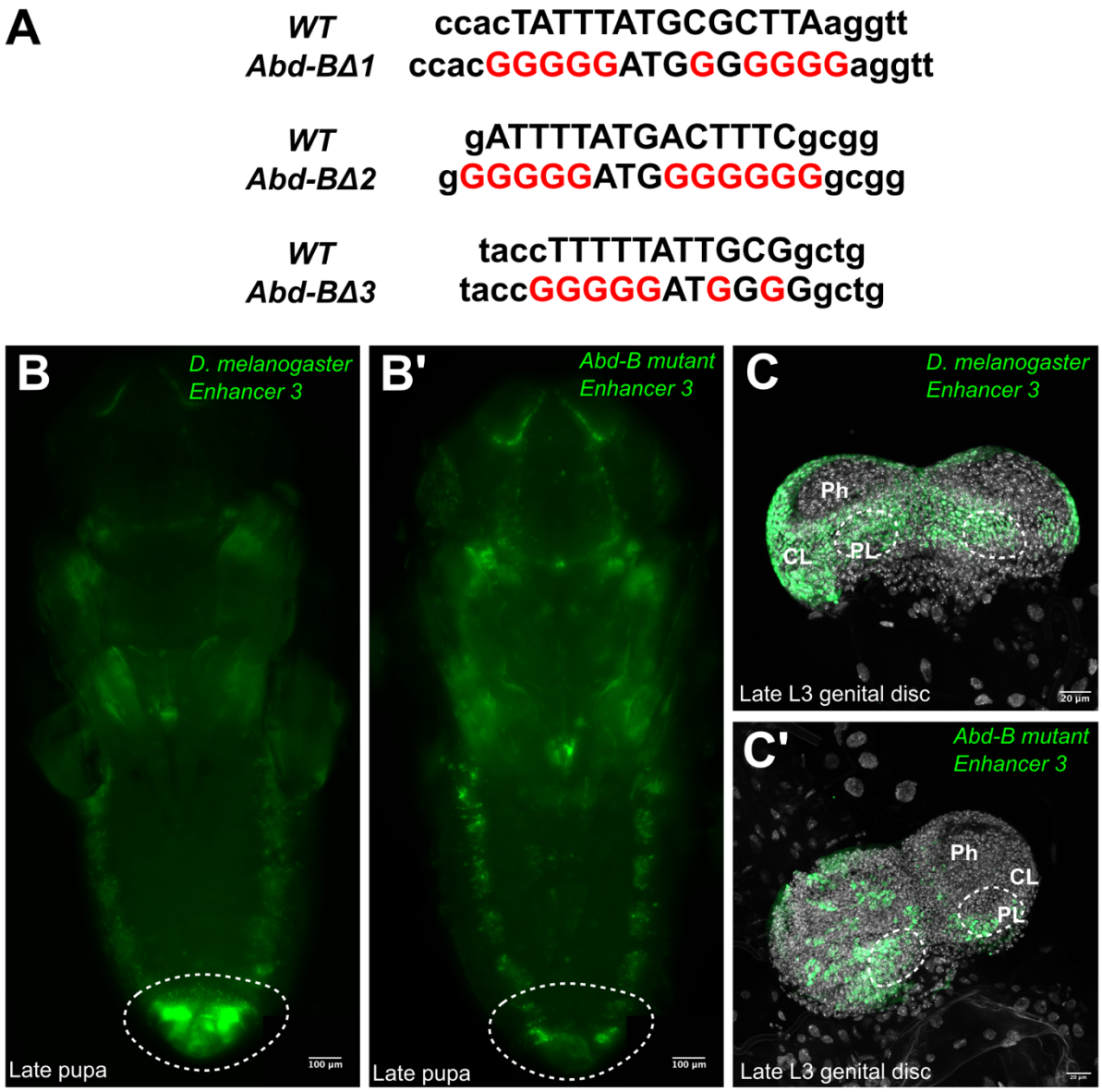


Figure 31: Mutant Abd-B binding sites in *D. melanogaster* Enhancer 3. (A) Mutations in the three mapped Abd-B binding sites in *D. melanogaster* Enhancer 3 (Delker *et al.*, 2019). Red text represents the mutated nucleotides. (B,B') White dashed ovals encircle the male pupal terminalia. (C,C') White dashed ovals encircle posterior lobe primordia in the male genital discs, both ventral views. (C') Mutated Abd-B enhancer 3 has less activity within the posterior lobe primordia compared to (C). Both discs are DAPI stained. (B') n=2, (C') n=5. CL: Clasper, Ph: Phallus, and PL: Posterior lobe.

I initially characterised the activity of the Abd-B mutant Enhancer 3 compared to wild-type *D. melanogaster* Enhancer 3 in live pupae (Figure 31B and B'). I observed a striking loss of activity in the terminalia (Figure 31B' and Supplementary Figure 12), while expression driven by this enhancer was sustained in all other regions of the pupa, consistent with *Abd-b* expression (Figure 31B') (Sanchez-Herrero *et al.*, 1985; Celniker *et al.*, 1990).

Staining for anti-GFP in male genital discs with the mutant Enhancer 3 was reduced compared to the wild-type Enhancer 3 (Figure 31C and C'). This included a reduction of expression driven by the mutant enhancer in the posterior lobe primordia where endogenous *Sox21b* expression is localised (Figure 8 and Figure 31C'). The disruption of Enhancer 3 when all three Abd-B motifs are mutated suggests that this TF plays a role in positively regulating *Sox21b* expression in the male genital disc.

4.4.3.1 Abd-B TF binding site conservation between species

I then compared the three Abd-B binding sites between *D. melanogaster*, *D. simulans*, and *D. mauritiana* Enhancer 3 sequences using MSA. This revealed that all three Abd-B binding motifs are conserved between these species (Supplementary Figure 13). The only notable change is observed in *D. mauritiana,* where upstream of the third Abd-B binding site there is a ten nucleotides deletion (Supplementary Figure 13). However, it is unknown if this deletion affects Abd-B binding or that of other TFs.

I next aligned sequences for Enhancer 3 from three non-lobed species (*D. yakuba, D. ananassae* and *D. biarmipes*) to *D. melanogaster* (Figure 32 and Supplementary Figure 14). The first mapped Abd-B motif was conserved between all lobed and non-lobed species tested (Figure 32A). The second mapped Abd-B motif contained just one SNP in *D. ananassae* (Figure 32B). Interestingly, the upstream flanking nucleotide adjacent to this motif was not conserved in *D. ananassae* and *D. biarmipes* compared to *D. melanogaster* (Figure 32B). The third Abd-B motif was not conserved in any of the three non-lobed species (Figure 32C). This

suggests that the third Abd-B motif may play a role in coordinating *Sox21b* led development of this morphological novelty.

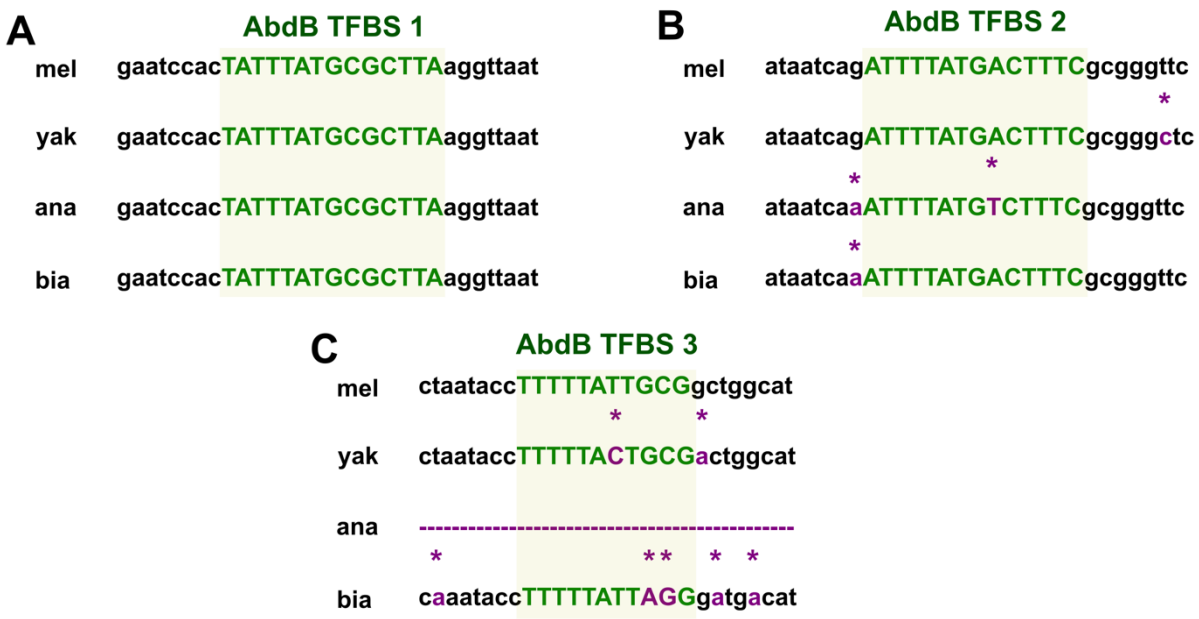


Figure 32: Enhancer 3 Abd-B binding sites in non-lobed species compared to lobed *D. melanogaster*. Capitalised letters and shaded boxes marks the Abd-B TF binding sites. Magenta coloured letters and asterisks marks single nucleotide polymorphisms between non-lobed species and lobed *D. melanogaster*. mel: *D. melanogaster*, yak: *D. yakuba*, ana: *D. ananassae*, bia: *D. biarmipes*. (C) *D. ananassae* contains a deletion in its sequence at this region of the genome.

4.5 Chapter Four Discussion

In this chapter I identified an evolved *Sox21b* enhancer expressed in the posterior lobe primordia cells of the male genital disc, and evidence that this element is positively regulated by Abd-B.

4.5.1 ATAC-seq and enhancer activity comparisons

I was able to use the ATAC-seq data alongside *Sox21b* enhancer reporter line activities to characterise the cis-regulatory network acting upon this gene during genital development. From the nine tested, six enhancer reporter line activities aligned with the chromatin accessibility profiles (Figure 18B and Figure 19). Of these six, four (GMR73B10, GMR73A05, GMR73B01 and GMR73B06) had no activity in the male terminalia consistent with closed chromatin spanning these regions (Figure 18B; Figure 19A, C, E and G). The other two,

Enhancer 1 and Enhancer 2, spanned accessible regions of chromatin and were active in the male terminalia (Figure 21).

Enhancer activity from three out of the nine reporter lines tested did not align with the ATAC-seq accessibility predictions (Figure 18B; Figure 19B, D and F). GMR73B12 and GMR73A11 spanned regions of accessible chromatin, but both enhancers did not drive genital expression during pupal development (Figure 18B; Figure 19B and F). It is thought that accessible chromatin with no detectable enhancer activities are signatures of poised enhancers occupied by histone proteins awaiting TF binding, or epigenetic shadows of enhancer activity at an earlier stage of development (Bonifer and Cockerill, 2017; Koenecke *et al.*, 2017; Ordoñez *et al.*, 2019). In contrast to this, GMR73B04 revealed activity in the distal head of the claspers yet occupied a region of inaccessible chromatin (Figure 20). Assaying for enhancer activity this way has its limitations: enhancers are tested in isolation alleviating the ability to test for synergistic and antagonistic interactions between them, and the sequence being tested is inserted into a region of the genome unrelated to its endogenous context. That said, GMR73B04, GMR73B12 and GMR73A11 enhancer activities may be regulated by flanking cis-acting regions in its native context that is not captured in this assay.

Analysis of ATAC-seq data in isolation hinders a comprehensive understanding of the chromatin environment regulating gene expression. I would endeavour to perform Chromatin Immunoprecipitation and sequencing (ChIP-Seq) against certain histone modifications in larval and pupal genital samples of *D. melanogaster, D. simulans* and *D. mauritiana* to better characterise the chromatin environment. For example, poised enhancers exhibit a unique epigenetic signature abundant in H3K4me1 and absent of H3K27ac (Koenecke *et al.*, 2017). H3K27ac presence is characteristic of an active enhancer (Creyghton *et al.*, 2010). By comparing the profiles for these two histone marks to the ATAC-seq data, this would enable me to better predict enhancer activities and functional importance.

The ATAC-seq data analysed in this chapter originated from the entire male terminalia including the surrounding abdomen (Hagen *et al.,* unpublished). Consequently, the chromatin profiles represent an average of heterogenous cell types, thus suppressing the identification of dynamic subcellular populations within the genitalia. To overcome this, I would generate single cell ATAC-seq (scATAC-seq) datasets for these three species in larval and pupal genital samples. scATAC-seq would help to identify posterior lobe specific *Sox21b* enhancers in genital discs. Interestingly, scATAC-seq of the eye-antennal discs in *Drosophila* has previously matched 85% of enhancer accessibility profiles to Janelia reporter line activity patterns, whereby the remaining 15% mapped to pioneer TF binding at primed region with no activity

(Bravo González-Blas *et al.*, 2020). As well as mapping *Sox21b* enhancers at a higher resolution, this would generate a dataset beneficial for the evo-devo *Drosophila* genitalia community.

4.5.2 Pupal genital activity of *Sox21b* enhancers

Enhancers 1, 2 and 3 all showed expression during pupal development despite *Sox21b* not being expressed at this stage, apart from at the hypandrium and the surrounding abdominal tissue (Figure 21 and Figure 27). There are a few reasons why this may be. Firstly, this could be a result of the GFP fluorescence half-life resulting in detectable protein past the time of temporal activation. The signal seen in pupa may be reminiscent of the larval expression captured. It should also be noted that there is a temporal delay between the transcription and translation of the GAL4 protein, to then binding the UAS sequence for activation of GFP transcription and translation. To overcome this obstacle, I would plan to use a destabilised version of the GFP fluorophore to capture better temporal precision of enhancer activity (Li *et al.*, 1998). Secondly, as the enhancer reporter line is tested for activity outside its endogenous location, repressor binding sites elsewhere in the locus might have been excluded. It has been shown that testing multiple enhancers simultaneously can reduce overall enhancer activity (Waymack, Gad and Wunderlich, 2021). To investigate the temporal specificity of putative *Sox21b* enhancers during genital development, I would aim to screen a series of overlapping enhancer fragments to help map repressive and active TF binding sites from observed activity comparisons.

4.5.3 Species-specific activity of *Sox21b* Enhancer 3

I used *D. mauritiana* and *D. simulans* versions of Enhancer 3 to reveal spatial activity differences within the male genital disc (Figure 28). The Enhancer 3 activity differences were consistent with the species difference previously seen for *Sox21b* expression (Figure 14 and Figure 28). This result suggests that *Sox21b* evolution between *D. mauritiana* and *D. simulans* derives from Enhancer 3 nucleotide differences. To prove this, I would look to generate versions of *D. simulans* Enhancer 3 where certain parts are of *D. mauritiana* identity, to help identify the region(s) driving differential activity between species. For example, *D. mauritiana* Enhancer 3 contains a 9 bp deletion at nucleotides 511 to 520, relative to *D. simulans* (Supplementary Figure 11). By designing and generating a *D. simulans* Enhancer 3 $^{\Delta 511\text{-}520}$ reporter line and comparing its activity with the *D. mauritiana* Enhancer 3 reporter line, this could help decode *Sox21b* evolution.

The species-specific Enhancer 3 reporter lines were assayed in *D. melanogaster* where the TF environment present may differ and thus enhancer expression is not a genuine representation of the species itself. Therefore, the species comparison of Enhancer 3 activity should be compared with caution. To overcome this, I would plan to generate Enhancer 3 reporter lines in the endogenous species. *D. simulans* and *D. mauritiana* transgenic lines carry attp2 landing sites (Stern *et al.*, 2017) would be co-injected with the species-specific Enhancer 3 GAL4 construct and a plasmid encoding *nanos > phiC31 integrase* for the creation of these lines. This would alleviate false positive and negative Enhancer 3 activity results in response to a *D. melanogaster* proteome.

4.5.4 Abd-B regulates *Sox21b* expression

I found evidence that the three predicted Abd-B binding sites in *Sox21b* Enhancer 3 are functional (Figure 30). Abd-B is a Hox TF coordinating the regulation of many genes during posterior lobe development (Glassford *et al.*, 2015), and has been shown to promote apoptotic activity in the female *Drosophila* nervous system with its cofactor, Dsx^F (Ghosh *et al.*, 2019). As *Sox21b* represses posterior lobe growth, it is possible that Abd-B may positively regulate this gene at Enhancer 3, to coordinate downstream cell-death, and hence repress lobe size. To test this theory, I would perform ChIP-Seq against Sox21b in male genital disc samples to map its binding sites within the genome and thus identify if predicted target genes are linked to apoptotic pathways.

By mutating three Abd-B motifs in *D. melanogaster* Enhancer 3, I showed that its activity was severely suppressed throughout the entire genital disc, including the posterior lobe primordia, compared with wild-type *D. melanogaster* Enhancer 3 reporter line activity (Figure 31C and C'). This observation suggests that Abd-B positively regulates *Sox21b* expression at Enhancer 3 in male genital discs. I would look to confirm this observation by mutating all three Abd-B motifs at the endogenous *D. melanogaster* Enhancer 3 sequence and screen for *Sox21b* expression differences. However, the Enhancer 3 Abd-B mutant retained partial posterior lobe primordia activity (Figure 31C'). This means that additional TFs, in addition to Abd-B, are required to coordinate Enhancer 3 activity. To characterise which TFs also bind to and affect Enhancer 3 activity, I would replicate the TFBS mutation approach used for Abd-B but for STAT92E and So mapped motifs within Enhancer 3 (Figure 32).

Multiple sequence alignment of Enhancer 3 between lobed and non-lobed *Drosophila* species drew attention to the third Abd-B binding site. This Abd-B TF binding site was the only one not conserved between lobed and non-lobed species (Figure 32C and Supplementary Figure 14). It would be insightful to understand whether these nucleotide differences could lead to differential Enhancer 3 activity between lobed and non-lobed flies. To assess this, I would generate *D. melanogaster* Enhancer 3 reporter lines, where the third Abd-B motif is replaced with a non-lobed species version (Figure 32C). *D. ananassae* does not contain the third Abd-B motif (Figure 32C). By deleting just this motif from Enhancer 3 reporter lines, it will identify the functional importance of this site compared to when all three motifs were abolished previously (Figure 31C'). The genetic origin of the posterior lobe novelty is not currently known. By extending my analysis of *Sox21b* Enhancer 3 to non-lobed species, this will help to decipher whether *Sox21b* plays a part in coordinating posterior lobe evolution between species outside of the *D. simulans* clade.

5. Chapter Five: Results

A comparative assessment of *C15*'s role during the development of the terminalia compared to the pretarsal segment of the leg

5.1 Background: *Drosophila* male genitalia and legs

The genital disc is composed of three primordia and is bilaterally symmetrical (Chen and Baker, 1997), whereas a leg disc arises from just one primordium and is asymmetric (Figure 33) (Serrano and O'Farrell, 1997). Differences between the genitalia and leg are clear when assessing which Hox genes determine the identity of each disc (Akam, 1995). *Abd-B* is a posterior Hox gene expressed in the genitalia but not in the developing legs (Akam, 1995; Feng *et al.*, 2022). Estrada *et al* (2001) investigated the function of *Abd-B* during the development of genital structures by using *Abd-B* mutant clones. Loss of *Abd-B* induced the homeotic transformation of the genitalia into leg and antennal structures (Estrada and Sánchez-Herrero, 2001). This suggests that *Abd-B* has hierarchical control over downstream factors during genitalia development and represses the adoption of a leg-determining developmental pathways. Interestingly, comparative genetic studies have shown that the development of the genital and leg disc are regulated by similar genes downstream of Hox factors (Figure 33) (Gorfinkiel, Sánchez and Guerrero, 1999; Estrada and Sánchez-Herrero, 2001; Keisman and Baker, 2001; Gorfinkiel, Sánchez and Guerrero, 2003; Foronda *et al.*, 2006).

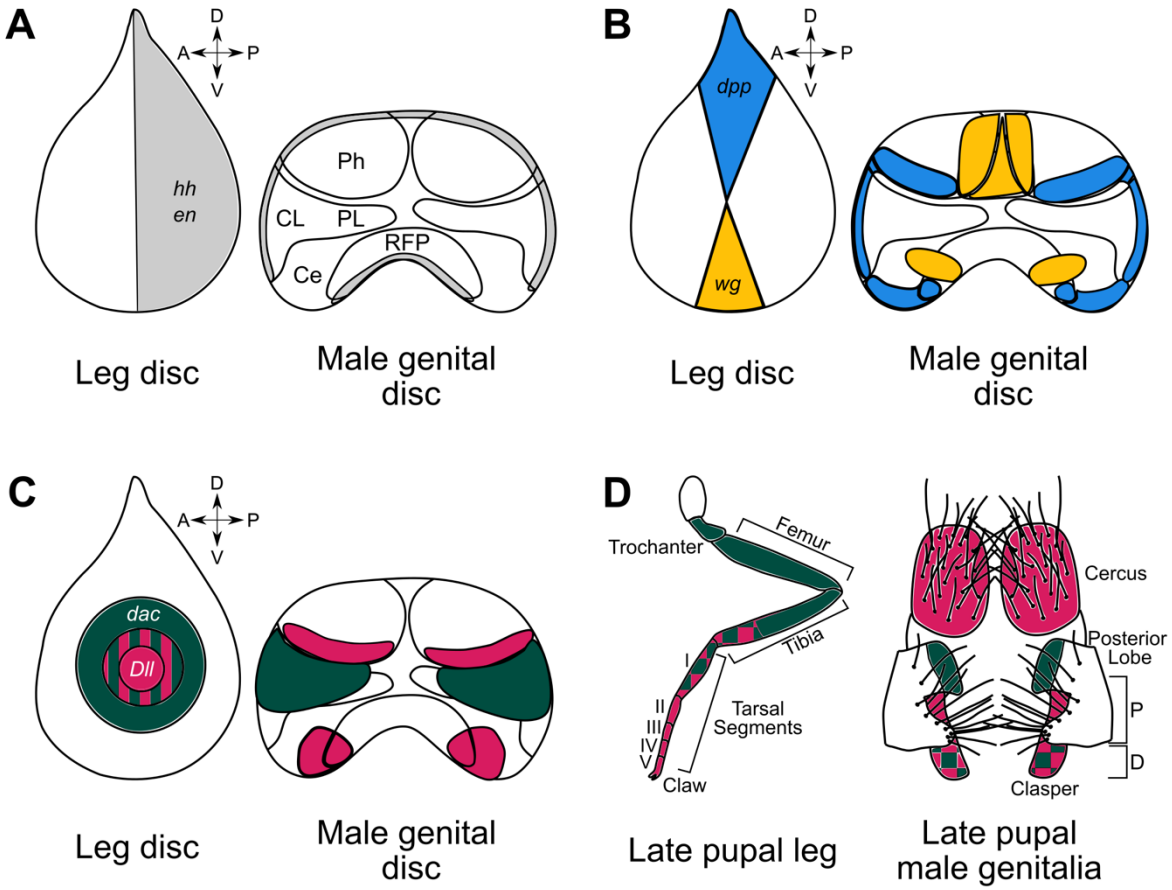


Figure 33: Expression patterns of key signalling and patterning genes in the developing leg versus the developing male genitalia. Figure assembled from data presented in Emerald and Roy, 1998; Gorfinkiel, Sánchez and Guerrero, 1999; Keisman and Baker, 2001; Gorfinkiel, Sánchez and Guerrero, 2003; Ruiz-Losada et al., 2018. The Dorsal/Ventral (D/V) and Anterior/Posterior (A/P) axes are relative to the leg disc. Genital discs are of a dorsal view and the annotations in (A) mark the following: CL: Clasper, Ph: Phallus, PL: Posterior lobe, Ce: Cercus, and RFP: Repressed female primordium. (A) *Hedgehog* (*hh)* is regulated by *engrailed* (*en*), and establishes the A/P boundary in the leg disc. The genital disc is composed of three primordia (A8: RFP, A9: male primordium and A10: anal primordium), each with their own polarity. *hh* and *en* are expressed in the RFP and surrounding tissue of the male genital and cercus primordia. (B) In response to *hh*, *dpp* and *wg* are expressed across the A/P boundary in opposing dorsal and ventral regions of the leg disc, and do not exhibit overlapping expression patterns. *wg* and *dpp* also do not overlap in expression in the genital disc, where *dpp* is expressed in the cercus, phallic and clasper primordia and *wg* is expressed in the phallic primordium and RFP of the disc. (C,D) In response to the highest concentrations of *wg* and *dpp*, *Dll* is expressed in the centre of the leg disc, where it subsequently regulates the development of tarsal leg segments. *dac* expression is regulated by lower expression levels of *dpp* and *wg*, and marks the fifth tarsal segment, tibia, femur, and trochanter. In the genital disc, *Dll* is expressed in the phallic and cercus primordia where *dpp* and *wg* are highly expressed, but during late pupal development *Dll* expression localises to the claspers and cercus. *dac* expression in the genital discs marks the clasper and posterior

lobe primordia and is sustained during pupal development. Note, the internal genitalia are not shown in (D) for simplicity. Diagrams are not to scale. P (proximal) and D (distal) clasper domains.

Early axis specification and patterning of genital and leg discs during larval development exhibit many similarities (Figure 33) (Tabin, Carroll and Panganiban, 1999; Ruiz-Losada *et al.*, 2018). Here, I focus on the male genital disc in comparison to the leg disc. The first similarity between leg and genital development are the spatiotemporal expression dynamics of key morphogens (Figure 33). In leg discs, *hedgehog* (*hh*) and engrailed (*en)* determine the posterior compartment (Figure 33A) (Basler and Struhl, 1994; Kopp, Muskavitch and Duncan, 1997). This is also the case for each of the three primordia in the male genital disc (Figure 33A) (Emerald and Roy, 1998). Both discs subsequently express *dpp* and *wg* which provide dorsal-ventral positional information (Figure 33B) (Jiang and Struhl, 1996; Emerald and Roy, 1998). These two morphogens maintain mutually exclusive expression domains in both discs (Figure 33B). Axis formation in the two discs is followed by expression of TFs, such as *dac* and *Dll*, which organise the discs into distinct regulatory regions required to coordinate the development of specific structures, such as male claspers or the leg tibia (Figure 33C and D) (Gorfinkiel, Sánchez and Guerrero, 1999; Keisman and Baker, 2001).

Comparisons of TF expression patterns between genital and leg discs have revealed striking similarities (Sánchez *et al.*, 1997; Gorfinkiel, Sánchez and Guerrero, 1999; Estrada and Sánchez-Herrero, 2001). For example, *Dll* expression in the four distal tarsal segments in the leg is reflected in the proximal domain of the clasper and throughout the cercus in the male genitalia (Figure 33D), while overlap of *dac* and *Dll* in the first tarsal segment and part of the tibia, is similar to their expression in the distal clasper domain (Figure 33D). This suggests there is homology between the proximal clasper domain and cercus to tarsal segments V to II and the claw, the distal clasper to the first tarsal segment and proximal tibia, and the posterior lobe tip to the rest of the tibia, femur and trochanter (Figure 33) Emerald and Roy, 1998; Gorfinkiel, Sánchez and Guerrero, 1999; Keisman and Baker, 2001; Gorfinkiel, Sánchez and Guerrero, 2003; Ruiz-Losada et al., 2018.

5.1.1 C15 is a pretarsus determination factor

The pretarsal segment of *Drosophila* legs is distinguished by a protruding claw that resembles two ‘hooks’ perpendicular to the leg (Figure 34A) (Chovnick and Lefkowitz, 1956). The homeobox gene *aristaless* (*al*) is expressed in the pretarsal primordium, and *al* mutants are missing the pretarsal claw (Campbell, Weaver and Tomlinson, 1993; Campbell, 2002; Galindo *et al.*, 2002; Beermann and Schröder, 2004; Kojima, Tsuji and Saigo, 2005; Newcomb *et al.*,

2018). However, an additional pretarsal factor, *C15* (*clawless*), was identified by screening for pretarsal phenotypic defects in weak *al* allelic mutants (Campbell, 2005). *C15* mutants show near complete loss of *al* expression in the centre of the leg disc and vice versa (Campbell, 2005). Additional mutant studies revealed a third pretarsal marker, *lim1* (Pueyo *et al.*, 2000; Pueyo and Couso, 2004; Campbell, 2005; Kojima, Tsuji and Saigo, 2005). *C15*, like *al* and *lim1*, is positively regulated by Epidermal Growth Factor Receptor (EGFR) signalling at the centre of the leg disc (Galindo *et al.*, 2002; Galindo, Bishop and Couso, 2005; Newcomb *et al.*, 2018). C15 then both promotes and represses target gene expression to precisely define TF expression domains throughout the pretarsal segment and surrounding tibia segments (Figure 34B) (Campbell, 2005; Mojica-Vázquez *et al.*, 2017). Furthermore, Kojima *et al* (2005) found that C15 and Al can bind cooperatively, for example to repress *bar* expression (Figure 34B) (Miyazono *et al.*, 2010), illustrating the functional importance of *al* and *C15* expression overlap in the centre of the leg discs.

Pretarsal TFs downstream of *C15* (*bowl, bab2* and *ap*) have previously been shown to have expression in the claspers during at least one developmental stage (Supplementary Figure 15), whereas *al* did not (Estrada and Sánchez-Herrero, 2001; Keisman and Baker, 2001; Chatterjee *et al.*, 2011; Vincent *et al.*, 2019). *C15* is one of the top 100 expressed TFs in the developing *D. melanogaster* male genitalia, and this gene is expressed in part of the clasper and the cercus at 48 hAPF (Supplementary Figure 15A) (Vincent *et al.*, 2019). To date, however, the function of *C15* has not been tested in the developing male genitalia.

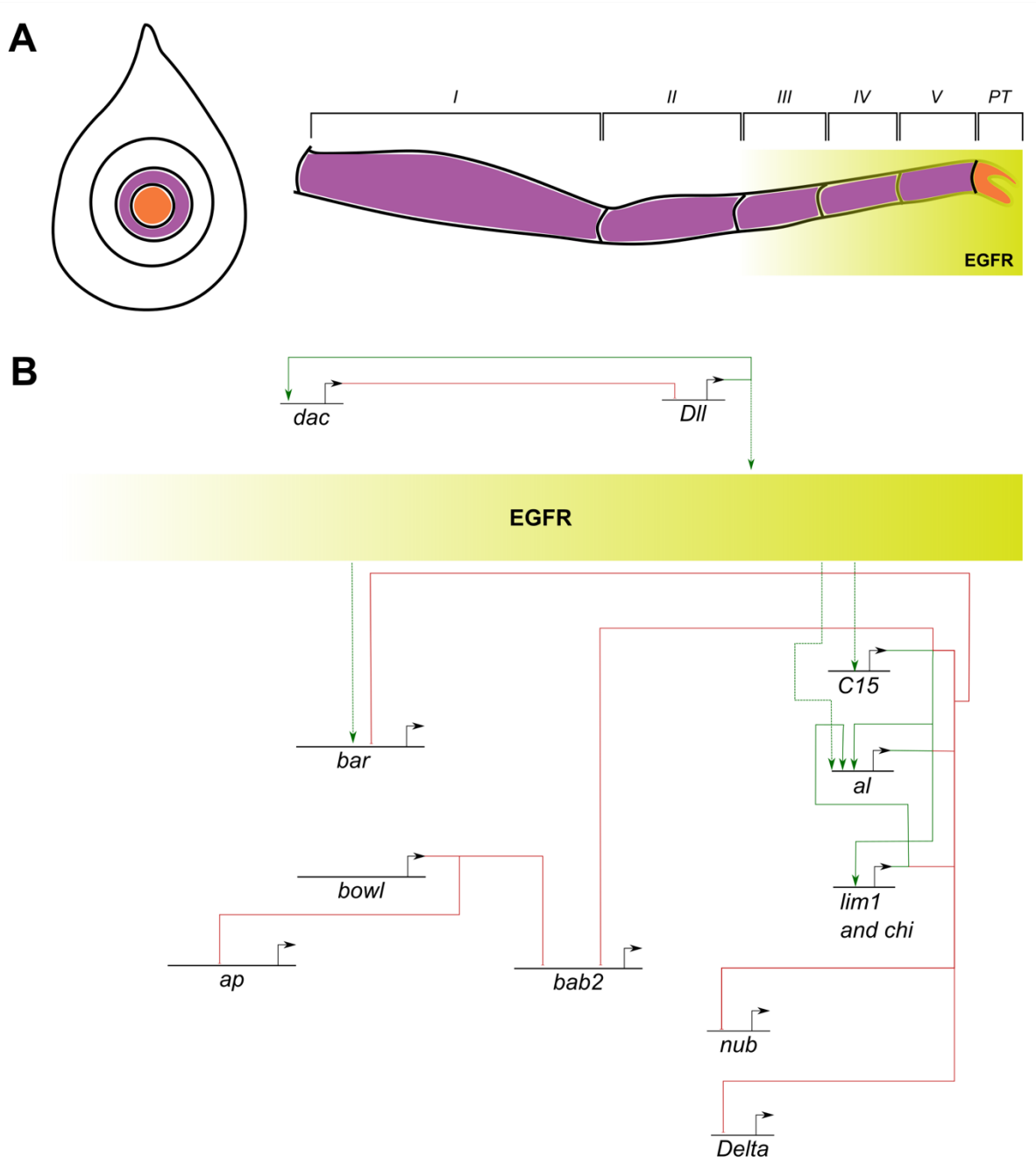


Figure 34: A simplified schematic highlighting the pivotal interactions originating from *C15* expression in the pretarsal segment of the leg. This figure has been assembled from data presented in (de Celis Ibeas and Bray, 2003; Pueyo and Couso, 2004; Campbell, 2005; Kojima, Tsuji and Saigo, 2005; Miyazono *et al.*, 2010; Natori *et al.*, 2012; Baanannou *et al.*, 2013; Kachhap, Priyadarshini and Singh, 2017; Mojica-Vázquez *et al.*, 2017; Ruiz-Losada *et al.*, 2018; Tse J *et al.*, 2022). (A) The pretarsal (PT) segment is represented by the orange central circle in the leg disc, with tarsal segments I – V represented by the surrounding purple concentric ring. Upon metamorphosis, the leg disc begins to protrude outwards to form the adult leg with the segments determined by specific gene expression domains. (B) A simplified GRN showing *C15*'s interactions with other genes. Upstream of *C15*, *Dll* promotes *dac* expression, and *dac* represses *Dll* expression. *Dll* is one of the TFs required to promote

EGFR signalling in the leg pretarsus, which then acts to regulate *C15* and *bar* expression. EGFR has different target genes dependent on the concentration; a high concentration activates *C15* and *al*, whereas a low concentration activates *bar*. *C15* acts with *al* cooperatively to regulate *lim1*, *nub*, *delta* and *bar*, and alone to regulate *al* and *bab2*. Chi is a Lim1 protein interaction partner, known to also interact with other LIM proteins such as *ap*. Downstream targets of *C15* target genes are not shown.

5.1.2 Chapter five aims

Legs and external genital structures have long been speculated to share an ancestral GRN and may even be derived homologous structures (Casares *et al.*, 1997; Gorfinkiel, Sánchez and Guerrero, 1999; Estrada and Sánchez-Herrero, 2001; Boudinot, 2018). In this chapter, I compared the development of male and female genital structures to the leg pretarsus in order to better understand differences and similarities between their GRNs. This was achieved by:

- Characterising the expression and function of *C15* and other known leg genes in the developing genital structures.
    - Antibody staining and HCR-IS were used to assess gene expression patterns in male and female genital structures.
    - RNAi knockdown and UAS overexpression crosses were generated to identify those required for genital development.

- Testing whether characterised pretarsus gene interactions are conserved in the developing genitalia.
    - RNAi knockdowns were stained to assess changes in *C15* and other pretarsal gene expression patterns.

- Identifying and validating *C15* enhancers and their activity.
    - Enhancer reporter lines were screened to characterise activity patterns in larval discs.
    - TF binding sites were mapped to an identified male genital *C15* enhancer to understand its upstream regulation.

5.2 Results: Characterisation of the pretarsal *C15* GRN in the male claspers

5.2.1 *wg* and *dpp* expression in the male pupal claspers

Expression of *wg* and *dpp* overlaps in the centre of the leg disc, which subsequently induces *dac* and *Dll* expression, and thus leads to pretarsal *C15* expression (Diaz-Benjumea, Cohen and Cohen, 1994; Estella and Mann, 2008; Newcomb *et al.*, 2018). I first sought to determine whether *wg* and *dpp* upstream specification of *C15* expression in the leg were replicated at the male claspers.

Although *wg* is known to be expressed in the cercus and phallic male primordia in the larval genital disc, its expression during pupal stages was not previously characterised with respect to clasper development (Figure 33B) (Chen and Baker, 1997). I therefore conducted an HCR-IS against *wg* in the male pupal terminalia (Figure 35A). At 40 hAPF, *wg* is expressed in the pupal cercus and phallus consistent with its genital disc expression (Figure 35A). I found that *wg* is also expressed in the clasper surrounding the developing bristles, with the most abundant expression localised in the proximal clasper domain (Figure 35A).

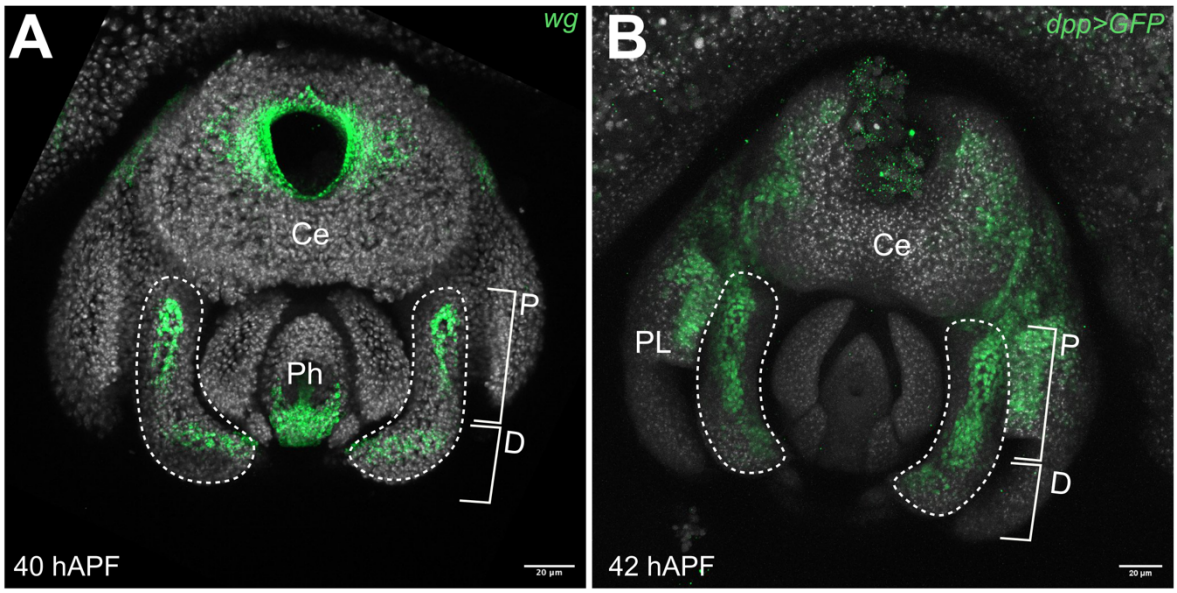


Figure 35: *wg* and *dpp* are both expressed in the male claspers during male pupal genital development. (A) HCR-IS of *wg* reveals expression localised around clasper bristle primordia, part of the phallic region and part of the cerci. (B) Anti-GFP staining in *dpp-GAL4>UAS-GFP* fixed male terminalia showed *dpp* expression surrounding the bristles in the clasper proximal domain, the tips of the posterior lobes and periphery of the cerci. (A) n=4, (B) n=1. All samples were stained with DAPI. White dashed outlines mark the claspers. Ph: Phallus, PL: Posterior lobe and Ce: Cercus. Distal (D) and proximal (P) mark the two clasper domains.

dpp expression was previously described in the claspers, phallic, cerci and repressed female primordia (RFP) of male genital discs (Figure 33B). I used a *dpp-GAL4* line crossed to *UAS-*

GFP, to assay where this gene is expressed in the genitalia during pupal development. This showed that the claspers and cerci sustained *dpp* larval expression (Figure 35B). Within the claspers, *dpp* expression predominantly localised to the proximal clasper domain (Figure 35B).

5.2.2 *dac* and *Dll* expression hubs in the male claspers

As *dpp* and *wg* are both expressed in the clasper during pupal development (Figure 35), I next looked to test whether *dac* and *Dll* are also expressed in the clasper, as demonstrated during leg development (Estella and Mann, 2008).

I used a line carrying a hypomorphic allele of *dac* with a GAL4 insertion crossed to *UAS-GFP* to assay *dac* expression (Figure 36A) (Tavsanli *et al.*, 2004). *dac* was shown to predominantly localise to the distal clasper domain, where weaker expression can be seen in the proximal domain (Figure 36A). I phenotyped the *dac* hypomorphic mutants and observed a loss of bristles in the distal domain of the clasper (Figure 36C and Supplementary Table 7), which supported its role in promoting bristle development, and was consistent with expression data (Figure 36A and Supplementary Figure 16B). The bristles also had a more elongated appearance, similar to that of the single mechanosensory bristle present in wild type claspers (Figure 36C; Supplementary Figure 16A and B). These findings suggest that *dac* aids clasper bristle development but may also act to repress bristle extension and thus limit bristle length.

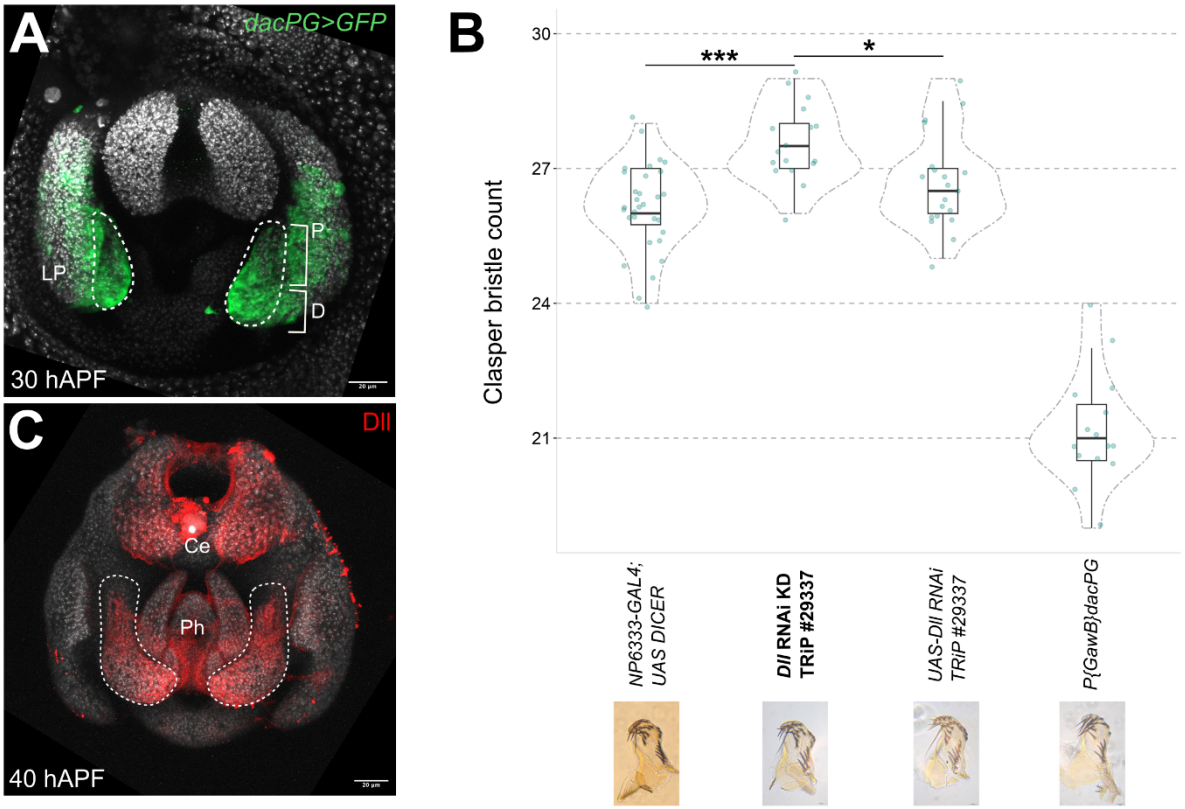


Figure 36: *dac* and Dll expression and function during male clasper development. (A) *dac* is expressed in the distal and proximal clasper domains and at the boundary with the lateral plate. P: Proximal and D: Distal clasper domains. LP: Lateral plate. (B) During pupal development, Dll is expressed throughout the claspers, phallus and cerci (Gorfinkiel, Sánchez and Guerrero, 1999). Ph: Phallus, Ce: Cercus. (C) *Dll* and *dac* are both required for clasper bristle development. The lines adjoining the *Dll* RNAi knockdown to the parental *UAS-RNAi* and *NP6333-GAL4, UAS-DICER* lines indicates the level of statistical significance between each pair (Supplementary Table 7). The *P{GawB}dacPG* line is not statistically compared to other lines due to the absence of parental lines being readily available. *** $p < 0.001$ and * $p < 0.05$. $n > 14$ for all lines phenotyped.

To stain for Dll protein expression, I used an anti-Dll antibody on fixed male terminalia samples. In the male genital disc, *Dll* has been shown to localise to the cercus and phallic primordia (Figure 33C) (Gorfinkiel, Sánchez and Guerrero, 1999), which I also observed during pupal development (Figure 36B). However, I found that Dll was also expressed in the clasper during pupal development in both the proximal and distal domains (Figure 33C and Figure 36B), where RNAi knockdown of *Dll* generated a 4% increase in clasper bristles (Figure 36C and Supplementary Figure 16C) (Gorfinkiel, Sánchez and Guerrero, 1999). In addition, *Dll* knockdown decreased the cercus size and bristle count, consistent with previous *Dll* mutant phenotypes (Supplementary Table 7) (Gorfinkiel, Sánchez and Guerrero, 1999).

In summary, I have found that the two TFs upstream of *C15* in the pretarsal leg GRN, Dll and *dac,* are both expressed in the male claspers and help regulate bristle development on these structures (Figure 34B and Figure 36). I next aimed to characterise the expression of the pretarsal determining factor, *C15,* downstream of Wg and Dpp, in the claspers during male genital development.

5.2.3 C15 protein localisation across development of the male terminalia

I assayed for C15 expression in male genital samples by staining with anti-C15 during larval and pupal development (Figure 37) (Campbell, 2005). I double stained this with Opa which is known to mark the entire developing claspers (Vincent *et al.*, 2019). I found that C15 is localised to the clasper primordium in larval genital discs as inferred from a partial overlap with Opa (Figure 37A-A"). In pupae, C15 remains localised to the clasper but is restricted to cells in the proximal domain that do not develop bristles (Figure 37B-D and Supplementary Figure 16D). This is consistent with previous analysis of the mRNA expression of this gene (Vincent *et al.*, 2019).

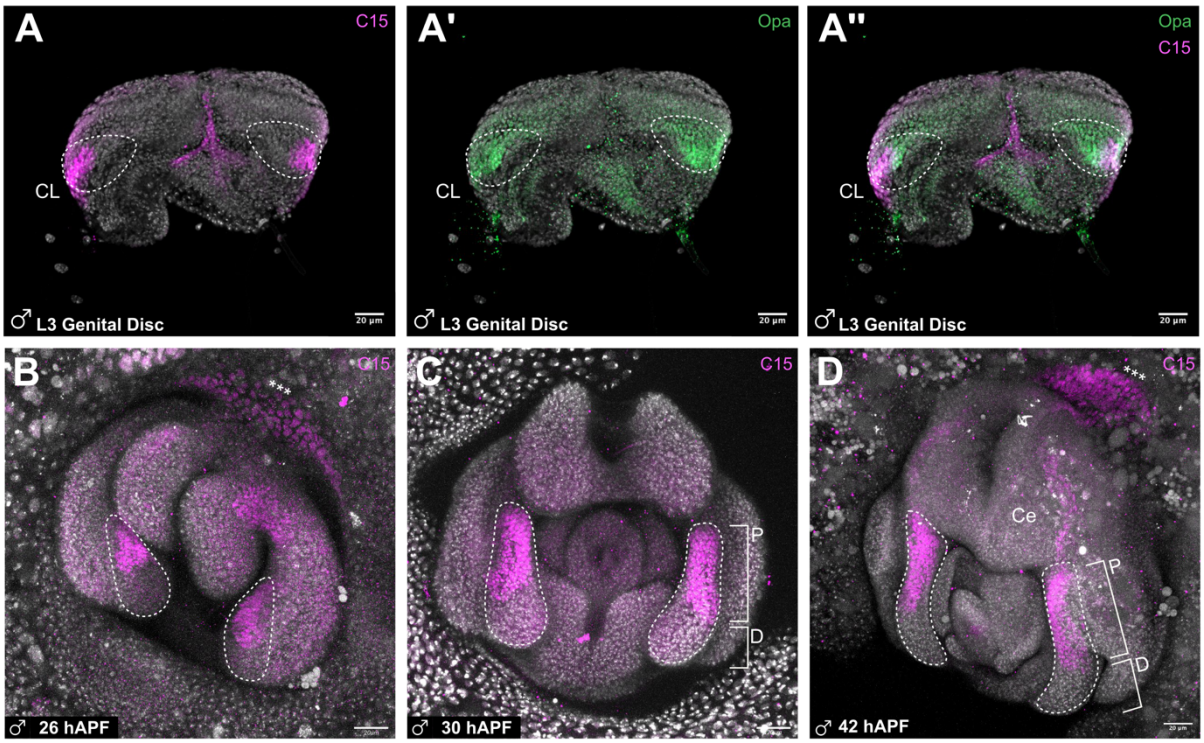


Figure 37: C15 expression in male genital and cercus structures during development. (A-A") C15 expression partially overlaps with the clasper marker Opa in the male genital disc. CL: Clasper. (B-D) C15 expression is observed throughout the developing male pupal tissue. C15 is expressed in the proximal clasper domain and peripheral boundaries of the cerci throughout development. All stains have

been carried out in *D. melanogaster* w^{1118}. White dashed outlines mark the developing claspers. (C and D) Proximal (P) and distal (D) domains of the clasper are labelled. Ce: Cercus. (B and D) *** marks the abdominal dorsal tissue. (A-A") n=4, (B) n=4, (C) n=2, (D) n=4. All samples were stained with DAPI.

C15 expression also extends from the proximal domain of the claspers to the periphery of the cerci from as early as 26 hAPF (Figure 37B). In later pupal stages, peripheral cerci C15 expression is 3 to 4 cells wide (Figure 37D). I observed that C15 is also expressed in the abdominal tissue dorsal to the terminalia (Figure 37B and D).

5.2.4 RNAi knockdown of *C15* in male terminalia

I then attempted to characterise the role of *C15* in the male terminalia using the GAL4-UAS system to knockdown expression of this gene using RNAi.

Knockdown of *C15* caused a 17% increase in clasper bristles (Figure 38A and Supplementary Table 7). The ectopic bristles developed at the proximal domain of the clasper, a region normally void of bristles (Figure 38D-E') (Smith and Sondhi, 1961), consistent with C15 expression in this region (Figure 37). This result indicates that C15 normally represses clasper bristle development, in contrast to its known role in activating bristle development elsewhere in the fly (Campbell, 2005; Kojima, Tsuji and Saigo, 2005). I overexpressed *C15* with the same GAL4 driver as used for RNAi knockdowns, but no clasper bristle effect was observed (Supplementary Table 7). Note that it was not possible to measure clasper size because of the 3D nature of the tissue preventing consistent mounting.

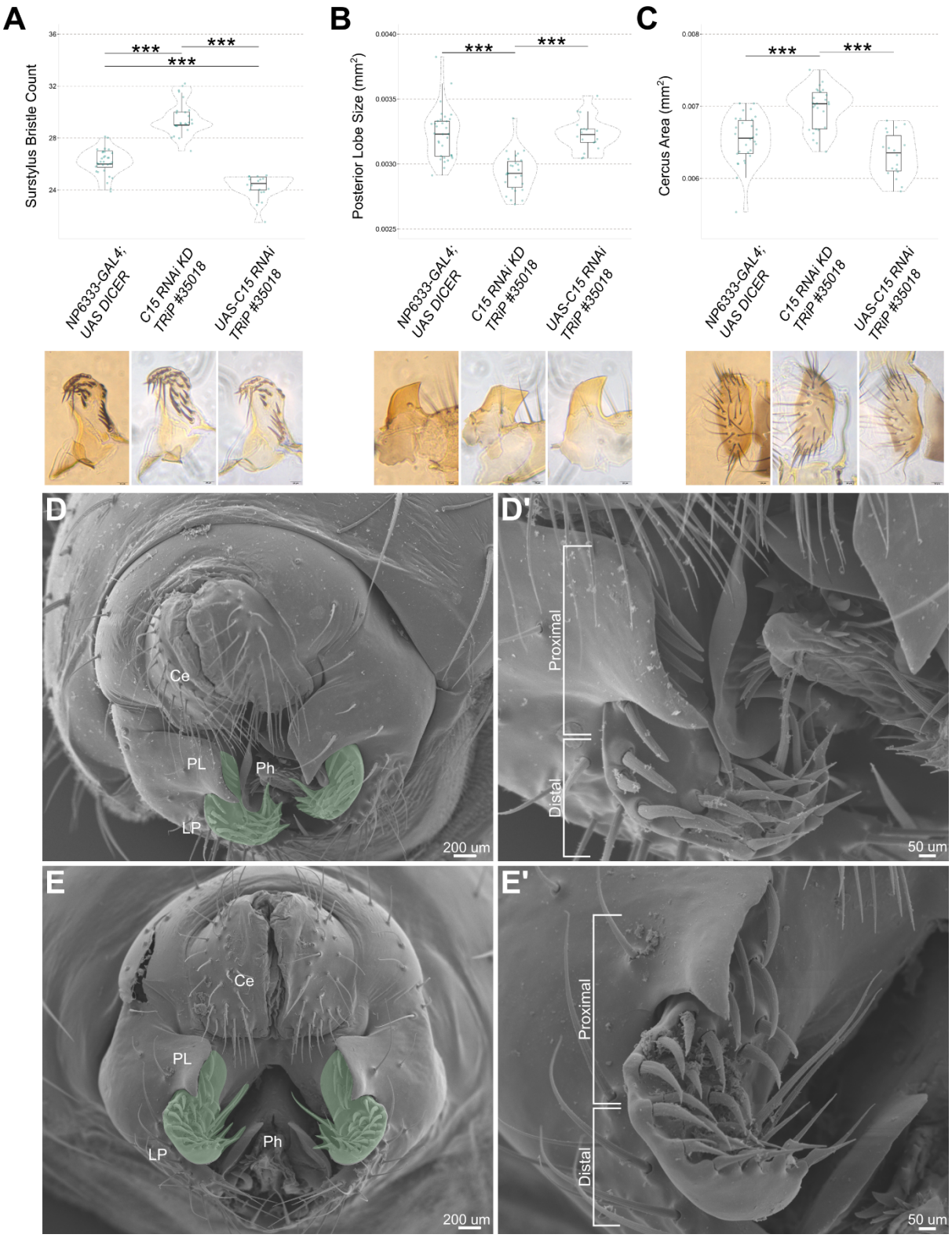


Figure 38: RNAi knockdown of *C15* in male *D. melanogaster* developing genitalia. (A) Knockdown of *C15* expression in the developing male genitalia led to a significant increase in clasper bristle number, specifically localised to the proximal domain, (B) reduction in posterior lobe area and (C) increase in cercus size. (A-C) Scale bars are 20 µm, *** represents the statistical significance level of $p < 0.001$, n>17 (Supplementary Table 7). (D,D') SEM images of *NP6333-GAL4, UAS-DICER* parental control and (E, E') of *NP6333-GAL4, UAS-DICER* > *UAS C15* RNAi knockdown. (E') An increase in clasper bristles

can be seen at the proximal clasper domain compared to (D'). (D, E) Green highlighted structures are of the claspers. Scale bars are 200 μm (D,E) and 50 μm (D', E'). (D', E') Proximal and distal regions of the clasper annotated.

Alongside *C15*'s clasper role, RNAi knockdown also revealed significant posterior lobe and cercus size phenotypes (Figure 38B, C, D and E; Supplementary Table 7), despite only being expressed at the peripheral boundary to these structures (Figure 37). No difference in cercus bristle number was detected upon *C15* knockdown (Supplementary Figure 17C).

5.2.5 Al and Lim1 expression during male pupal genital development

C15, *al* and *lim1* occupy the top of the GRN coordinating pretarsal claw development (Tsuji *et al.*, 2000; Campbell, 2005; Kojima, Tsuji and Saigo, 2005). I next looked to see whether the overlap in expression of these three TFs in the leg (Supplementary Figure 18A and B) (Campbell, 2005; Kojima, Tsuji and Saigo, 2005), was replicated in the male clasper during genital development.

During male genital development, I did not detect any Al expression in the clasper primordia during larval and pupal stages, despite confirmed C15 clasper expression (Figure 37; Figure 39A and A').

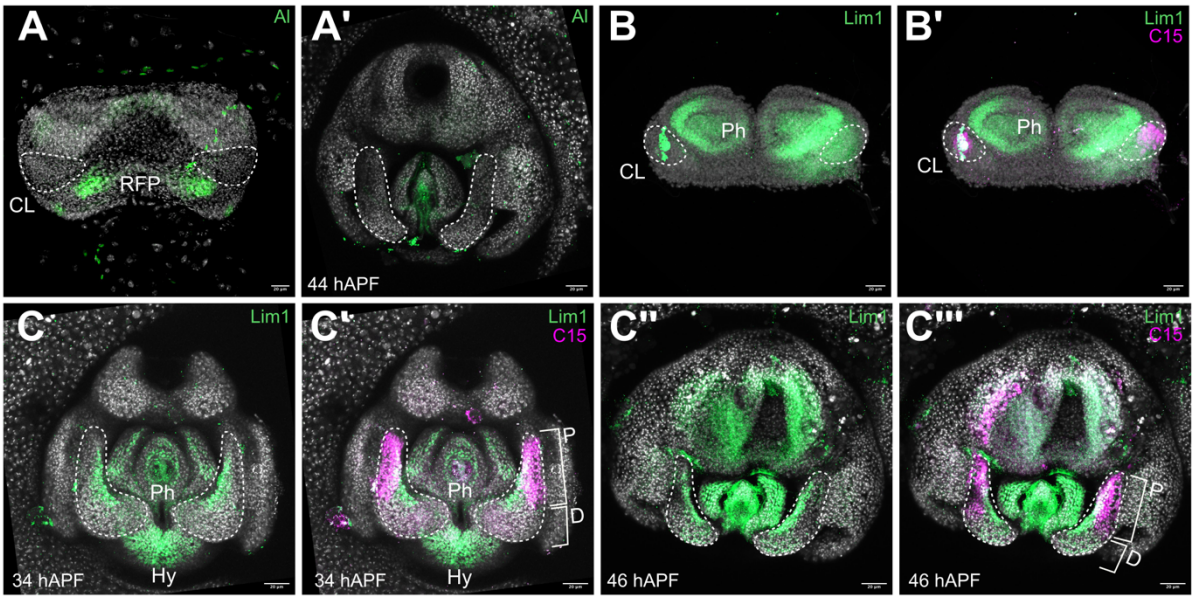


Figure 39: Expression of Al and Lim1 in the developing male genitalia. (A) Al expression is shown within the RFP of the male genital disc. Dashed white outlines indicate where endogenous C15 expression would be. (A') Al expression is absent in all pupal genital structures. (B,B') Lim1 expression is shown within the phallic region of the male genital disc, and does not overlap with C15 expression.

The green Lim1 stain in the left-hand side clasper primordium of the disc is precipitate secondary and not specific to Lim1's expression pattern. (C-C''') Lim1 expression during pupal development is shown within the inner region of the proximal and distal clasper domains where it overlaps with C15 in a restricted region on the proximal domain. Lim1 is also expressed throughout the phallic structures. (A) n=4, (A') n=1, (B,B') n=5, (C,C') n=3, (C'',C''') n=2. All samples have been stained with DAPI. White dashed outlines mark the claspers. CL: Clasper, Ph: Phallus, Hy: Hypandrium, RFP: Repressed female primordium. Distal (D) and proximal (P) marks the two clasper domains.

Lim1 has not previously been characterised in the genitalia and therefore I carried out a double antibody stain against Lim1 and C15 (Figure 39B-C'''). I found that C15 and Lim1 expression did not overlap in the male genital disc (Figure 39B and B'), in contrast to the leg disc and arista of the antenna (Campbell, 2005; Roignant *et al.*, 2010). At 34 hAPF, the clasper and lateral plate primordia are separating, and at this stage Lim1 expression localises to a thin strip of cells at the inner region of the clasper, overlapping slightly with C15 in the proximal domain (Figure 39C and C'), again in contrast to their complete overlap in the pupal leg (Supplementary Figure 18A and C). In older pupae, Lim1 expression stretches further into the distal clasper domain retaining some overlap with C15 in the proximal domain (Figure 39C'' and C'''). Interestingly, the pupal expression pattern of Lim1 marks the region of the genitalia that gives rise to the leg upon complete loss of *Abd-B* (Estrada and Sánchez-Herrero, 2001). Lim1 exhibits phallic expression throughout metamorphosis, yet the clasper expression in the pupa seems to only appear after invagination of the genital disc (Figure 39B-C''').

The absence of Al clasper expression, and minimal overlap between C15 and Lim1 differs from the pretarsal segment of the leg. I next aimed to functionally test the roles of Al and Lim1 during male genital development.

5.2.6 Al and Lim1 function during male pupal genital development

RNAi knockdown of *al* in the claspers showed no obvious phenotype when compared to both parental lines (Figure 40A), consistent with Al expression patterns (Figure 39A and A'). I overexpressed *al* in the male terminalia using the same GAL4 driver line, but resultant progeny did not survive post-pupation.

lim1 RNAi knockdown resulted in 15% fewer bristles in the distal clasper domain compared to the average of both parental lines (Figure 40B; Supplementary Figure 16A and E). Interestingly, *lim1* knockdown also generated significantly larger posterior lobes (Figure 40C and Supplementary Table 7). These *lim1* knockdown phenotypes are the opposite of what I

saw with *C15* RNAi knockdown (Figure 38A and B). Both *lim1* RNAi knockdown results were not consistent with Lim1 expression patterns (Figure 39B-C'''; Figure 40B and C), which could suggest a nonautonomous mode of regulation.

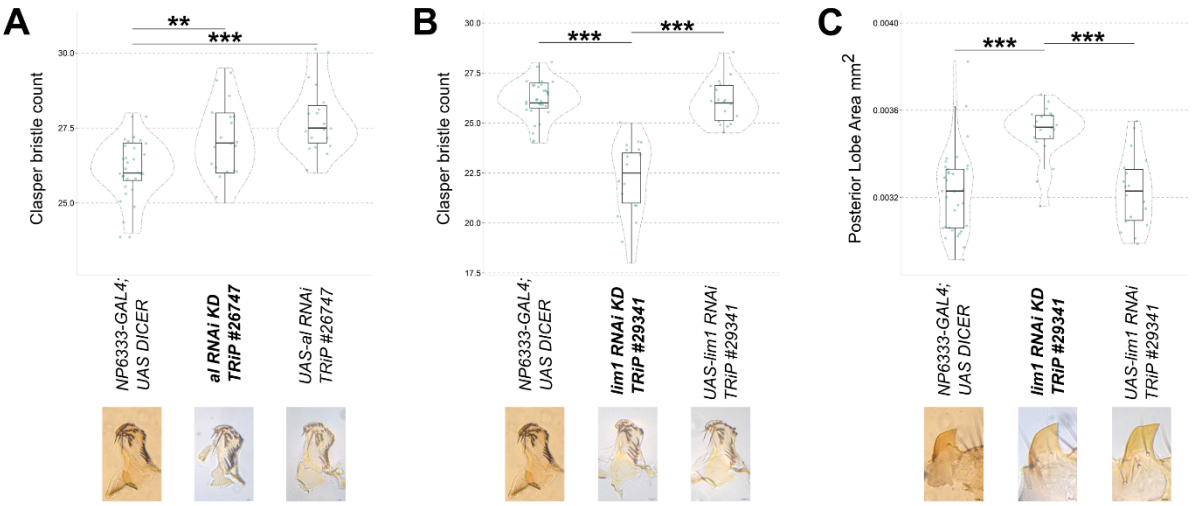


Figure 40: RNAi knockdown of *al* and *lim1* in *D. melanogaster* male claspers and posterior lobes. (A) *al* RNAi knockdown did not affect clasper morphology, (B) whereas *lim1* RNAi knockdown significantly reduced the bristle number. (C) *lim1* RNAi knockdown also significantly increased the size of the posterior lobes. The lines adjoining the RNAi knockdown to the parental *UAS-RNAi* and *NP6333-GAL4, UAS-DICER* line, indicate the level of statistical significance between each pair. *** $p < 0.001$ and ** $p < 0.01$ (Supplementary Table 7). $n > 14$ for all lines phenotyped.

5.2.7 C15, Lim1 and Al interactions during male clasper development

I next tested for an interaction between *C15* and *lim1,* and *C15* and *al* in the male claspers by carrying out reciprocal RNAi knockdown and antibody stainings, as all three have been shown to interact in the pretarsus of the leg (Campbell, 2005; Kojima, Tsuji and Saigo, 2005). *al* RNAi knockdown during pupal development, as expected, showed no disruption to C15 localisation (Figure 41A-B' and D).

lim1 mutants resulted in a mild reduction of *C15* expression during larval leg development (Campbell, 2005). During early pupal development, RNAi knockdown of *lim1* did not seem to disrupt C15 localisation compared to wild type C15 expression (Figure 41A and C). However, during later pupal development after the clasper has split from the lateral plate, RNAi knockdown of *lim1* revealed a slight expansion of C15 at the inner region of the proximal clasper domain, which does not usually express *C15* (Figure 41D-E').

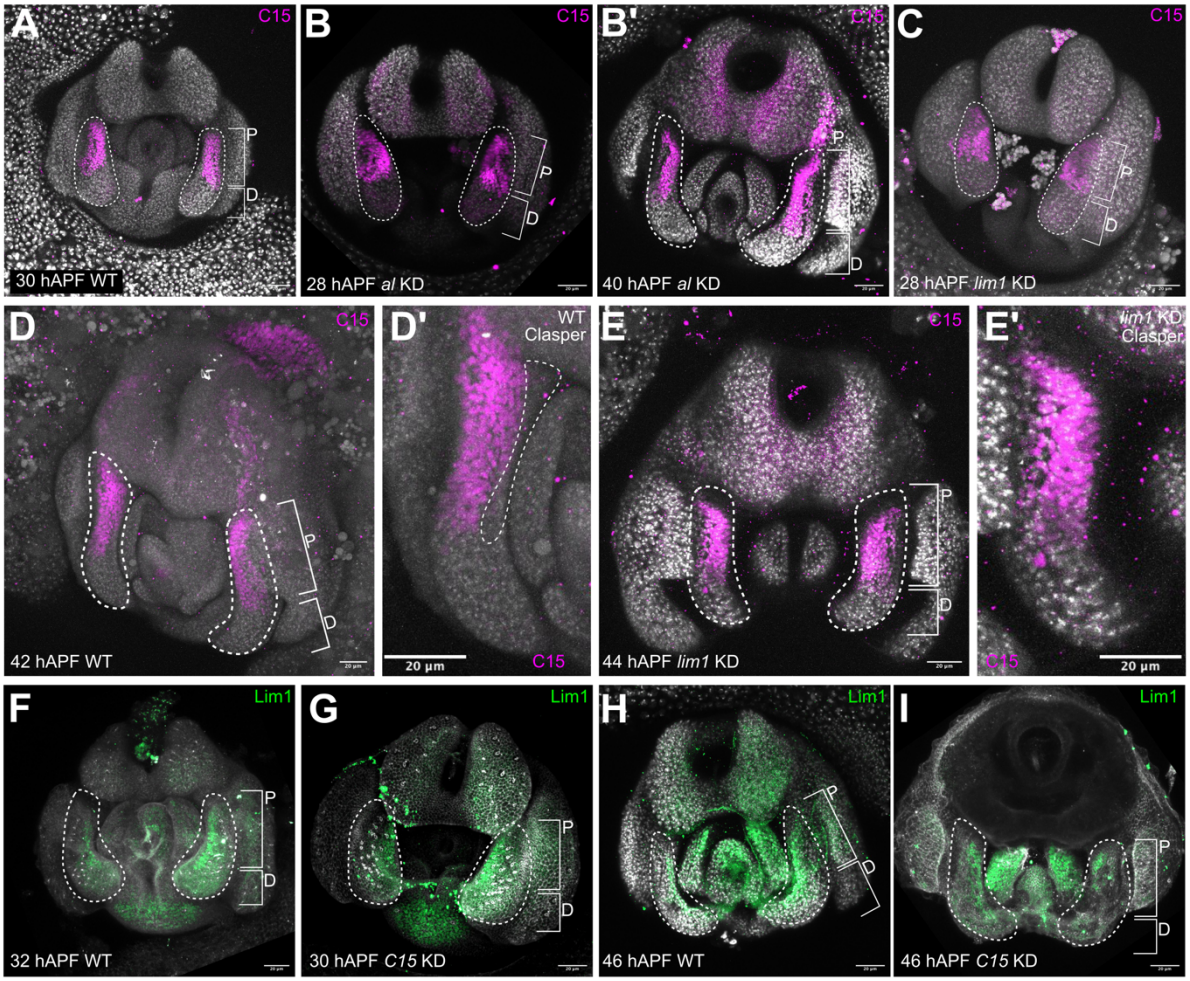


Figure 41: *al*, *lim1* and *C15* RNAi knockdown and reciprocal stainings to test known pretarsal interacting partners during male pupal genitalia development. (A,D,D') Wild-type (WT) C15 expression is observed at the proximal clasper domain. Peripheral cerci expression cannot be seen from this angle. (B,B') *al* RNAi knockdown pupae stained for C15. (C,E,E') *lim1* RNAi knockdown pupae stained for C15. (D') In WT pupae, C15 expression looks to be absent from the inner proximal clasper domain as highlighted by the dashed white outline. (F,H) WT Lim1 expression is observed within the inner region of the clasper across both domains. (G,I) *C15* RNAi knockdown pupae stained for Lim1. All samples were stained with DAPI apart from staining against DCAD2 in (F-I). (B) n=1, (B') n=1, (C) n=2, (E) n=4, (G) n=3, and (I) n=2. P: Proximal and D: Distal clasper domains.

In the leg, *C15* mutants result in the complete loss of *lim1* expression (Campbell, 2005). To understand if C15 may regulate *lim1*, I examined Lim1 protein expression in *C15* RNAi knockdown lines (Figure 41F-I). I observed no visible disruption to Lim1 expression throughout the male genitalia during pupal development (Figure 41G and I).

With respect to the reciprocal *lim1* and *C15* RNAi knockdown antibody stains, slight expansion of C15 protein localisation upon *lim1* RNAi knockdown suggests that Lim1 may inhibit *C15*

expression (Figure 41D-E'), despite minimal expression overlap between the two TFs at the inner region of the proximal clasper domain (Figure 39C'). The claspers of the male terminalia exhibit expression of a similar set of genes to the pretarsal leg GRN but they differ in their relative expression and interactions. I next decided to investigate how other genes required during leg and genital development interact with *C15* in the male claspers, to further decipher clasper GRN interactions and its comparability with the pretarsal leg.

5.2.7.1 C15 interactions with *trn*, *Abd-B* and *Dll*

I next investigated whether RNAi knockdown of *trn* altered C15 localisation (Figure 42A-D) (Hagen *et al.*, 2019). However, no observable alteration to C15 protein localisation was observed in the developing male terminalia or legs upon *trn* RNAi knockdown (Figure 42A-D).

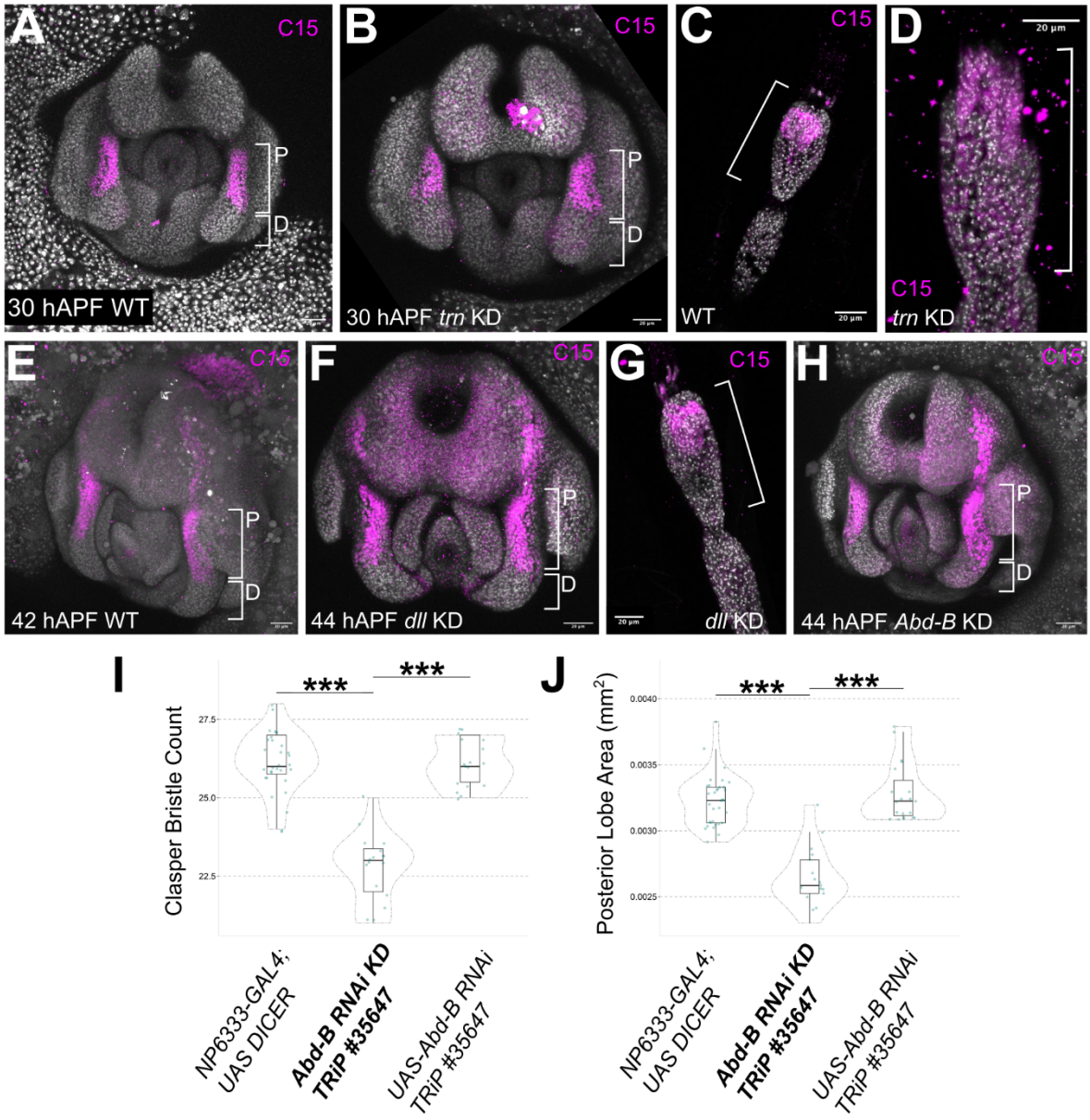


Figure 42: C15 expression upon RNAi knockdown of known male genitalia factors. (A, C and E) WT C15 expression captured by anti-C15 staining in the developing male genitalia and pupal leg. (B) Although *trn* RNAi knockdown is known to decrease clasper size and bristle count, the C15 expression domain remains restricted to the proximal clasper domain (Hagen *et al.*, 2019). (D) C15 expression did not get affected by *trn* RNAi knockdown at the pretarsal segment. (F,G) *Dll* RNAi knockdown did not alter C15 localisation in the developing claspers. (H) Reduction of *Abd-B* expression extended C15 expression from the proximal clasper domain to the distal clasper domain, (I) reduced overall clasper bristle count, (J) and decreased the posterior lobe area (Supplementary Table 7). *** represents the statistical significance level of $p < 0.001$, n>16. P (proximal) and D (distal) clasper domain labelled. (C,D,G) The marked regions indicate the pretarsal segment. (B) n=2, (C) n=2, (D) n=2, (F) n=2, (G) n=1 and (H) n=3.

Similarly, RNAi knockdown of *Dll* did not alter C15 expression in the proximal clasper domain (Figure 42E and F), nor did it affect C15 expression in pupal legs (Figure 42C and G).

I next tested if Abd-B regulated *C15* expression in the developing genitalia. RNAi knockdown of *Abd-B* resulted in the expansion of C15 expression into part of the distal clasper domain, as well as peripheral to the cerci (Figure 42E and H). In addition, *Abd-B* RNAi resulted in a significant decrease in clasper bristles, consistent with described C15 expression patterns changes (Figure 42H and I; Supplementary Table 7). The posterior lobe area also decreased because of *Abd-B* RNAi knockdown (Figure 42J). As RNAi knockdown of *C15* increased the posterior lobe area (Figure 38B), the decreased lobe size observed from *Abd-B* knockdown is also consistent with the observed expansion of C15 expression (Figure 42H and J).

I have shown that the pretarsal factor C15 regulates clasper bristle development. As certain genes and interactions within the pretarsus GRN are not required in the clasper, I next investigated the expression patterns and functions of downstream pretarsal genes to *C15* in the clasper.

5.2.8 The role of Bab2, Bowl and Ap during male clasper development

During leg development, C15 represses *bab2* expression (Figure 34B), which generates a ring of *bab2* expression encircling central *C15* expression (Mojica-Vázquez *et al.*, 2017). I assayed for Bab2 protein expression in fixed male pupal terminalia using anti-GFP in a GFP-tagged *bab2* line (Kudron *et al.*, 2018), which expressed in the leg disc (Supplementary Figure 19A) (Bourbon *et al.*, 2022). In the male genital disc, there was no overlap of Bab2 and C15 expression (Figure 43A). This spatial exclusivity was maintained during pupal development, where Bab2 was concentrated only at the distal domain of the clasper (Figure 43B; Supplementary Figure 19B and B'). The expression pattern previously described by *ISH* for *bab2* mRNA (Vincent *et al.*, 2019), was consistent with Bab2 protein expression in the cerci and distal domain of the claspers, however, mRNA expression at the base of the posterior lobes was not captured by my antibody stain (Figure 43B and Supplementary Figure 15C). I was unable to conduct RNAi knockdown against *bab2* as the *UAS-RNAi* line was not viable.

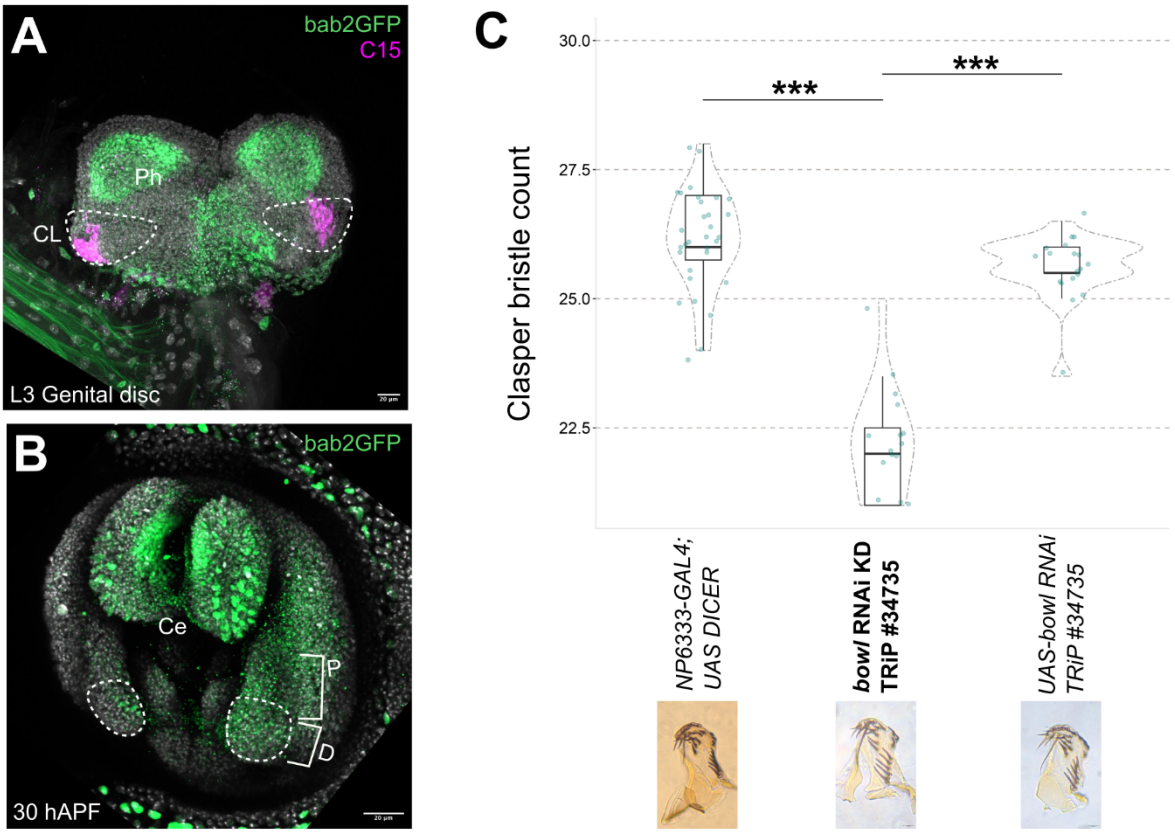


Figure 43: Pretarsal factors Bab2 and *bowl* expression and function during clasper development. (A) Bab2GFP is expressed within the phallic primordium and RFP, lacking overlap with C15 and the clasper primordia (dashed white ovals) (Couderc *et al.*, 2002). (B) The dashed white ovals mark *bab2-GFP* expression in the distal clasper domain. Both samples have been stained with DAPI. Ph: Phallus, CL: Clasper, RFP: Repressed Female Primordium and Ce: Cercus. Proximal (P) and distal (D) clasper domains marked. (A) n=2, (B) n=3. (C) RNAi knockdown of *bowl* significantly reduced clasper bristle count. The lines adjoining the RNAi knockdown to the parental *UAS-RNAi* and *NP6333-GAL4, UAS-DICER* lines indicate the level of statistical significance between each pair. *** $p < 0.001$ (Supplementary Table 7). Within each violin plot, the following values are represented: the median as the bold horizontal line, the box as the interquartile range, and the range as the vertical line. $n > 14$ for all lines phenotyped.

bowl mutants have fused tarsal segments and varying extents of truncation in the legs (Shellenbarger and Mohler, 1978; de Celis Ibeas and Bray, 2003). mRNA *ISH* against *bowl* has previously revealed expression at the distal clasper domain (Vincent *et al.*, 2019). However, I could not stain for Bowl expression in the developing male genitalia because of the high background with Anti-Bowl stains and a GFP-tagged line generated no detectable signal (Kudron *et al.*, 2018). In the clasper, RNAi knockdown of *bowl* decreased bristle count by 14% (Figure 43C). This effect was localised to the distal domain of the clasper (Supplementary Figure 16F), consistent with the previously reported spatial expression of this gene (Vincent *et al.*, 2019). I also overexpressed *bowl* in the male terminalia using the same GAL4 driver

line, but resultant progeny did not survive during pupation. Therefore, both *bab2* and *bowl* are expressed in the distal clasper domain and have no overlap with *C15* expression in the proximal domain, as is also the case during leg development (de Celis Ibeas and Bray, 2003; Mojica-Vázquez *et al.*, 2017).

Previous *ISH* showed *ap* expression at the base of the clasper adjacent to the cercus, as well as within the distal clasper domain (Vincent *et al.*, 2019). I found that *ap* GFP-tagged lines did not generate a detectable signal (Kudron *et al.*, 2018). Furthermore, RNAi knockdown of *ap* during male genital development did not significantly alter the clasper bristle count (Supplementary Table 7).

bab2, bowl and *ap* are all expressed in the clasper during male genital development (Figure 43) (Vincent *et al.*, 2019). From my experiments and previous data, it can be concluded that the clasper is split into two regulatory domains: the distal and proximal domains (Supplementary Figure 16) (Vincent *et al.*, 2019). By comparing the male clasper to the pretarsal leg GRN, I have identified *lim1*, *Dll*, *dac* and *bowl* as TFs required for distal clasper regulation, and *C15* for proximal clasper domain regulation (Supplementary Figure 16). Phenotypic analysis suggests that only *C15* and *Dll* repress bristle development in the clasper (Figure 38A and Figure 36B), whereas *dac*, *lim1* and *bowl* promote it (Figure 36B, Figure 40B and Figure 43C).

5.3 Characterisation of the pretarsal *C15* GRN in the female epiproct

Following characterisation of C15 and associated pretarsal factors in the development of the male claspers, I next replicated this experimental approach in females.

5.3.1 *wg* and *dpp* expression in the female pupal genitalia

I began by assaying where *dpp* and *wg* are expressed in the developing female terminalia. I conducted HCR-IS against *wg* and observed a gradient of expression extending from the ventral tip of both cerci plates (epiproct and hypoproct), as well as the internal ovipositor (Figure 44A). The cerci in female terminalia are rotated 90 degrees with respect to males, and the dorsal plate is the epiproct and the ventral one is the hypoproct (McQueen *et al.*, 2022). *dpp* expression was observed in the epiproct, distal edges of the hypoproct, and dorsal region of the ovipositor (Figure 44B and B'). Therefore, like the male clasper, the female epiproct and ovipositor both show expression of these morphogens (Figure 35 and Figure 44). As this is

known to induce *dac* and *Dll* expression, and thus promote *C15* expression in both the pretarsal segment (Galindo *et al.*, 2002; Estella and Mann, 2008; Mojica-Vázquez *et al.*, 2017; Newcomb *et al.*, 2018), and the male claspers (Figure 35), it seemed plausible that *C15* would also be expressed in the epiproct and the ovipositor.

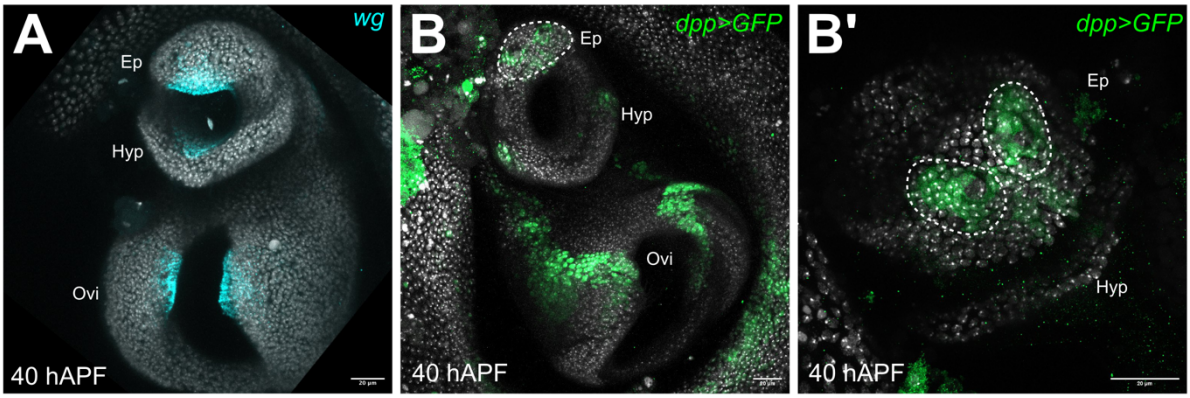


Figure 44: *wg* and *dpp* are both expressed in the female epiproct during pupal genital development. (A) HCR-IS of *wg* in female pupa reveals four origins of the morphogen gradient in the epiproct, hypoproct and ovipositor. (B,B') Anti-GFP staining in *dpp-GAL4>UAS-GFP* fixed female terminalia showed expression around the epiproctal sensillum, as well as the dorsal region of the ovipositor. (B') *dpp* expression surrounds the two prominent epiproctal bristle precursor cells. (A) n=2, (B,B') n=3. All samples have been stained with DAPI. Dashed white ovals mark the epiproctal sensillum. Ep: Epiproct, Hyp: Hypoproct, Ovi: Ovipositor.

5.3.2 Dll regulatory role in the female epiproct

Dll is a key TF coordinating the development of the male cerci and claspers (Figure 36B and Supplementary Table 7), and overlaps in expression with C15 in the latter (Figure 36C and Figure 37). I next sought to validate the function of Dll during female genital development. I used the *NP6333-GAL4* driver to drive *Dll* RNAi knockdown and found that epiproct bristle number was reduced by 72%, and in many cases, the two cerci structures did not separate during development, consistent with previous published data (Figure 45) (Gorfinkiel, Sánchez and Guerrero, 1999). As Dll is required for epiproct development, and *wg* and *dpp* are both expressed in this structure, I next looked to see whether this combination of expressed genes also marked the region of C15 expression as demonstrated in the pretarsus and male claspers.

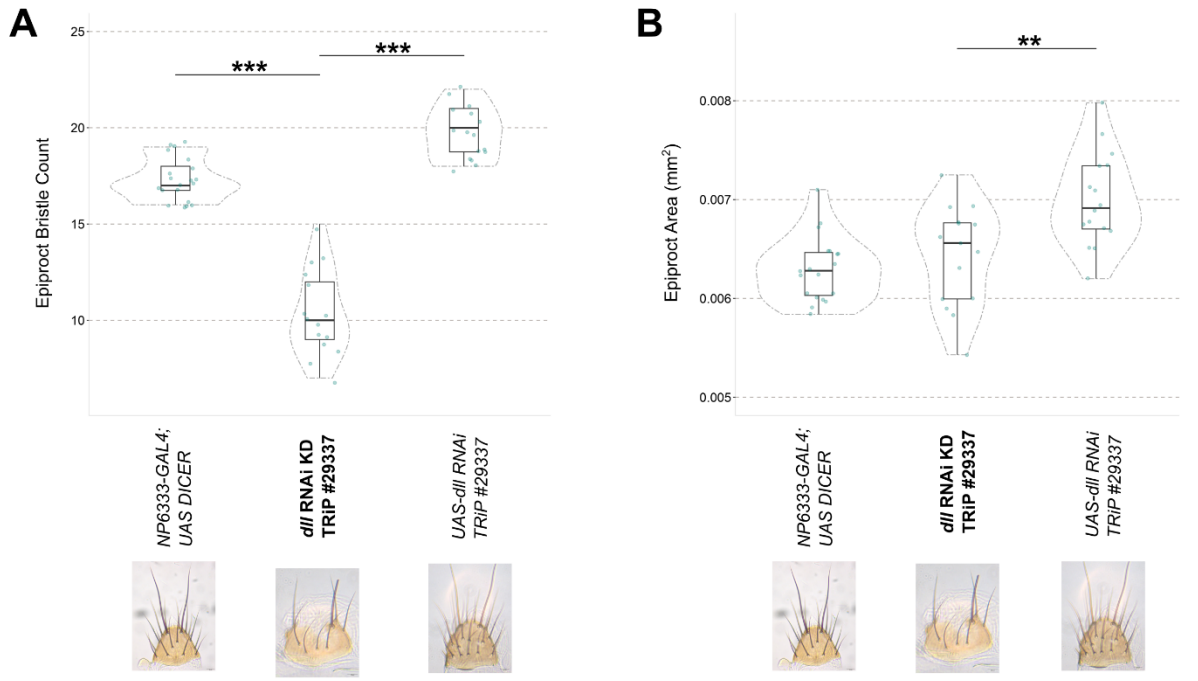


Figure 45: RNAi knockdown of *Dll* in the developing *D. melanogaster* female epiproct. (A) *Dll* RNAi knockdown decreased overall epiproct bristle number, (B) but had no effect on the epiproct area, despite visible missing tissue in the centre. The lines adjoining the RNAi knockdown to the parental *UAS-RNAi* and *NP6333-GAL4, UAS-DICER* lines, indicates the level of statistical significance between each pair. *** $p < 0.001$ and ** $p < 0.01$. $n > 14$ for all lines phenotyped. No significant hypoproct difference between the *Dll* RNAi knockdown and both parental lines were found (Supplementary Table 7).

5.3.3 C15 protein localisation during female terminalia development

In females, I found that C15 is expressed in two well-defined cellular clusters in the genital disc composed of approximately 7 cells each (Figure 46A). This expression pattern is maintained during pupal development and surrounds the two prominent epiproctal sensilla, as identified by the two concentrated DAPI loci central to the C15 rings of expression and latter light microscopy images (Figure 46B and C; Figure 47A).

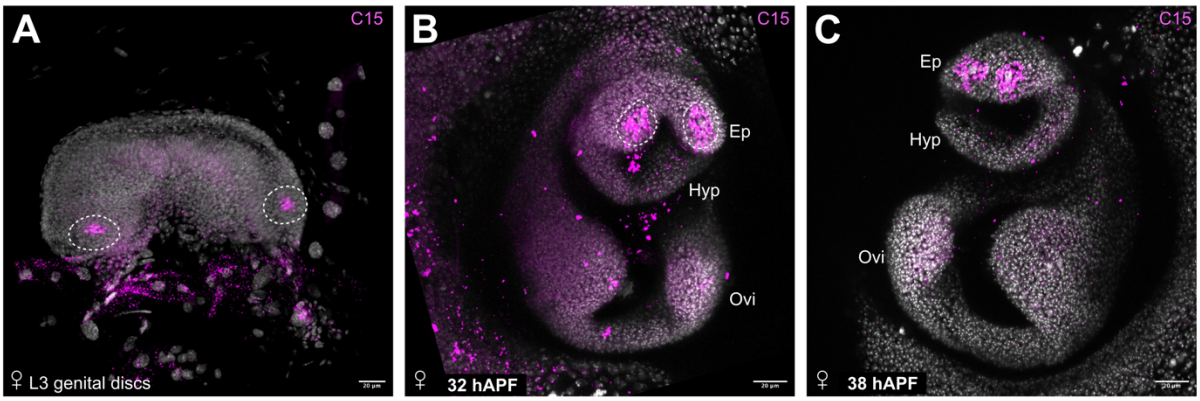


Figure 46: Anti-C15 staining in female genital and cercus structures during development. (A) C15 is expressed within the cercus primordium. (B,C) C15 is also expression in the cercus during pupal development. Expression is localised to the epiproct, encircling the location of the two prominent sensilla that exhibit a longer length and different structure to the other epiproctal sensilla. All stains have been carried out in *D. melanogaster* w^{1118}. (B) Dashed white ovals mark the two prominent epiproctal sensillum. (A) n=5, (B) n=3 and (C) n=3. All samples have been stained with DAPI. Ep: Epiproct, Hyp: Hypoproct, Ovi: Ovipositor.

5.3.4 RNAi knockdown of *C15* in female terminalia

To knockdown *C15* in female terminalia, I used the same GAL4 driver and *UAS-RNAi* line as utilised in males (Figure 38). *C15* RNAi knockdown in females eliminated the two prominent epiproctal sensilla (Figure 47A, A', C and C'). This phenotype is consistent with the roles of *C15* elsewhere in the fly (Kojima, Tsuji and Saigo, 2005), yet differs from the repressive regulatory role found during male clasper development (Figure 38A). However, *C15* RNAi knockdown revealed no significant difference in epiproct bristle count compared to parental controls, although the area of the epiproct did significantly decrease (Figure 47B and B'; Supplementary Table 8).

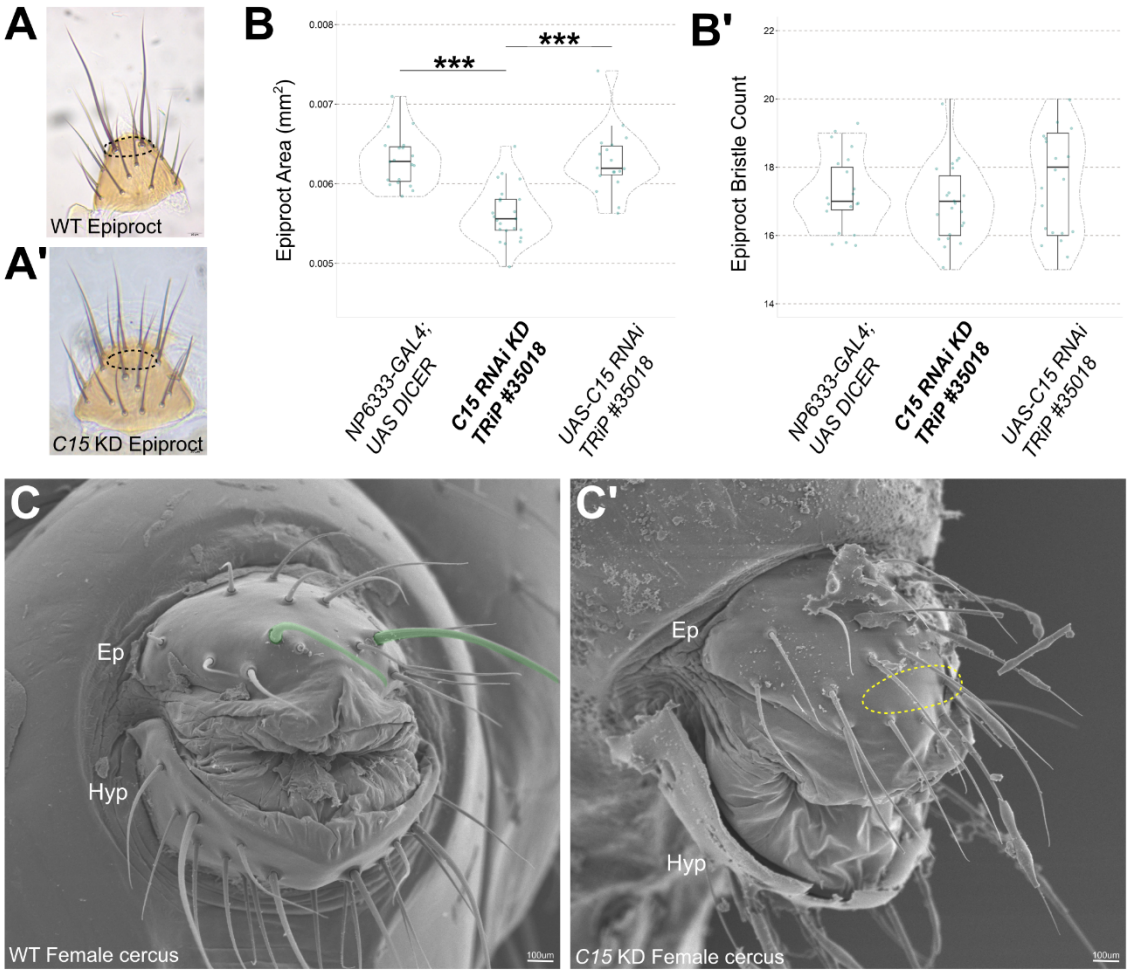


Figure 47: RNAi knockdown of *C15* in female *D. melanogaster* terminalia. Reduction of *C15* expression levels in the developing female terminalia led to the loss of the two longer epiproctal sensilla on the analia. (A, A') Light microscopy images of the epiproct, where the dashed black ovals mark the location of the two longer epiproctal sensilla. Scale bars are 20 μm. (B) RNAi knockdown of *C15* led to a significant reduction of the epiproct area, (B') yet no effect on the overall epiproct bristle count. *** represents the statistical significance level of $p < 0.001$, $n>17$ (Supplementary Table 8). (C) SEM images of *NP6333-GAL4, UAS-DICER* parental control, where the epiproctal sensilla are highlighted in green, and (C') of *C15* RNAi knockdown where the dashed ovals mark the area absent of these structures. Scale bars are 100 μm. Ep: Epiproct and Hyp: Hypoproct.

No effect was seen on the hypoproct or tibia length, or ovipositor bristle patterning in response to *C15* RNAi knockdown, consistent with its expression pattern (Supplementary Figure 20). I also overexpressed *C15* using the *NP6333-GAL4* driver in female terminalia and saw no effect on the epiproct area or bristle number (Supplementary Table 8). These results show an epiproctal-specific role for *C15* in females, where it appears to positively regulate the development of the two prominent epiproctal sensilla, and final epiproctal size.

5.3.5 Al and Lim1 expression during female pupal genital development

I next looked to see whether the overlap in expression of *C15, al* and *lim1* in the pretarsus were replicated in the female epiproct during genital development.

In the female genital disc, Al is not expressed in the epiproctal primordium marked by C15 (Figure 37A and Figure 48A). However, post-metamorphosis, Al is expressed ventral to the two prominent epiproctal sensilla in pupae (Figure 48B and B'). I could not validate overlap in expression between C15 and Al as both antibodies originated from rat. However, both Al and C15 are expressed in the same region of the epiproct, suggesting that in this context these TFs may share more similarity to the pretarsal segment, than the male claspers.

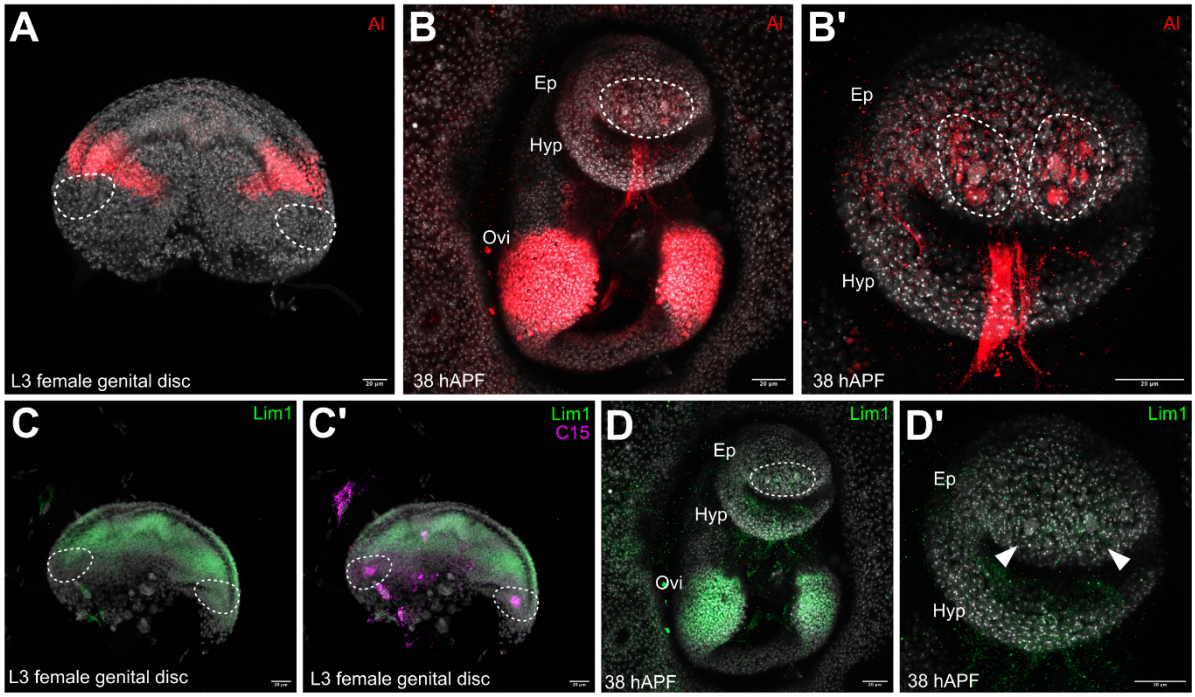


Figure 48: Expression of Al and Lim1 in the developing female genitalia. (A) Al is expressed at a defined region of the hypogynium in the female genital disc (Chatterjee *et al.*, 2011). Dashed white ovals marks the epiproct primordia where Al signal is absent. (B,B') During female pupal development, Al is expressed at the ovipositor and in the epiproct surrounding where the two long epiproctal sensilla protrude from. (C,C') Lim1 is expressed within the hypogynium primordium of the female genital disc, but does not overlap with C15 expression in the cercus primordium. (D,D') Lim1 is expressed in the ovipositor, and is not expressed in the epiproct (white dashed oval) where C15 and Al were shown to be expressed (two white arrows). (A) n=2, (B,B') n=2, (C,C') n=2 and (D,D') n=2. All samples have been stained with DAPI. Ep: Epiproct, Hyp: Hypoproct, Ovi: Ovipositor.

There was no observable Lim1 protein localisation in the epiproct during larval and pupal developmental (Figure 48C-D'), yet Lim1 was expressed in the ovipositor (Figure 48D). Therefore, unlike in the pretarsus and male claspers, Lim1 and C15 show no overlap in expression during female epiproct development (Figure 39C' and Figure 48C').

Although C15 and Al are both expressed in the epiproct, the lack of Lim1 expression does not replicate that seen in the pretarsus.

5.3.6 Al and Lim1 function during female pupal genital development

To understand the role of Lim1 and Al during female genital development, I next characterised their functions using RNAi knockdown.

RNAi knockdown against *al* decreased epiproct bristle count by 8%, yet the two prominent sensilla specified by C15 were still present (Figure 49A). The epiproct area also decreased upon knockdown of *al* expression compared to both parental controls (Figure 49B), consistent with its expression pattern (Figure 48B). In addition, *al* RNAi knockdown increased the hypoproct bristle number, despite no protein localisation being observed in this structure during development (Figure 48B and Supplementary Table 8). As Al was shown to be expressed in the ovipositor, I also quantified ovipositor bristle number in the RNAi knockdown and the *UAS-RNAi* parental lines but found no significant difference (Supplementary Table 8). I overexpressed *al* in the female terminalia using the same GAL4 driver line, but resultant progeny did not survive during pupation.

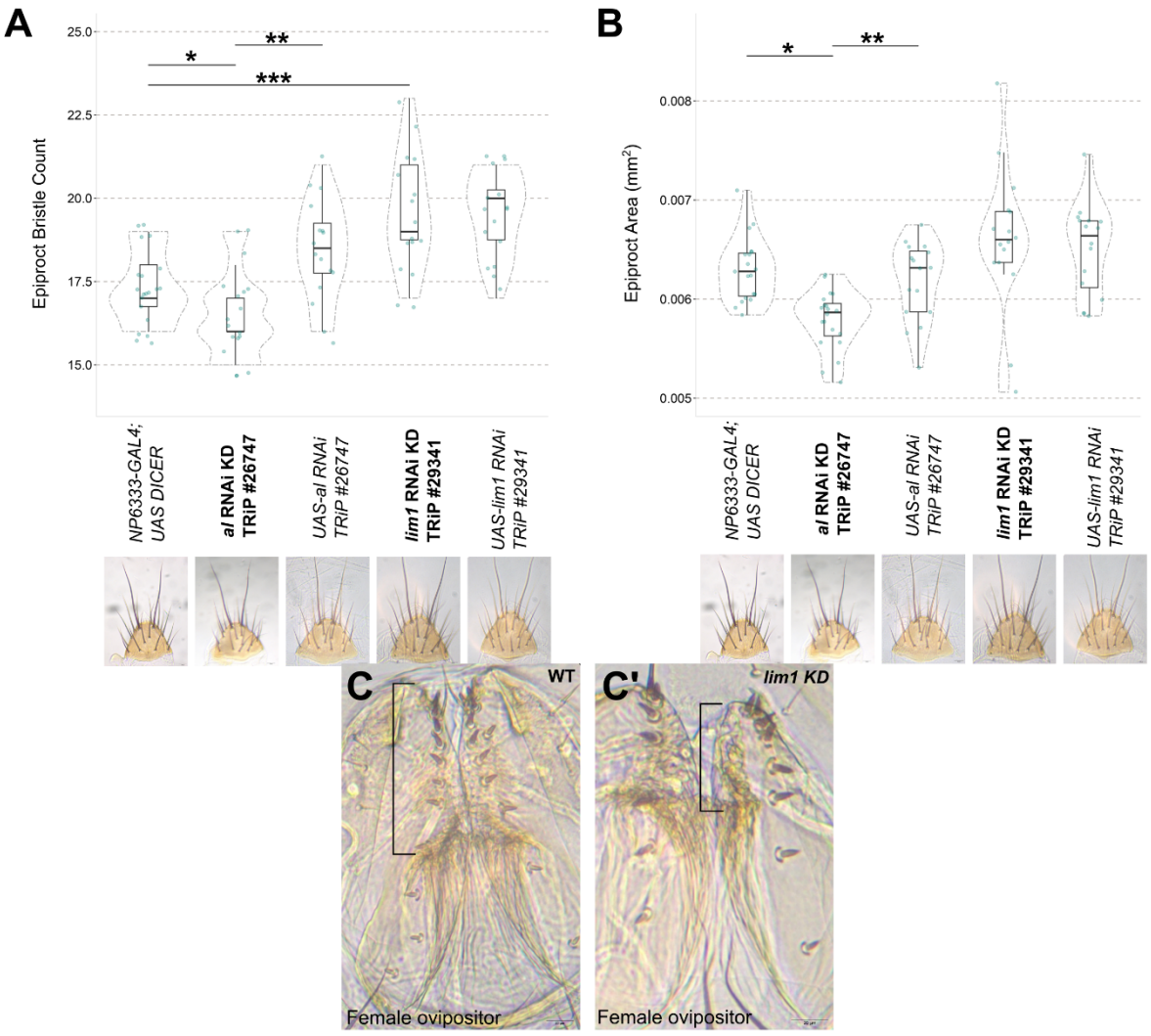


Figure 49: RNAi knockdown of *al* and *lim1* in female *D. melanogaster* terminalia. (A,B) *al* RNAi knockdown reduced overall epiproct bristle number, and area. *lim1* RNAi knockdown had no effect on the epiproct bristle count or area. The lines adjoining the RNAi knockdown to the parental *UAS-RNAi* and *NP6333-GAL4, UAS-DICER* lines, indicate the level of statistical significance between each pair. *** p < 0.001, ** p < 0.01, * p < 0.05 (Supplementary Table 8). n > 14 for all lines phenotyped. (C,C') Light microscopy images of the female ovipositor in (C) WT and (C') *lim1* RNAi knockdown, where bristle number has been reduced in the region marked by the extending enclosed lines (Supplementary Table 8).

lim1 RNAi knockdown in females had no effect on epiproct bristle or area compared to both parental controls (Figure 49A and B). However, *lim1* RNAi knockdown did produce an observable decrease in ovipositor bristles from approximately 26 to 18, when compared to the *UAS-RNAi* parental control (Figure 49C and C'; Supplementary Table 8).

In summary, C15 and Al both positively regulate epiproct size and bristle number during female genital development.

5.3.7 C15 and Al interaction during female epiproct development

As Al and C15 were shown to localise to similar regions of the epiproct, surrounding the two prominent sensilla, I next sought to see whether the two TFs interact with one another. In the pretarsal region of the leg disc, mutant *C15* or *al* alleles lead to the near complete loss of one another (Campbell, 2005). To interrogate whether this positive feedback loop exists in the female epiproct, I conducted antibody staining against Al in *C15* RNAi knockdowns, and vice versa (Figure 50).

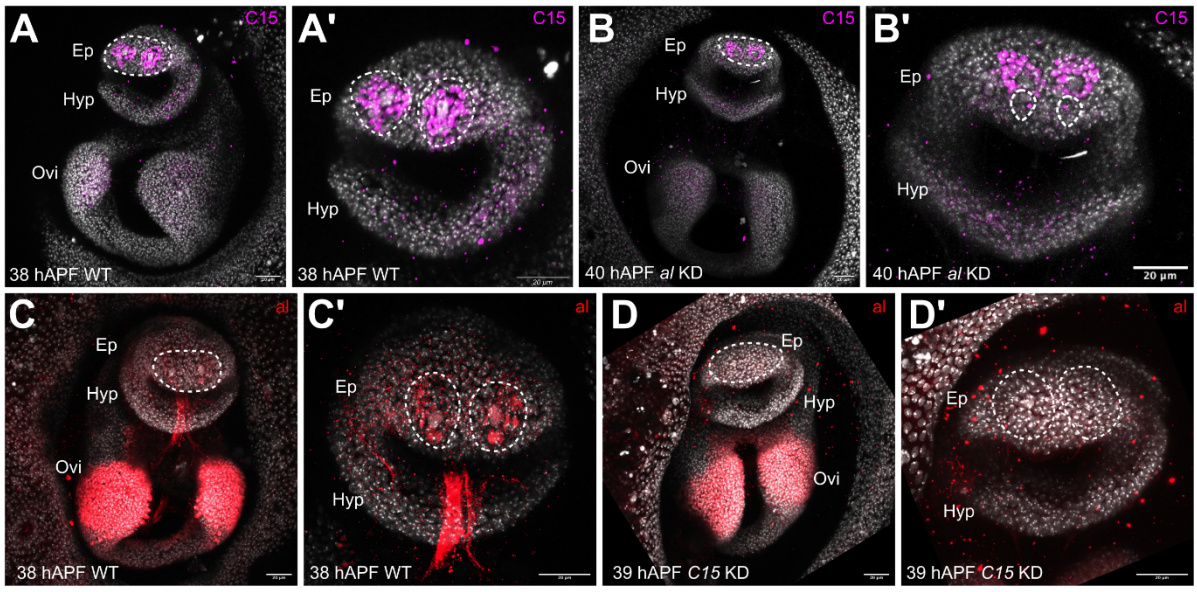


Figure 50: RNAi knockdown of *al* and *C15* in the female developing terminalia alters the expression of C15 and Al, respectively. (A,A') WT C15 expression surrounds the two longer epiproctal sensilla. (C,C') WT Al expression is ventral to the epiproctal sensilla. (B,B') RNAi knockdown (KD) of *al* eliminates C15 ventral to the epiproctal sensilla, as highlighted by the dotted white ovals in B'. (D,D') RNAi KD of *C15* completely eliminates Al in the entire epiproct, as highlighted by the dashed white ovals in D'. All samples have been stained with DAPI. (B) n=2 and (D) n=5. Ep: Epiproct, Hyp: Hypoproct, Ovi: Ovipositor.

Knockdown of *al* expression resulted in loss of C15 in the cells ventral to the epiproctal sensilla (Figure 50A-B'), consistent with where they overlap in expression (Figure 50A' and C'). Consistent with this, RNAi knockdown of *C15* completely abolished Al expression in the epiproct (Figure 50C-D'). Therefore, this confirms there is positive regulation between *al* and *C15* in the female epiproct (Campbell, 2005; Kojima, Tsuji and Saigo, 2005; Miyazono *et al.*, 2010).

5.4 Regulation of *C15* expression in the male claspers

I next investigated whether *C15* is regulated in the leg and genital structures by the same enhancer(s). To date, the enhancer(s) regulating *C15* expression in the pretarsal segment are not known.

5.4.1 ATAC-seq comparisons of the *C15* locus in leg and male genital samples

I chose to screen for putative enhancers within the *C15* locus, defined by the region between the two adjacent neighbouring genes, *CG7956* and *Fancm* (Figure 51). I used T1 and T2 larval leg disc ATAC-seq data from *D. melanogaster*, and male pupal genital ATAC-seq data from *D. simulans* and *D. mauritiana* (Figure 51) (Feng et al., 2022 and Hagen et al., unpublished). At the time of writing, *D. melanogaster* male genital ATAC-seq data was unavailable, however, use of *D. simulans* and *D. mauritiana* tracks instead should not hinder enhancer discovery as antibody stains against C15 in both species developing male terminalia showed no observable expression differences to *D. melanogaster* (Supplementary Figure 21). Note, female genital ATAC-seq data was also unavailable at the time of writing.

In all four ATAC-seq tracks, the largest accessible peak located 5' of *C15* is the transcriptional start site (Figure 51) (Hagen et al., 2021; Feng et al., 2022). Comparisons between leg and genital ATAC-Seq data identified four and three tissue specific peaks, respectively (Figure 51). The leg-specific peak closest to *Fancm* was only present in T1 leg disc data (Figure 51). The genital-specific peak at the third *C15* intron is accessible only in *D. simulans* and not in *D. mauritiana,* suggesting a possible interspecies *C15* regulatory difference (Figure 51).

Three *C15* enhancer reporter lines, spanning approximately 3 kb regions of *C15* intron one and two, were available (Figure 51) (Jenett *et al.*, 2012). The three enhancer reporter lines are referred to as intron 1 enhancer, intron 2 enhancer 1, and intron 2 enhancer 2 (Figure 51). Prior to screening the reporter lines, I compared between the leg and genital chromatin environments to help predict tissue-specific and shared enhancer activity (Figure 51). The intron 1 enhancer spanned inaccessible chromatin in both leg and genital samples and was therefore predicted to not be active (Figure 51). Intron 2 enhancer 1 contains two well-defined peaks in both genital and the T1 leg disc track (Figure 51) that may mark active cis-regulatory regions capable of driving *C15* expression. Intron 2 enhancer 2, which overlaps with intron 2 enhancer 1 by 1007 bp, spans inaccessible chromatin in both tissues and therefore should

show no activity (Figure 51). Interestingly, intron 2 enhancer 2 contains the previously mapped embryonic enhancer of *C15* (Lin *et al.*, 2006).

I screened the enhancer-GAL4 reporter lines by crossing them to *UAS-GFP*, to visualise enhancer activity in genitalia and legs during larval and pupal development.

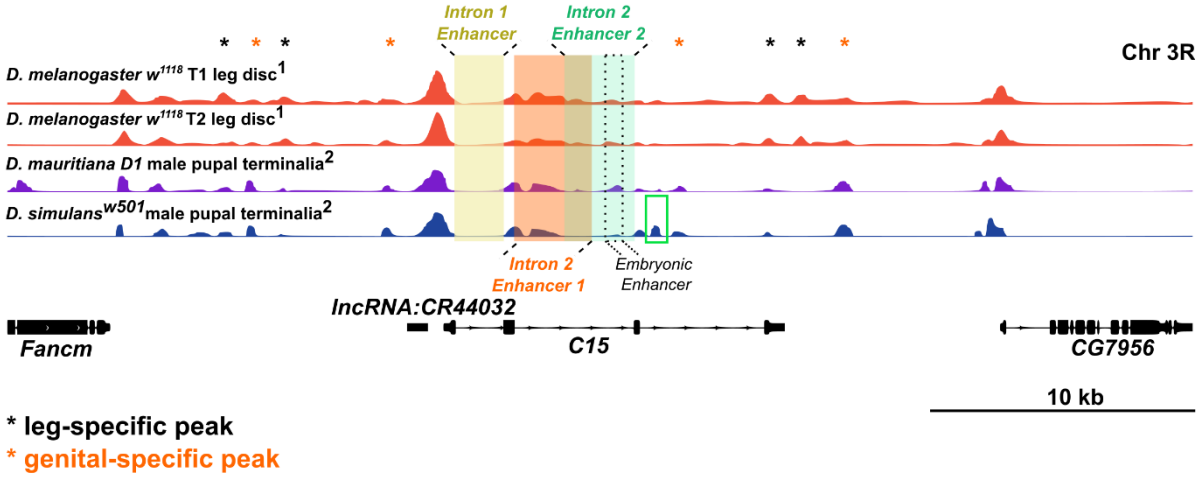


Figure 51: ATAC-seq for *C15* locus in larval leg discs and male pupal terminalia tissue. [1]ATAC-seq of T1 and T2 leg discs from *D. melanogaster* w^{1118} generated by (Feng *et al.*, 2022). [2]ATAC-seq of male pupal terminalia at 30 – 36 hAPF generated from *D. simulans*^{w501} and *D. mauritiana* D1 by Hagen *et al., (unpublished)*. Coloured boxes mark available GAL4 enhancer driver lines (Jenett *et al.*, 2012). The dashed box marks the known embryonic *C15* enhancer (Lin *et al.*, 2006). *D. mauritiana* and *D. simulans* tracks were flipped to align with the positive orientation of *D. melanogaster*. Black asterisks are regions of accessible chromatin biased to leg discs, whereas orange asterisks represent male genital-biased regions of accessibility. The green box present only in the *D. simulans* male terminalia and not *D. mauritiana* highlights a region of accessible chromatin specific to just this species and tissue.

5.4.2 *C15* Intron Enhancer screen in male and female live pupae

I first screened activity of the reporter lines crossed to *UAS-GFP* in live pupae. *C15* intron 1 enhancer was not active in leg or genital pupal structures as predicted from the lack of accessibility observed in the ATAC-seq profiles (Figure 51; Figure 52A and A'). The segmental pattern captured across the pupal abdomen for the intron 1 region activity is similar to pigmentation gene expression patterns in this tissue (Figure 52) (Camino *et al.*, 2015). In larvae, the leg and female genital discs had no observable intron 1 enhancer activity, whereas the male genital disc revealed activity in the RFP (Supplementary Figure 22A-C'), which does not recapitulate endogenous *C15* expression.

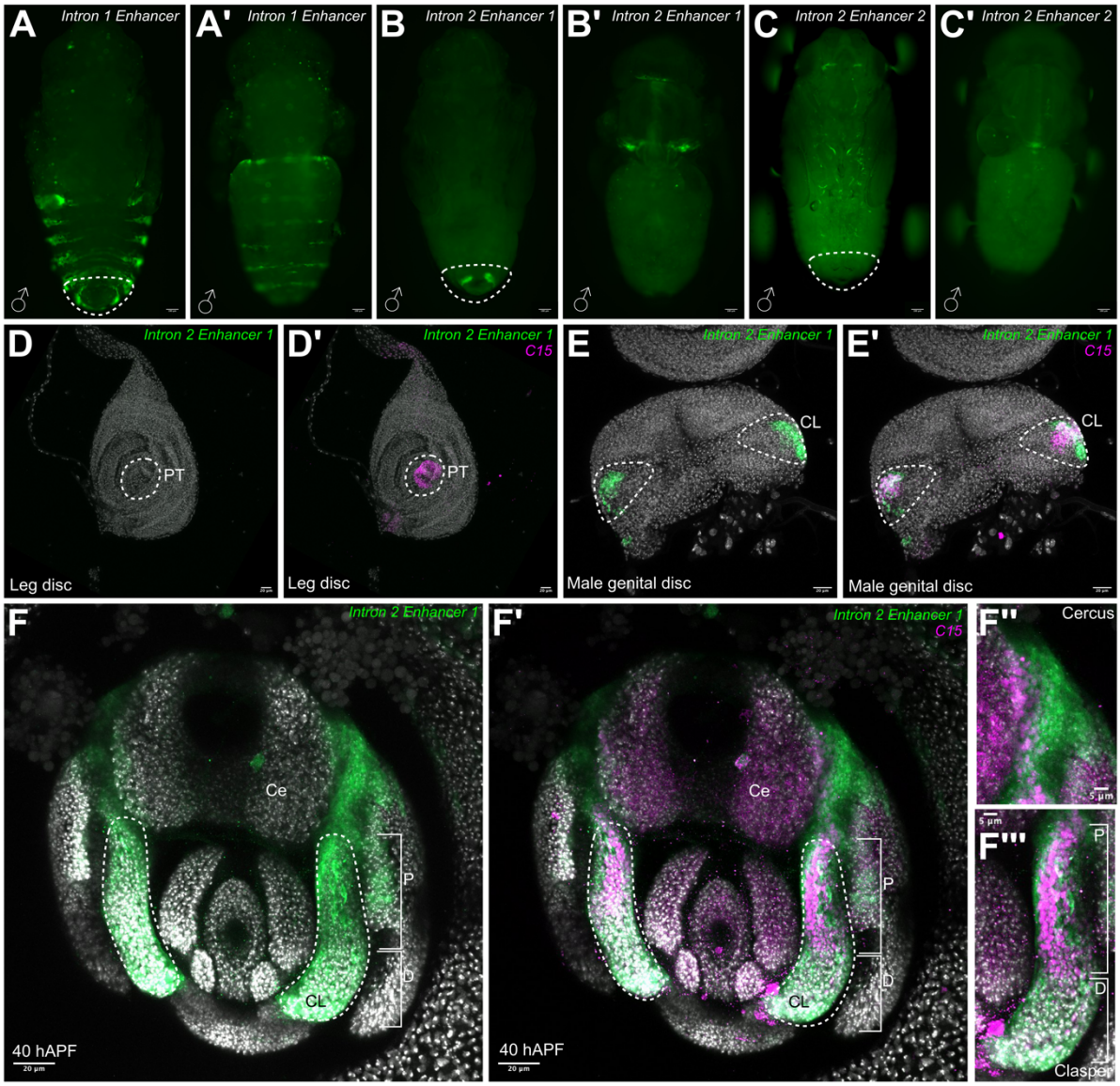


Figure 52: C15 Intron enhancer reporter line activities in live and fixed pupae structures. (A-C') All three *C15* intron reporter lines revealed activity throughout live pupae. Male terminalia is marked by white dashed ovals. (D-F''') Intron 2 enhancer 1 GAL4 line driving *UAS-GFP* was screened for activity in fixed larval and pupal structures. (D,D') This enhancer was not active in the leg disc. (E,E') This enhancer was active in male genital discs and overlapped with endogenous *C15* expression at the clasper primordia. (F) This enhancer was active in the claspers and periphery of the cerci at 40 hAPF, (F') and is shown to overlap with endogenous C15 expression. (F'',F''') Zoomed in images of the overlap between C15 expression and Intron 2 enhancer 1 activity around the cercus and in the proximal clasper domains. (A) n=3, (B) n=1, (C) n=3, (D) n=7, (E) n=4, and (F) n=4. All samples have been stained with DAPI. PT: Pretarsus, CL: Clasper, Ce: Cercus, P: Proximal clasper domain, D: Distal clasper domain.

Initial screening of intron 2 enhancer 1 in live pupae revealed activity in only the male terminalia and not in the female counterpart (Figure 52B and B'; Supplementary Figure 23A-B). Upon closer inspection, this enhancer was active in the clasper primordia of male genital

discs and recapitulated endogenous *C15* expression, which was sustained during late pupal development (Figure 52E-F'''). However, this enhancer was active in the distal clasper domain despite *C15* not being expressed there (Figure 52F'''). Although this enhancer is active in the male genitalia, it did not drive expression in leg discs (Figure 52D and D'). Alongside male genital activity, intron 2 enhancer 1 was also shown to be active in the ocellar primordium of the eye disc and notum in the wing disc (Supplementary Figure 23C-D'').

In live pupae, intron 2 enhancer 2 showed activity at the antennal structures and above the dorsal abdomen but lacked activity in leg and genital tissues (Figure 52C and C'), also seen in larval samples (Supplementary Figure 22D-E'). However, this enhancer was active in the notum of the wing which overlapped with endogenous C15 expression (Supplementary Figure 22F) (Del Signore, Hayashi and Hatini, 2012).

In summary, *C15* expression in the pretarsus is not regulated by intron 1 or intron 2 enhancers (Figure 52D; Supplementary Figure 22A and D). However, in the male genitalia *C15* expression is coordinated partly by intron 2 enhancer 1 (Figure 52E-F'''). Therefore, this screening has confirmed that *C15* is regulated by different enhancer(s) in the male genitalia compared to the pretarsus.

5.4.3 Motif mapping of genital expressed TFs in *C15* intron 2 enhancer 1

I next investigated which TFs may bind to the *C15* intron 2 enhancer 1 during genital development. I first selected for TFs only expressed in the genitalia and not legs, as the observed differences in enhancer activity may arise from which activating TFs are present to bind the sequence and thus drive expression (Figure 52 and Supplementary Figure 22) (Zeitlinger, 2020).

To do so, I used RNA-seq data generated from T1 leg discs and identified 7,162 genes from 17,559 that were not expressed (Feng *et al.*, 2022). I then filtered for genes encoding TFs (Hagen et al., 2021), to reveal 108 TFs that are not expressed during larval T1 leg development (Feng *et al.*, 2022). I repeated this approach using RNA-seq data generated from T2 and T3 leg discs (Feng *et al.*, 2022), as well as pupal legs (Kittelmann *et al.*, 2018), to ensure that the 108 TFs not expressed in T1 leg discs were also not expressed in other legs and later in development. Finally, I examined whether the 108 unexpressed leg TFs are expressed during genital development (Hagen et al., 2021). This filtering generated nine TFs for motif mapping (Supplementary Table 9 and Supplementary Table 10); *cad, eve, hb, Sry-*

beta, bap, byn, otp, HLH54F and *HGTX*. Upon comparison to the *Drosophila* terminalia *ISH* database, I chose to not analyse *cad* any further as its expression pattern was localised to the cerci, with no visible expression in the claspers (Vincent *et al.*, 2019).

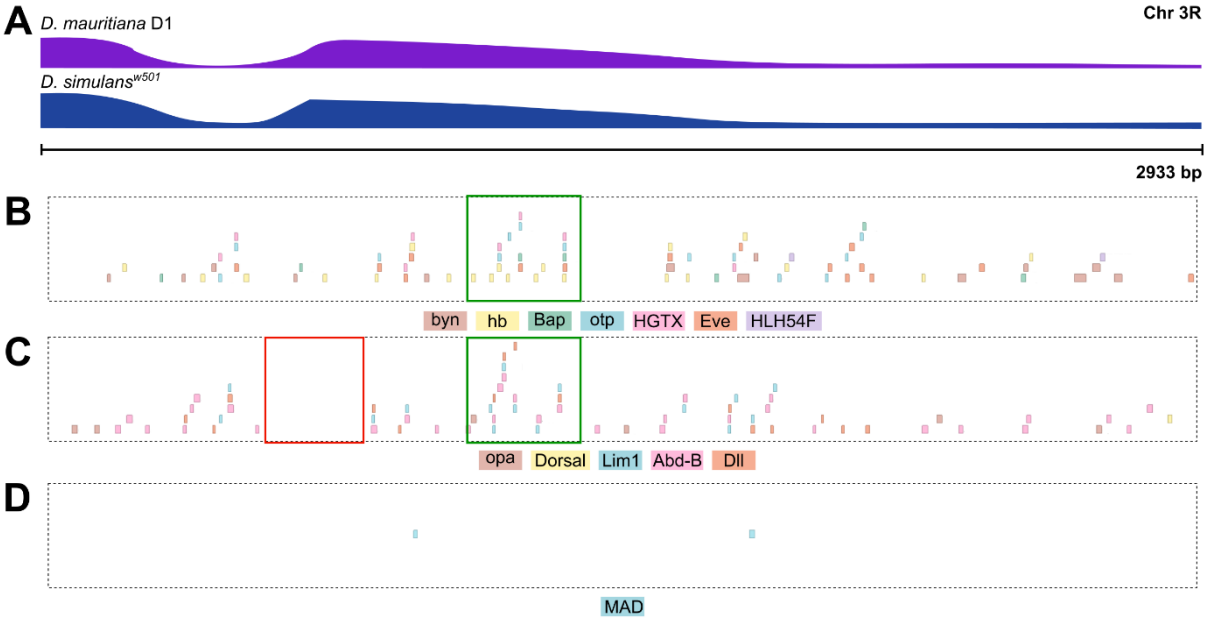


Figure 53: Mapped TF binding sites within *C15* Intron 2 Enhancer 1. Binding sites of predicted TFs regulating *C15* intron 2 enhancer 1 activity (Supplementary Table 10). (A) A magnified schematic of *D. mauritiana* and *D. simulans C15* Intron 2 Enhancer 1 ATAC-seq data from developing male terminalia (Hagen *et al., unpublished*). (B) Mapped binding sites using TFs only expressed in the developing genitalia, that are also not expressed in the developing leg. Note, *eve* expression has been reported before in the adult clasper (Gorfinkiel, Sánchez and Guerrero, 1999). (C) Mapped binding sites using TFs predicted to regulate *C15* in the male genitalia in accordance with functional and spatial expression data presented within this chapter. The red box marks the region absent of any mapped TF binding sites. (B,C) The green box marks the most clustered TF binding site region. (D) Mapped MAD binding sites to compare to *C15*'s regulation in the embryo (Lin *et al.*, 2006). All mapped sites were filtered for a p-value of < 0.001, and binding motifs were collected from the CIS-BP database (Weirauch *et al.*, 2014).

I used the compiled PWMs for each TF to scan intron 2 enhancer 1 for significantly mapped binding sites (Supplementary Figure 24) (Weirauch *et al.*, 2014). All TFs apart from Sry-beta revealed motifs within this enhancer (Figure 53B), where Otp and HGTX would repeatedly cluster together due to high similarity in binding motifs (Figure 53B and Supplementary Figure 24). Although motifs are mapped throughout the enhancer, a prominent cluster of potential TF binding sites are located within the second ATAC-seq peak (Figure 53A and B). This approach has predicted a different trans-regulation, as well as cis regulation, of *C15* in the male genitalia compared to the legs.

5.4.4 Motif mapping of identified clasper-acting TFs in *C15* intron 2 enhancer 1

Dll, Lim1 and Abd-B were previously shown to regulate clasper development and overlap with C15 expression (Figure 36B and C, Figure 39C', Figure 40B, Figure 42H and I, and Supplementary Figure 9A). In addition, Opa revealed clasper-specific expression in the developing male genitalia (Figure 37A'). I next sought to test whether binding sites for these four TFs were predicted within intron 2 enhancer 1 and thus directly regulate *C15* clasper expression.

Mapped TF binding sites spanned the entire enhancer, where Lim1 and Abd-B both mapped to the same thirteen motifs (Figure 53C). A prominent cluster of binding sites were mapped to the same position as described in Section 5.4.3, indicating that this sequence may be important for enhancer regulation (Figure 53B and C). Interestingly, the sequence aligned with the ATAC-seq 'dip' between the two peaks did not map to any TF binding sites (Figure 53A and C). This complementary approach has mapped four upstream TFs that could also be regulating *C15* expression at intron 2 enhancer 1.

5.4.4.1 MAD motif mapping within intron 2 enhancer 1 of *C15*

The only known study that sought to understand how *C15* expression is regulated during *Drosophila* development was during embryogenesis (Lin *et al.*, 2006). The enhancer regulating *C15* expression during embryogenesis has been annotated within intron 2, consisting of multiple MAD and Zen binding sites that modulate *C15* expression in response to the Dpp gradient (Lin *et al.*, 2006). Interestingly, I showed previously that C15 and Dpp were both expressed in the proximal clasper domain (Figure 35B and Figure 37). Upon inspection of *zen* and *MAD* expression in the developing male genitalia, only the latter was shown to be expressed (Hagen et al., 2021). In addition, *MAD* has been shown to be expressed in the proximal clasper domain like C15 (Figure 37) (Vincent *et al.*, 2019). Therefore, I next looked to see if embryonic MAD regulation of *C15* is replicated during genital development. I found two MAD binding sites within intron 2 enhancer 1, whereby only one aligned to accessible chromatin at the second peak (Figure 53A and D). These results indicate that *C15* may be regulated in response to the Dpp threshold via MAD as seen during embryogenesis (Lin *et al.*, 2006).

5.5 Chapter Five Discussion

In this chapter I characterised the pretarsal *C15* GRN similarities and differences during male clasper and female epiproct development. I was able to identify genes and interactions that were conserved between all three developmental contexts and importantly those that differed. The differences predominantly stemmed from *C15*, *al* and *lim1.* I also identified a *C15* enhancer that drives *C15* expression in the developing male genitalia but not the legs suggesting network differences may stem from its regulation. In addition, this comparative assessment revealed sex-specific modifications to this network within the cerci.

5.5.1 Pretarsal genes and their deployment compared to male genitalia

Upon investigation of the pretarsal GRN deployment during male clasper development, I found that upstream expression patterns of morphogens and patterning factors relative to *C15* were remarkably similar (Figure 35 and Figure 36). This was also the case for *bowl*, *bab2* and *ap* downstream of *C15* (Figure 43 and Supplementary Table 7). However, absence of al clasper expression, and minimal overlap between C15 and Lim1 revealed key network differences between different developmental contexts (Figure 39). *svb* GRN comparisons between larval and adult leg trichome patterning showed that the developmental context can greatly impact which nodes are open to evolution and thus drive phenotypic divergence between species (Kittelmann *et al.*, 2018). To expand this analysis, it would be insightful to depict the architecture of the pretarsal GRN during antennal arista and ocellar bristle development too, and identify which genes and interactions are conserved and those that are not. This would generate candidates for putative hotspot genes underlying phenotypic divergence between male claspers, for example.

5.5.2 C15 and Lim1 boundary formation in male genital structures

RNAi knockdown of *C15* generated a significant decrease in the posterior lobe size (Figure 38B) and increase in clasper size as insinuated from the increase in bristles (Figure 38A) (Hagen 2019). In contrast, *lim1* RNAi knockdown increased lobe size and decreased clasper bristle count (Figure 40B and C; Supplementary Table 7). The opposing phenotypic changes of the posterior lobe and clasper from *C15* and *lim1* RNAi knockdown insinuates that these TFs may regulate the specification of the boundary determining the relative split of these two structures during early pupal development.

Boundary specification between the clasper and lateral plate prior to the split has been previously associated with Trn and Ten-M (Hagen et al., 2019, Hood *et al.,* unpublished). Both membrane proteins encircle C15 clasper expression (Hagen et al., 2019, Hood *et al.,* unpublished), where Trn also encircles pretarsal C15 expression in the leg disc (Sakurai *et al.*, 2007). *C15* RNAi knockdown in the leg significantly disrupted accumulation of the Trn and Ten-M interacting scaffold protein, Par-3 (Sakurai et al., 2007), inferring that C15 may regulate their gene expressions. It could be thought that C15 regulates the same genes in the clasper, and ultimately determines the separation between the clasper and the lateral plate. This could be tested for by staining against these three factors in *C15* RNAi knockdown lines to see whether their expression patterns are disrupted. To complement this, it would be insightful to generate RNA-seq data from the male terminalia of control versus *C15* knockdown lines to assess whether *trn*, *ten-m* and/or *par-3* expression is significantly altered.

5.5.3 *C15* intron 2 enhancer 1

Intron 2 enhancer 1 activity overlapped with endogenous *C15* expression in the male genitalia, wing notum and ocelli (Figure 52E-F''' and Supplementary Figure 23C-D''), yet showed no activity in the leg disc or female epiproct (Figure 52D and Supplementary Figure 23B). In these three structures, enhancer activity also spans parts of the tissue where *C15* is not expressed (Figure 52E' and F'; Supplementary Figure 23). This observation suggests that a repressor site is missing, in combination with intron 2 enhancer 1, to precisely recapitulate *C15* expression. Enhancers can be composed of many modular sequences that work together to integrate the chromatin environment and TF binding combinations for target gene regulation (Gómez-Skarmeta *et al.*, 1995; Koshikawa, 2015). It would be useful to generate new enhancer reporter constructs that use part of intron 2 enhancer 1 and extend 5' and 3' of the sequence to identify the location of the missing repressor site(s), assuming they are positioned closely.

5.5.3.1 Dpp may coordinate *C15* expression in the male genitalia

The identification of two MAD binding sites in intron 2 enhancer 1, as well as *MAD* and *dpp* expression in the male claspers, resembles embryonic regulation of *C15* (Figure 35B and Figure 53D) (Lin *et al.*, 2006; Vincent *et al.*, 2019). Lin *et al.,* (2006) generated a combination of single, double and triple MAD-binding site mutants in the embryonic *C15* enhancer to test how this affected endogenous *C15* expression. Single mutants caused only small deviations in spatial expression, whereas the triple mutant led to a significant loss in *C15* expression (Lin

et al., 2006). It would be interesting to test this approach with the mapped MAD binding sites within intron 2 enhancer 1 to assess their regulatory role individually and in combination upon *C15* expression in the male genitalia.

Severe *dpp* allelic mutants in *Drosophila* lack pretarsal claws and male genitalia (Spencer, Hoffmann and Gelbart, 1982). In contrast, the female genitalia is not altered (Spencer, Hoffmann and Gelbart, 1982). The threshold response of *C15* to Dpp regulation via MAD is an area of particular interest required for the early characterisation of how male claspers and pretarsal segments are specified.

5.5.4 Sex-specific C15 expression in the cerci

I showed that C15 is only expressed in the female cerci and not in the male cerci (Figure 37 and Figure 46). C15 has been shown to promote the development of the two longer female epiproct bristles that are visibly distinguishable from the other bristles covering the epiproct and hypoproct (Figure 47A). A certain type of sensillum is regulated by C15 in the female epiproct, however the sensory function and role during copulation is currently unknown. *D. melanogaster* subgroup males have bristles of uniform appearance throughout the cerci, however *D. elegans* and *D. mimetica* males have a visibly different bristle on the cercus alike to *D. melanogaster* subgroup females (Figure 2B and Figure 47A). It would be fascinating to see whether *C15* is expressed in these male cercus structures by performing antibody stains and assess whether expression localises around the bristle protrusion zone as shown in females.

Cerci in male and female flies derive from the same originating A10 primordium, yet are phenotypically distinct (Sánchez and Guerrero, 2001; G. Rice *et al.*, 2019; McQueen *et al.*, 2022). The genetic changes to facilitate cercus sex-differences may depend upon morphogen signals from the A8 (female) and A9 (male) primordia during development (Gorfinkiel, Sánchez and Guerrero, 2003). In addition, some have reported that the epiproct of females is homologous to the male cercus, and the hypoproct has no homolog, suggesting it may develop from a hypothesised A11 primordium repressed in males (Sánchez and Guerrero, 2001).

I have shown that the cerci are distinguished between males and females by C15 expression (Figure 37 and Figure 46). It has been shown before than when transgenic flies express the male and female isoforms of *dsx* simultaneously, the cercus adopts an intersex phenotype (Sánchez and Guerrero, 2001). To understand how *C15* is only expressed in female cerci, but

repressed in the male counterpart, I would next look to see whether mutant clones of *dsx* in the female terminalia affects *C15* expression using antibody staining.

6. General Discussion

6.1 *Sox21b* evolutionary implications

I have shown that Sox21b is an early-acting TF that represses posterior lobe growth during male genital development (Figure 6, Figure 8 and Figure 54). Importantly, I found that this gene has evolved between *D. simulans* and *D. mauritiana* and thus contributes towards posterior lobe species differences, as illustrated using a reciprocal hemizygosity test and HCR stains (Figure 12A, Figure 14 and Figure 54). Previous literature describing posterior lobe development has mainly focused on pupal expression patterns, when the lobe protrudes from the lateral plate, as a key timepoint for identifying posterior lobe candidate genes (Glassford *et al.*, 2015; Smith, Davidson and Rebeiz, 2020). Masly *et al.,* (2011) however, used larval genital disc RNA-seq datasets to identify candidate genes underlying posterior lobe differences between *D. mauritiana* and *D. sechellia*. The identification of the role of *Sox21b* supports the likelihood that posterior lobe morphology is at least in part determined prior to metamorphosis.

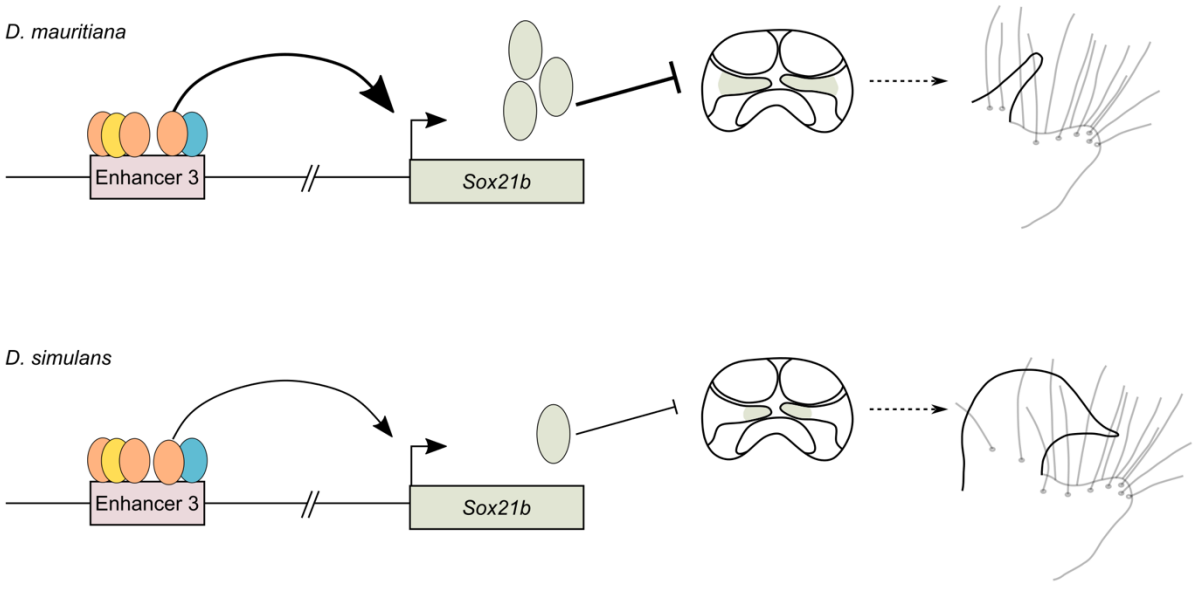


Figure 54: *Sox21b* expression differences contributing to posterior lobe divergence between *D. mauritiana* and *D. simulans*. Schematic summarising the data presented in Chapter 3 and 4. Orange ovals represent the three Abd-B proteins bound at their predicted binding sites in Enhancer 3 conserved between both species. Yellow and blue ovals represent unidentified additional TFs thought to also be regulating Enhancer 3 activity. It has been assumed that Enhancer 3 coordinates the higher expression of *Sox21b* in *D. mauritiana*, that localises to a broader region of the genital disc than in *D. simulans*. The evolved component of Enhancer 3 has not yet been uncovered.

6.1.1 *Sox21b* and copulation

The posterior lobe is required during copulation to assist male attachment to the female (Robertson, 1988; Kamimura and Mitsumoto, 2011; Yassin and Orgogozo, 2013). In accordance with the reciprocal hemizygote work presented, behavioural assays were carried out (by Emily Hood) to investigate whether the working species allele of *Sox21b* affected copulation when paired with *D. simulans* females, in a non-competitive environment (Ridgway *et al.*, 2024). Remarkably, it was found that when the paired male expressed the *D. mauritiana Sox21b* allele and consequently developed smaller lobes, copulation duration was significantly shorter compared with the reciprocal *D. simulans* working *Sox21b* allele (Ridgway *et al.*, 2024). Accordingly, it has been suggested that comparably larger *D. simulans* posterior lobes offer a copulatory advantage for males when coupling with conspecific females (Jagadeeshan and Singh, 2006; Ridgway *et al.*, 2024). The precise mechanisms of sexual selection driving posterior lobe diversity is unclear, but it seems that pre and post copulatory selection pressures both play a part (Frazee and Masly, 2015; LeVasseur-Viens, Polak and Moehring, 2015). For example, it has been suggested that the precise fit of the lobe with the female 'pouch' may impact upon the female's post copulatory response (Frazee and Masly, 2015). Although *Sox21b* has contributed to subtle size and shape differences between *D. simulans* and *D. mauritiana* their impact on reproductive success is not known. It would be interesting to perform long-term behavioural assays comparing sperm transfer, fertilisation, and the number of progenies between the reciprocal hemizygotes to assess whether they are affected by *Sox21b*'s derived lobe deviations.

6.1.2 *Sox21b*'s posterior lobe GRN

Although *Sox21b* has evolved to contribute towards posterior lobe differences between species (Figure 12 and Figure 13), this genital structure is polygenic (Chatterjee *et al.*, 2011; Glassford *et al.*, 2015; Hackett *et al.*, 2016; Smith, Davidson and Rebeiz, 2020), which means additional unidentified evolved genes must also be facilitating species shape and size differences. Characterisation of the posterior lobe GRN would help decode the genetics driving size and shape evolution. The posterior lobe GRN was shown to be co-opted from posterior spiracle development (Glassford *et al.*, 2015), yet *Sox21b* was not expressed in the latter during embryogenesis (Figure 10). I previously discussed how ChIP-Seq against Sox21b could map downstream target genes (See Chapter 4.5.4). To characterise upstream factors regulating *Sox21b* expression (Figure 54), I would use the paired yeast one hybrid (pY1H)

assay (Berenson *et al.*, 2023). The *D. mauritiana* and *D. simulans* versions of the *Sox21b* Enhancer 3 sequence would be exposed to random combinations of male genital-expressed TFs, where binding is validated from the reporter readout (Berenson *et al.*, 2023). As well as detecting individual binding sites, this assay incorporates the detection of cooperative and antagonistic TF binding mechanisms (Berenson *et al.*, 2023). ChIP-Seq and pY1H would both help determine other genes and their position within the *Sox21b* posterior lobe GRN, as well as better predict the origin and impact of gene expression evolution between species.

6.2 *Sox21b* enhancer activity divergence between species

I identified Enhancer 3 as a cis-regulatory element that recapitulated endogenous *Sox21b* genital disc expression at the posterior lobe primordia (Figure 27). Furthermore, *D. mauritiana* and *D. simulans* Enhancer 3 reporter lines showed differential activity at this region, consistent with *Sox21b* expression differences (Figure 14, Figure 28 and Figure 54). Evolved cis-acting sequences are a common mechanism underlying morphological diversity between species (Carroll, 2008), and therefore *Sox21b* represents another example of this.

6.2.1 Enhancer 3 pleiotropy

Alterations to the TF DNA-Binding domains, protein-protein interaction domains and/or expression patterns can all drive TF evolution (Chesmore *et al.*, 2016). In the case of *Sox21b*, it seems that changes to the expression between *D. simulans* and *D. mauritiana* underlie posterior lobe morphology differences (Figure 14). Although *Sox21b* is a pleiotropic TF expressed in other tissues (Crémazy, Berta and Girard, 2001), enhancer modularity can overcome this and drive tissue-specific evolution (Lynch and Wagner, 2008). For example, the wings and eyes between these two species express *Sox21b* and also differ in morphology (Gidaszewski, Baylac and Klingenberg, 2009; Posnien *et al.*, 2012). To follow up, it would be interesting to investigate whether *Sox21b* Enhancer 3 is also pleiotropic and exhibits differential activity in those tissues between *D. simulans* and *D. mauritiana*.

6.2.2 Enhancer 3 evolution

The approach used in this study to assay differences in enhancer activity has its limitations as discussed previously. Future experiments should focus on testing differences evolved at the endogenous *Sox21b* locus in the focal species rather than relying upon attP integration in *D. melanogaster*. For example, I would swap the *D. mauritiana* Enhancer 3 sequence with the *D.*

simulans version, and vice versa, and monitor for *Sox21b* expression pattern differences using HCR-IS. Species swaps of the coding sequence for a gene encoding a microsatellite-binding protein has been investigated previously between *D. simulans* and *D. melanogaster* (Brand and Levine, 2022). This method provided evidence for adaptive gene evolution between species (Brand and Levine, 2022).

6.2.3 Enhancer activity quantification

I screened for *Sox21b* enhancers using the GAL4-UAS reporter driver lines and identified just three out of nine that expressed in the male pupa genital structures (Figure 18 and Figure 21). Although I could validate tissue specific expression qualitatively, a limitation of this technique arose from the inability to quantify expression. A high throughput sequencing technique known as 'STARR-Seq' provides a method to quantify enhancer expression (Arnold *et al.*, 2013). Enhancer sequences are cloned downstream of a minimal promoter, whereby the enhancer is self-transcribed and then sequenced (Arnold *et al.*, 2013), which allows quantification of enhancer activity. Although this technique has been applied to cultured cells taken from embryos, larval brains and adult ovaries, it has the potential to expand to the genitalia. By conducting STARR-Seq on larval and pupal genital samples, I would be able to quantify *Sox21b*'s regulatory dynamics across development, as well as precisely measure for Enhancer 3 species differences (Aldaz, Escudero and Freeman, 2010).

6.3 C15 disc-dependent regulatory role

I have demonstrated that the GRNs required to coordinate the development of the male claspers and female epiproct is derived from the pretarsus but differs in key interactions involving C15 (Figure 55). I tested for interactions between C15 and known pretarsal partners in the epiproct and claspers by using RNAi knockdowns, overexpression assays and staining against each factor, to illustrate the similarities and differences between all three networks. As discussed, C15 positively regulates the development of pretarsal claw, epiproctal sensilla, arista, ocellar postvertical bristles, and medial thorax bristles, as shown by mutant *C15* alleles eradicating all structures (Campbell, 2005; Kojima, Tsuji and Saigo, 2005). The male claspers are the only example so far where C15 represses bristle development (Figure 55B).

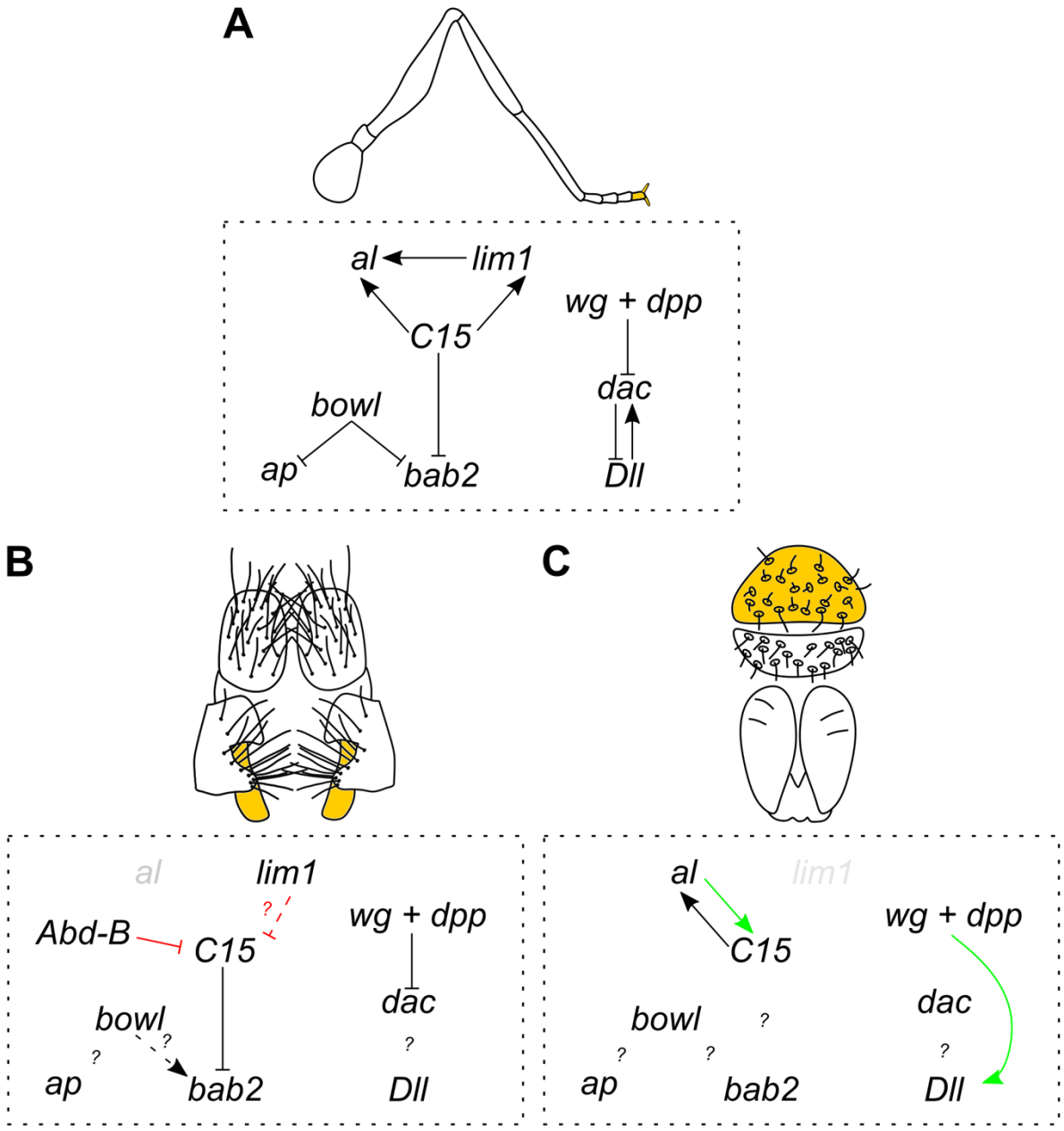


Figure 55: Summary of C15-central GRNs in the pretarsus of the leg, claspers of the male genitalia, and epiproct of the female cercus. (A) Pretarsus GRN schematic as shown in Figure 34B. (B) Clasper GRN schematic. Dashed red blunt-ended arrow represents suggested repression of *C15* by Lim1, not present in the pretarsus. *al* is not expressed in the claspers. C15 is suggested to repress *bab2* expression as both TFs do not overlap in expression during pupal development. Bowl is suggested to positively regulate *bab2* in the claspers as functional and expression data, respectively, are both localised to the distal clasper domain. *wg* and *dpp* do not completely overlap with *dac* expression, consistent with the pretarsus. The interactions connecting *ap* and *Dll* were not investigated. Abd-B is proposed to negatively regulate *C15* expression in the male claspers from staining data presented in this chapter. (C) Epiproct GRN schematic. Green arrows represent activating regulatory interactions present in the epiproct that are not seen in the pretarsus. Al and C15 positively regulate one another's expression, suggesting a positive feedback loop. *lim1* is not expressed in the epiproct. The interactions

connecting *bab2*, *bowl*, *ap* and *dac* were not investigated. Positive regulation of *Dll* by wg and Dpp has been predicted in the female epiproct from expression pattern results. The GRNs presented in this schematic have aimed to summarise the staining and functional data presented throughout Chapter Five.

6.3.1 C15 target gene differences

The genetic and molecular components causing C15 bristle repression in the male claspers are unknown. This regulatory role of C15 may derive from the lack of *al* expression (Miyazono *et al.*, 2010). Interestingly, the two structures known to only express *C15* and not *al* are the male claspers and ocelli (Figure 39A and A') (Schneitz, Spielmann and Noll, 1993), but C15 is only seen to repress bristle growth in the former (Kojima, Tsuji and Saigo, 2005). This implies that the absence of *al* expression in the male clasper is not the cause for C15's unique repressive bristle role. To investigate C15's clasper-specific role further, it would be insightful to perform CUT&RUN against C15 in pretarsal and male genital tissues (Skene and Henikoff, 2017). CUT&RUN would enable the isolation of C15-bound DNA for downstream sequencing and mapping back to the genome (Skene and Henikoff, 2017). I would also generate RNA-seq data from *C15* RNAi knockdown and GAL4 parental control pretarsus and clasper tissues. From this, I would be able to identify differentially expressed genes between the knockdown and control in both tissues, to infer which genes are directly regulated by C15. Comparisons between the RNA-seq and CUT&RUN data would identify shared and tissue specific C15 target genes that may shed insight on how C15 mediates its unique repressive role during clasper bristle development.

6.4 Evolutionary perspective of *C15*

I stained for C15 in *D. simulans* and *D. mauritiana*, and observed localisation at the proximal clasper domains and peripheral to the cerci, as was also the case in *D. melanogaster* (Figure 37 and Supplementary Figure 21). However, *D. simulans* exhibited a narrower and more elongated region of C15 clasper expression compared to *D. mauritiana*, consistent with clasper morphology differences between these species. A difference was also observed between the *C15* male genital chromatin environments, where at intron three, only *D. simulans* had a peak whereas *D. mauritiana* did not (Figure 51). Although I did not directly test whether *C15* has evolved between these species, expression pattern and ATAC-seq differences imply that this TF may play a part in coordinating clasper phenotypic variation.

6.4.1 *Tribolium C15* and *al*

The pretarsal claw is conserved amongst *Drosophila* species. Moreso, this structure is conserved across distantly related insects such as cockroaches and dock bugs, as well as the beetle, *Tribolium* (Angelini, Smith and Jockusch, 2012; Schmitt and Betz, 2017; Rebora *et al.*, 2021). *Tribolium* have been shown to express *C15* at the distal appendage tips, like in *Drosophila* (Angelini, Smith and Jockusch, 2012; Cande *et al.*, 2012). However, RNAi knockdown of *C15* in *Tribolium* did not eliminate the claw, whereas it did in *Drosophila* (Angelini, Smith and Jockusch, 2012; Cande *et al.*, 2012). It could be that RNAi knockdown in *Tribolium* did not completely abolish *C15* expression, thus it cannot be ruled out whether C15 is the master coordinator of claw development outside of *Drosophila. al* expression overlaps with *C15* in the distal appendage tips of *Tribolium*, and RNAi knockdown of *al* eliminates the claw structures (Beermann and Schröder, 2004), consistent with its role during *Drosophila* pretarsus development. I would be intrigued to see whether a chimeric swap of *C15* and *al* *Drosophila* loci, including the coding sequence and regulatory regions, with *Tribolium*'s would retain homologous pretarsal claw development. Broadly speaking, this would be an insightful experiment to understand evolutionary constraints upon appendage-specifying genes.

6.4.2 Clasper diversity amongst *Drosophila* species

D. teissieri diverged from *D. melanogaster* approximately eight million years ago (Chyb and Gompel, 2013). Interestingly, the claspers of this species are completely covered in bristles, including the proximal clasper domain that is shown to be naked in the *D. melanogaster* subgroup (Chyb and Gompel, 2013). I revealed that C15 protein marks the proximal clasper domain and does not overlap with bristle cell precursors (Figure 37), and also identified the enhancer determining *C15* clasper expression (Figure 52E and F). It would be very interesting to screen for *C15* expression in *D. teissieri* to see whether the ubiquitous coverage of bristles in the proximal clasper domain results from the loss of *C15* expression. In addition, pairwise sequence alignment between the *D. teissieri* and *D. melanogaster C15* clasper enhancer would help identify sequence differences between the two species that may give rise to any *C15* expression differences observed.

6.5 Conclusions

In my thesis, I have investigated the genetic basis underpinning genital morphological diversity in *Drosophila*. This work has identified *Sox21b* as an evolved gene causative of male posterior

lobe shape and size differences between *D. simulans* and *D. mauritiana*. This is the first evolved posterior lobe gene to be identified between these species thus far. In addition, I have demonstrated that the male clasper and female epiproct developmental GRNs are derived from the *C15* leg pretarsus network. In summary, my PhD research has contributed hugely to our understanding of *Drosophila* genital development and evolution and has provided imperative data for continuing to decode morphological diversity.

7. Bibliography

Acebes, A., Cobb, M. and Ferveur, J.-F. (2003) 'Species-specific effects of single sensillum ablation on mating position in *Drosophila*', *Journal of Experimental Biology*, 206(17), pp. 3095–3100. Available at: https://doi.org/10.1242/jeb.00522.

Affolter, M. and Basler, K. (2007) 'The Decapentaplegic morphogen gradient: from pattern formation to growth regulation', *Nature Reviews Genetics*, 8(9), pp. 663–674.

Affolter, M., Slattery, M. and Mann, R.S. (2008) 'A lexicon for homeodomain-DNA recognition', *Cell*, 133(7), pp. 1133–1135.

Akam, M. (1995) 'Hox Genes and the Evolution of Diverse Body Plans', *Philosophical Transactions: Biological Sciences*, 349(1329), pp. 313–319. Available at: http://www.jstor.org/stable/56229.

Aldaz, S., Escudero, L.M. and Freeman, M. (2010) 'Live imaging of *Drosophila* imaginal disc development', *Proceedings of the National Academy of Sciences*, 107(32), pp. 14217–14222. Available at: https://doi.org/10.1073/pnas.1008623107.

Angelini, D.R., Smith, F.W. and Jockusch, E.L. (2012) 'Extent With Modification: Leg Patterning in the Beetle *Tribolium castaneum* and the Evolution of Serial Homologs', *G3 Genes|Genomes|Genetics*, 2(2), pp. 235–248. Available at: https://doi.org/10.1534/g3.111.001537.

Arif, S., Kittelmann, S. and McGregor, A.P. (2015) 'From shavenbaby to the naked valley: trichome formation as a model for evolutionary developmental biology', *Evolution & Development*, 17(1), pp. 120–126. Available at: https://doi.org/10.1111/ede.12113.

Arnold, C.D. *et al.* (2013) 'Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq', *Science*, 339(6123), pp. 1074–1077. Available at: https://doi.org/10.1126/science.1232542.

Arnosti, D.N. *et al.* (1996) 'The eve stripe 2 enhancer employs multiple modes of transcriptional synergy', *Development*, 122(1), pp. 205–214. Available at: https://doi.org/10.1242/dev.122.1.205.

Arnoult, L. *et al.* (2013) 'Emergence and diversification of fly pigmentation through evolution of a gene regulatory module', *Science*, 339(6126), pp. 1423–1426.

Arnqvist, G. (1997) 'The evolution of animal genitalia: distinguishing between hypotheses by single species studies', *Biological Journal of the Linnean Society*, 60(3), pp. 365–379. Available at: https://doi.org/10.1111/j.1095-8312.1997.tb01501.x.

Arnqvist, G. and Danielsson, I. (1999) 'Copulatory behavior, genital morphology, and male fertilization success in water striders', *Evolution*, 53(1), pp. 147–156.

Baanannou, A. *et al.* (2013) '*Drosophila* Distal-less and Rotund Bind a Single Enhancer Ensuring Reliable and Robust *bric-a-brac2* Expression in Distinct Limb Morphogenetic Fields', *PLOS Genetics*, 9(6), pp. e1003581-. Available at: https://doi.org/10.1371/journal.pgen.1003581.

Bainbridge, S.P. and Bownes, M. (1981) 'Staging the metamorphosis of *Drosophila melanogaster*', *Development*, 66(1), pp. 57–80. Available at: https://doi.org/10.1242/dev.66.1.57.

Banerji, J., Rusconi, S. and Schaffner, W. (1981) 'Expression of a β-globin gene is enhanced by remote SV40 DNA sequences', *Cell*, 27(2), pp. 299–308.

Barski, A. *et al.* (2007) 'High-resolution profiling of histone methylations in the human genome', *Cell*, 129(4), pp. 823–837.

Basler, K. and Struhl, G. (1994) 'Compartment boundaries and the control of *Drosophila* limb pattern by hedgehog protein', *Nature*, 368(6468), pp. 208–214.

Beermann, A. and Schröder, R. (2004) 'Functional stability of the aristaless gene in appendage tip formation during evolution', *Development Genes and Evolution*, 214(6), pp. 303–308. Available at: https://doi.org/10.1007/s00427-004-0411-7.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), pp. 289–300. Available at: https://doi.org/https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bentsen, M. *et al.* (2020) 'ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation', *Nature Communications*, 11(1), p. 4267. Available at: https://doi.org/10.1038/s41467-020-18035-1.

Berenson, A. *et al.* (2023) 'Paired yeast one-hybrid assays to detect DNA-binding cooperativity and antagonism across transcription factors', *Nature Communications*, 14(1), p. 6570. Available at: https://doi.org/10.1038/s41467-023-42445-6.

Berman, B.P. *et al.* (2002) 'Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome', *Proceedings of the National Academy of Sciences*, 99(2), pp. 757–762.

Bernard, P. and Harley, V.R. (2010) 'Acquisition of SOX transcription factor specificity through protein–protein interaction, modulation of Wnt signalling and post-translational modification', *The international journal of biochemistry & cell biology*, 42(3), pp. 400–410.

Bischof, J. *et al.* (2013) 'A versatile platform for creating a comprehensive UAS-ORFeome library in *Drosophila*', *Development*, 140(11), pp. 2434–2442. Available at: https://doi.org/10.1242/dev.088757.

Blackman, R.K. *et al.* (1991) 'An extensive 3' cis-regulatory region directs the imaginal disk expression of decapentaplegic, a member of the TGF-β family in *Drosophila*', *Development*, 111(3), pp. 657–666.

Blake, A.J. *et al.* (2017) 'RNAi-Based Techniques for the Analysis of Gene Function in *Drosophila* Germline Stem Cells', in B. Zhang (ed.) *RNAi and Small Regulatory RNAs in Stem Cells: Methods and Protocols*. New York, NY: Springer New York, pp. 161–184. Available at: https://doi.org/10.1007/978-1-4939-7108-4_13.

Boll, W. and Noll, M. (2002) 'The *Drosophila* Pox neuro gene: control of male courtship behavior and fertility as revealed by a complete dissection of all enhancers', *Development*, 129(24), pp. 5667–5681. Available at: https://doi.org/10.1242/dev.00157.

Bonifer, C. and Cockerill, P.N. (2017) 'Chromatin priming of genes in development: Concepts, mechanisms and consequences', *Experimental Hematology*, 49, pp. 1–8. Available at: https://doi.org/https://doi.org/10.1016/j.exphem.2017.01.003.

Bonini, N.M., Leiserson, W.M. and Benzer, S. (1998) 'Multiple Roles of the eyes absent Gene in *Drosophila*', *Developmental Biology*, 196(1), pp. 42–57. Available at: https://doi.org/https://doi.org/10.1006/dbio.1997.8845.

Bothma, J.P. *et al.* (2014) 'Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living *Drosophila* embryos', *Proceedings of the National Academy of Sciences*, 111(29), pp. 10598–10603. Available at: https://doi.org/10.1073/pnas.1410022111.

Boudinot, B.E. (2018) 'A general theory of genital homologies for the Hexapoda (Pancrustacea) derived from skeletomuscular correspondences, with emphasis on the Endopterygota', *Arthropod Structure & Development*, 47(6), pp. 563–613. Available at: https://doi.org/https://doi.org/10.1016/j.asd.2018.11.001.

Bourbon, H.-M.G. *et al.* (2022) 'A shared ancient enhancer element differentially regulates the *bric-a-brac* tandem gene duplicates in the developing *Drosophila* leg', *PLOS Genetics*, 18(3), pp. e1010083-. Available at: https://doi.org/10.1371/journal.pgen.1010083.

Brand, C.L. and Levine, M.T. (2022) 'Cross-species incompatibility between a DNA satellite and the *Drosophila* Spartan homolog poisons germline genome integrity', *Current Biology*, 32(13), pp. 2962-2971.e4. Available at: https://doi.org/10.1016/j.cub.2022.05.009.

Bravo González-Blas, C. *et al.* (2020) 'Identification of genomic enhancers through spatial integration of single-cell transcriptomics and epigenomics', *Molecular Systems Biology*, 16(5), p. e9438. Available at: https://doi.org/10.15252/msb.20209438.

Brennan, P.L.R. and Prum, R.O. (2015) 'Mechanisms and evidence of genital coevolution: the roles of natural selection, mate choice, and sexual conflict', *Cold Spring Harbor Perspectives in Biology*, 7(7), p. a017749.

Bridges, C.B. and Morgan, T.H. (1923) *The third-chromosome group of mutant characters of* Drosophila melanogaster. Carnegie Institution of Washington.

Buchberger, E. *et al.* (2021) 'Variation in Pleiotropic Hub Gene Expression Is Associated with Interspecific Differences in Head Shape and Eye Size in *Drosophila*', *Molecular Biology and Evolution*, 38(5), p. 1924. Available at: https://doi.org/10.1093/MOLBEV/MSAA335.

Calleja, M. *et al.* (2002) 'How to pattern an epithelium: lessons from *achaete-scute* regulation on the notum of *Drosophila*', *Gene*, 292(1–2), pp. 1–12.

Camino, E. *et al.* (2015) 'The Evolutionary Origination and Diversification of a Dimorphic Gene Regulatory Network through Parallel Innovations in cis and trans', *PLoS genetics*, 11, p. e1005136. Available at: https://doi.org/10.1371/journal.pgen.1005136.

Campbell, G. (2002) 'Distalization of the *Drosophila* leg by graded EGF-receptor activity', *Nature*, 418(6899), pp. 781–785. Available at: https://doi.org/10.1038/nature00971.

Campbell, G. (2005) 'Regulation of gene expression in the distal region of the *Drosophila* leg by the Hox11 homolog, C15', *Developmental Biology*, 278(2), pp. 607–618. Available at: https://doi.org/https://doi.org/10.1016/j.ydbio.2004.12.009.

Campbell, G., Weaver, T. and Tomlinson, A. (1993) 'Axis specification in the developing *Drosophila* appendage: The role of *wingless*, *decapentaplegic*, and the homeobox gene *aristaless*', *Cell*, 74(6), pp. 1113–1123. Available at: https://doi.org/https://doi.org/10.1016/0092-8674(93)90732-6.

Cande, J. *et al.* (2012) 'Evolution of multiple additive loci caused divergence between *Drosophila yakuba* and *D. santomea* in wing rowing during male courtship', *PloS one*. 2012/08/30, 7(8), pp. e43888–e43888. Available at: https://doi.org/10.1371/journal.pone.0043888.

Caple, J., Byrd, J. and Stephan, C.N. (2017) 'Elliptical Fourier analysis: fundamentals, applications, and value for forensic anthropology', *International Journal of Legal Medicine*, 131(6), pp. 1675–1690. Available at: https://doi.org/10.1007/s00414-017-1555-0.

Carl, S.H. and Russell, S. (2015) 'Common binding by redundant group B Sox proteins is evolutionarily conserved in *Drosophila*', *BMC Genomics*, 16(1), p. 292. Available at: https://doi.org/10.1186/s12864-015-1495-3.

Carroll, S.B. (2008) 'Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution', *Cell*, 134(1), pp. 25–36. Available at: https://doi.org/10.1016/j.cell.2008.06.030.

Carroll, S.B., Grenier, J.K. and Weatherbee, S.D. (2009) *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Wiley. Available at: https://books.google.co.uk/books?id=tayrCszYKdkC.

Casares, F. *et al.* (1997) 'The genital disc of *Drosophila melanogaster*', *Development Genes and Evolution*, 207(4), pp. 216–228. Available at: https://doi.org/10.1007/s004270050110.

Castro-Mondragon, J.A. *et al.* (2022) 'JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles', *Nucleic Acids Research*, 50(D1), pp. D165–D173. Available at: https://doi.org/10.1093/nar/gkab1113.

de Celis Ibeas, J.M. and Bray, S.J. (2003) 'Bowl is required downstream of Notch for elaboration of distal limb patterning', *Development*, 130(24), pp. 5943–5952. Available at: https://doi.org/10.1242/dev.00833.

Celniker, S.E. *et al.* (1990) 'The molecular genetics of the bithorax complex of *Drosophila*: cis-regulation in the Abdominal-B domain.', *The EMBO Journal*, 9(13), pp. 4277–4286. Available at: https://doi.org/https://doi.org/10.1002/j.1460-2075.1990.tb07876.x.

Chan, I.Z.W., Stevens, M. and Todd, P.A. (2019) 'pat-geom: A software package for the analysis of animal patterns', *Methods in Ecology and Evolution*, 10(4), pp. 591–600. Available at: https://doi.org/https://doi.org/10.1111/2041-210X.13131.

Chan, Y.F. *et al.* (2010) 'Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer', *Science*, 327(5963), pp. 302–305. Available at: https://doi.org/10.1126/science.1182213.

Chang, Z. *et al.* (1993) 'Molecular and genetic characterization of the *Drosophila tartan* gene', *Developmental biology*, 160(2), pp. 315–332.

Chatterjee, S.S. *et al.* (2011) 'The female-specific Doublesex isoform regulates pleiotropic transcription factors to pattern genital development in *Drosophila*', *Development*, 138(6), p. 1099. Available at: https://doi.org/10.1242/dev.055731.

Chen, E.H. and Baker, B.S. (1997) 'Compartmental organization of the *Drosophila* genital imaginal discs', *Development*, 124(1), pp. 205–218. Available at: https://doi.org/10.1242/dev.124.1.205.

Chesmore, K.N. *et al.* (2016) 'Complex Patterns of Association between Pleiotropy and Transcription Factor Evolution', *Genome Biology and Evolution*, 8(10), pp. 3159–3170. Available at: https://doi.org/10.1093/gbe/evw228.

Cheung, L.S. *et al.* (2013) 'Dynamic model for the coordination of two enhancers of broad by EGFR signaling', *Proceedings of the National Academy of Sciences*, 110(44), pp. 17939–17944. Available at: https://doi.org/10.1073/pnas.1304753110.

Choi, H.M.T. *et al.* (2018) 'Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust', *Development*, 145(12), p. dev165753.

Chovnick, A. and Lefkowitz, R.J. (1956) 'A phenogenetic study of the lozenge pseudoalleles in *Drosophila melanogaster*. i. effects on the development of the tarsal claws in homozygotes', *Genetics*, 41(1), pp. 79–92. Available at: https://doi.org/10.1093/genetics/41.1.79.

Chyb, S. and Gompel, N. (2013) '*D. melanogaster* subgroup species', in S. Chyb and N. Gompel (eds) *Atlas of Drosophila Morphology*. San Diego: Academic Press, pp. 209–220. Available at: https://doi.org/https://doi.org/10.1016/B978-0-12-384688-4.00010-9.

Cohn, M.J. and Tickle, C. (1999) 'Developmental basis of limblessness and axial patterning in snakes', *Nature*, 399(6735), pp. 474–479. Available at: https://doi.org/10.1038/20944.

Couderc, J.-L. *et al.* (2002) 'The *bric à brac* locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in *Drosophila*', *Development*, 129(10), pp. 2419–2433. Available at: https://doi.org/10.1242/dev.129.10.2419.

Coyne, J.A. (1983) *Genetic Basis of Differences in Genital Morphology Among Three Sibling Species of Drosophila*.

Crémazy, F., Berta, P. and Girard, F. (2001) 'Genome-wide analysis of Sox genes in *Drosophila melanogaster*', *Mechanisms of Development*, 109(2), pp. 371–375. Available at: https://doi.org/https://doi.org/10.1016/S0925-4773(01)00529-9.

Creyghton, M.P. *et al.* (2010) 'Histone H3K27ac separates active from poised enhancers and predicts developmental state', *Proceedings of the National Academy of Sciences*, 107(50), pp. 21931–21936.

Deline, B. *et al.* (2018) 'Evolution of metazoan morphological disparity', *Proceedings of the National Academy of Sciences*, 115(38), pp. E8909–E8918. Available at: https://doi.org/10.1073/pnas.1810575115.

Delker, R.K. *et al.* (2019) 'Low affinity binding sites in an activating CRM mediate negative autoregulation of the *Drosophila* Hox gene Ultrabithorax', *PLOS Genetics*, 15(10), pp. e1008444-. Available at: https://doi.org/10.1371/journal.pgen.1008444.

Delon, I., Chanut-Delalande, H. and Payre, F. (2003) 'The Ovo/Shavenbaby transcription factor specifies actin remodelling during epidermal differentiation in *Drosophila*', *Mechanisms of Development*, 120(7), pp. 747–758. Available at: https://doi.org/https://doi.org/10.1016/S0925-4773(03)00081-9.

Dembeck, L.M. *et al.* (2015) 'Genetic architecture of abdominal pigmentation in *Drosophila melanogaster*', *PLoS genetics*, 11(5), p. e1005163.

Derelle, R. *et al.* (2013) 'Color differences among feral pigeons (*Columba livia*) are not attributable to sequence variation in the coding region of the melanocortin-1 receptor gene (*MC1R*)', *BMC Research Notes*, 6(1), p. 310. Available at: https://doi.org/10.1186/1756-0500-6-310.

Diaz-Benjumea, F.J., Cohen, B. and Cohen, S.M. (1994) 'Cell interaction between compartments establishes the proximal-distal axis of *Drosophila* legs', *Nature*, 372(6502), pp. 175–179. Available at: https://doi.org/10.1038/372175a0.

Dietzl, G. *et al.* (2007) 'A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*', *Nature*, 448(7150), pp. 151–156. Available at: https://doi.org/10.1038/nature05954.

Dreyer, A.P. and Shingleton, A.W. (2011) 'The Effect of Genetic and Environmental Variation on Genital Size in Male *Drosophila*: Canalized but Developmentally Unstable', *PLOS ONE*, 6(12), pp. e28278-. Available at: https://doi.org/10.1371/journal.pone.0028278.

Eberhard, W. (1985) *Sexual selection and animal genitalia*. Cambridge: Harvard University Press.

Emerald, B.S. and Roy, J.K. (1998) 'Organising activities of engrailed, hedgehog, wingless and decapentaplegic in the genital discs of *Drosophila melanogaster*', *Development Genes and Evolution*, 208(9), pp. 504–516. Available at: https://doi.org/10.1007/s004270050209.

Erwin, D.H. and Davidson, E.H. (2006) 'Response to comment on" Gene regulatory networks and the evolution of animal body plans"', *Science*, 313(5788), p. 761.

Estella, C. and Mann, R.S. (2008) 'Logic of Wg and Dpp induction of distal and medial fates in the *Drosophila* leg', *Development*, 135(4), pp. 627–636. Available at: https://doi.org/10.1242/dev.014670.

Estrada, B., Casares, F. and Sánchez-Herrero, E. (2003) 'Development of the genitalia in *Drosophila melanogaster*', *Differentiation*, 71(6), pp. 299–310. Available at: https://doi.org/https://doi.org/10.1046/j.1432-0436.2003.03017.x.

Estrada, B. and Sánchez-Herrero, E. (2001) 'The Hox gene *Abdominal-B* antagonizes appendage development in the genital disc of *Drosophila*', *Development*, 128(3), pp. 331–339. Available at: https://doi.org/10.1242/dev.128.3.331.

Feng, S. *et al.* (2022) 'Transcription factor paralogs orchestrate alternative gene regulatory networks by context-dependent cooperation with multiple cofactors', *Nature Communications*, 13(1), p. 3808. Available at: https://doi.org/10.1038/s41467-022-31501-2.

Foronda, D. *et al.* (2006) 'Requirement of abdominal-A and Abdominal-B in the developing genitalia of *Drosophila* breaks the posterior downregulation rule', *Development*, 133(1), pp. 117–127. Available at: https://doi.org/10.1242/dev.02173.

Frankel, N. *et al.* (2010) 'Phenotypic robustness conferred by apparently redundant transcriptional enhancers', *Nature*, 466(7305), pp. 490–493.

Frankel, N. *et al.* (2011) 'Morphological evolution caused by many subtle-effect substitutions in regulatory DNA', *Nature*, 474(7353), pp. 598–603. Available at: https://doi.org/10.1038/nature10200.

Frankel, N., Wang, S. and Stern, D.L. (2012) 'Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution', *Proceedings of the National Academy of Sciences*, 109(51), pp. 20975–20979.

Frazee, S.R. *et al.* (2021) 'Interspecific introgression reveals a role of male genital morphology during the evolution of reproductive isolation in *Drosophila*', *Evolution*, 75(5), pp. 989–1002. Available at: https://doi.org/10.1111/evo.14169.

Frazee, S.R. and Masly, J.P. (2015) 'Multiple sexual selection pressures drive the rapid evolution of complex morphology in a male secondary genital structure', *Ecology and evolution*, 5(19), pp. 4437–4450. Available at: https://doi.org/10.1002/ece3.1721.

Fuß, B. *et al.* (2001) 'Control of endoreduplication domains in the *Drosophila* gut by the *knirps* and *knirps-related* genes', *Mechanisms of Development*, 100(1), pp. 15–23. Available at: https://doi.org/https://doi.org/10.1016/S0925-4773(00)00512-8.

Galindo, M.I. *et al.* (2002) 'Leg patterning driven by proximal-distal interactions and EGFR signaling', *Science*, 297(5579), pp. 256–259.

Galindo, M.I., Bishop, S.A. and Couso, J.P. (2005) 'Dynamic EGFR-Ras signalling in *Drosophila* leg development', *Developmental Dynamics*, 233(4), pp. 1496–1508. Available at: https://doi.org/https://doi.org/10.1002/dvdy.20452.

Gao, J. *et al.* (2022) 'Secondary reversion to sexual monomorphism associated with tissue-specific loss of *doublesex* expression', *Evolution*, 76(9), pp. 2089–2104. Available at: https://doi.org/10.1111/evo.14564.

García-García, M.J. *et al.* (1999) 'Different contributions of pannier and wingless to the patterning of the dorsal mesothorax of *Drosophila*', *Development*, 126(16), pp. 3523–3532.

Garrigan, D. *et al.* (2012) 'Genome sequencing reveals complex speciation in the *Drosophila simulans* clade', *Genome Research*, 22(8), pp. 1499–1511. Available at: https://doi.org/10.1101/GR.130922.111.

Gaspar, P. *et al.* (2020) 'Characterization of the Genetic Architecture Underlying Eye Size Variation Within *Drosophila melanogaster* and *Drosophila simulans*', *G3 Genes|Genomes|Genetics*, 10(3), pp. 1005–1018. Available at: https://doi.org/10.1534/g3.119.400877.

Ghosh, N. *et al.* (2019) 'The Hox gene *Abdominal-B* uses DoublesexF as a cofactor to promote neuroblast apoptosis in the *Drosophila* central nervous system', *Development*, 146(16), p. dev175158. Available at: https://doi.org/10.1242/dev.175158.

Gibert, J.-M. *et al.* (2016) 'Phenotypic Plasticity through Transcriptional Regulation of the Evolutionary Hotspot Gene *tan* in *Drosophila melanogaster*', *PLOS Genetics*, 12(8), pp. e1006218-. Available at: https://doi.org/10.1371/journal.pgen.1006218.

Gidaszewski, N.A., Baylac, M. and Klingenberg, C.P. (2009) 'Evolution of sexual dimorphism of wing shape in the *Drosophila melanogaster* subgroup', *BMC Evolutionary Biology*, 9(1), p. 110. Available at: https://doi.org/10.1186/1471-2148-9-110.

Glassford, W.J. *et al.* (2015) 'Co-option of an Ancestral Hox-Regulated Network Underlies a Recently Evolved Morphological Novelty', *Developmental cell*. 2015/09/03, 34(5), pp. 520–531. Available at: https://doi.org/10.1016/j.devcel.2015.08.005.

Goldman, T.D. and Arbeitman, M.N. (2007) 'Genomic and Functional Studies of *Drosophila* Sex Hierarchy Regulated Gene Expression in Adult Head and Nervous System Tissues', *PLOS Genetics*, 3(11), pp. e216-. Available at: https://doi.org/10.1371/journal.pgen.0030216.

Gómez-Skarmeta, J.L. *et al.* (1995) 'Cis-regulation of *achaete* and *scute*: shared enhancer-like elements drive their coexpression in proneural clusters of the imaginal discs'.

Gompel, N. *et al.* (2005) 'Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*', *Nature*, 433(7025), pp. 481–487. Available at: https://doi.org/10.1038/nature03235.

Gompel, N. and Ayla Schröder, E. (2015) *Drosophila germline transformation*.

Gompel, N. and Carroll, S.B. (2003) 'Genetic mechanisms and constraints governing the evolution of correlated traits in Drosophilid flies', *Nature*, 424(6951), pp. 931–935.

Gorfinkiel, N., Sánchez, L. and Guerrero, I. (1999) '*Drosophila* terminalia as an appendage-like structure', *Mechanisms of Development*, 86(1), pp. 113–123. Available at: https://doi.org/https://doi.org/10.1016/S0925-4773(99)00122-7.

Gorfinkiel, N., Sánchez, L. and Guerrero, I. (2003) 'Development of the *Drosophila* genital disc requires interactions between its segmental primordia', *Development*, 130(2), pp. 295–305. Available at: https://doi.org/10.1242/DEV.00214.

Goto, T., Macdonald, P. and Maniatis, T. (1989) 'Early and late periodic patterns of *even skipped* expression are controlled by distinct regulatory elements that respond to different spatial cues', *Cell*, 57(3), pp. 413–422. Available at: https://doi.org/10.1016/0092-8674(89)90916-1.

Grant, C.E., Bailey, T.L. and Noble, W.S. (2011) 'FIMO: scanning for occurrences of a given motif', *Bioinformatics*, 27(7), pp. 1017–1018. Available at: https://doi.org/10.1093/bioinformatics/btr064.

Gratz, S.J. *et al.* (2014) 'Highly Specific and Efficient CRISPR/Cas9-Catalyzed Homology-Directed Repair in *Drosophila*', *Genetics*, 196(4), pp. 961–971. Available at: https://doi.org/10.1534/genetics.113.160713.

Gratz, S.J. *et al.* (2015) 'CRISPR-Cas9 Genome Editing in *Drosophila*', *Current protocols in molecular biology*, 111, pp. 31.2.1-31.2.20. Available at: https://doi.org/10.1002/0471142727.mb3102s111.

Das Gupta, M. and Tsiantis, M. (2018) 'Gene networks and the evolution of plant morphology', *Current Opinion in Plant Biology*, 45, pp. 82–87. Available at: https://doi.org/https://doi.org/10.1016/j.pbi.2018.05.011.

Gutiérrez, L. *et al.* (2003) 'The *Drosophila* trithorax group gene *tonalli* (*tna*) interacts genetically with the Brahma remodeling complex and encodes an SP-RING finger protein'.

Hackett, J.L. *et al.* (2016) 'Mapping QTL Contributing to Variation in Posterior Lobe Morphology between Strains of *Drosophila melanogaster*', *PloS one*, 11(9), pp. e0162573–e0162573. Available at: https://doi.org/10.1371/journal.pone.0162573.

Hagen, J.F.D. *et al.* (2019) '*tartan* underlies the evolution of *Drosophila* male genital morphology', *Proceedings of the National Academy of Sciences*, 116(38), p. 19025. Available at: https://doi.org/10.1073/pnas.1909829116.

Hagen, J.F.D. *et al.* (2020) 'Unraveling the Genetic Basis for the Rapid Diversification of Male Genitalia between *Drosophila* Species', *Molecular Biology and Evolution*, 38(2), pp. 437–448. Available at: https://doi.org/10.1093/MOLBEV/MSAA232.

Hoekstra, H.E. *et al.* (2006) 'A Single Amino Acid Mutation Contributes to Adaptive Beach Mouse Color Pattern', *Science*, 313(5783), pp. 101–104. Available at: https://doi.org/10.1126/science.1126121.

Hong, J.-W., Hendrix, D.A. and Levine, M.S. (2008) 'Shadow Enhancers as a Source of Evolutionary Novelty', *Science*, 321(5894), p. 1314. Available at: https://doi.org/10.1126/science.1160631.

Hosken, D.J. and Stockley, P. (2004) 'Sexual selection and genital evolution', *Trends in ecology & evolution*, 19(2), pp. 87–93.

Hothorn, T., Bretz, F. and Westfall, P. (2008) 'Simultaneous Inference in General Parametric Models', *Biometrical Journal*, 50(3), pp. 346–363.

House, C.M. *et al.* (2013) 'Sexual and Natural Selection Both Influence Male Genital Evolution', *PLOS ONE*, 8(5), p. e63807. Available at: https://doi.org/10.1371/JOURNAL.PONE.0063807.

Inaki, M., Liu, J. and Matsuno, K. (2016) 'Cell chirality: its origin and roles in left–right asymmetric development', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1710), p. 20150403. Available at: https://doi.org/10.1098/rstb.2015.0403.

Jagadeeshan, S. and Singh, R.S. (2006) 'A time-sequence functional analysis of mating behaviour and genital coupling in *Drosophila*: role of cryptic female choice and male sex-drive in the evolution of male genitalia', *Journal of Evolutionary Biology*, 19(4), pp. 1058–1070. Available at: https://doi.org/10.1111/j.1420-9101.2006.01099.x.

Jean-Marie Tremblay, B. (2021) 'universalmotif: Import, Modify, and Export Motifs with R'.

Jenett, A. *et al.* (2012) 'A GAL4-Driver Line Resource for *Drosophila* Neurobiology', *Cell Reports*, 2(4), pp. 991–1001. Available at: https://doi.org/https://doi.org/10.1016/j.celrep.2012.09.011.

Jiang, J. and Struhl, G. (1996) 'Complementary and mutually exclusive activities of decapentaplegic and wingless organize axial patterning during *Drosophila* leg development', *Cell*, 86(3), pp. 401–409.

Jois, S. *et al.* (2022) 'Sexually dimorphic peripheral sensory neurons regulate copulation duration and persistence in male *Drosophila*', *Scientific Reports*, 12(1), p. 6177. Available at: https://doi.org/10.1038/s41598-022-10247-3.

Kachhap, S., Priyadarshini, P. and Singh, B. (2017) 'Molecular dynamics simulations show altered secondary structure of clawless in binary complex with DNA providing insights into aristaless-clawless-DNA ternary complex formation', *Journal of Biomolecular Structure and Dynamics*, 35(6), pp. 1153–1167. Available at: https://doi.org/10.1080/07391102.2016.1175967.

Kalay, G. *et al.* (2019) 'Redundant and Cryptic Enhancer Activities of the *Drosophila yellow* Gene', *Genetics*, 212(1), pp. 343–360. Available at: https://doi.org/10.1534/genetics.119.301985.

Kamimura, Y. and Mitsumoto, H. (2011) 'Comparative copulation anatomy of the *Drosophila melanogaster* species complex (Diptera: Drosophilidae)', *Entomological Science*, 14(4), pp. 399–410. Available at: https://doi.org/https://doi.org/10.1111/j.1479-8298.2011.00467.x.

Kassambara, A. and Mundt, F. (2020) 'Extract and Visualize the Results of Multivariate Data Analyses'. Available at: https://cran.r-project.org/package=factoextra (Accessed: 5 August 2022).

Keisman, E.L. and Baker, B.S. (2001) 'The *Drosophila* sex determination hierarchy modulates wingless and decapentaplegic signaling to deploy dachshund sex-specifically in the genital imaginal disc', *Development*, 128(9), pp. 1643–1656. Available at: https://doi.org/10.1242/dev.128.9.1643.

Kerber, B. *et al.* (2001) 'The AP-2 transcription factor is required for joint formation and cell survival in *Drosophila* leg development', *Development*, 128(8), pp. 1231–1238. Available at: https://doi.org/10.1242/dev.128.8.1231.

Kittelmann, S. *et al.* (2018) 'Gene regulatory network architecture in different developmental contexts influences the genetic basis of morphological evolution', *PLoS Genetics*, 14(5). Available at: https://doi.org/10.1371/journal.pgen.1007375.

Kittelmann, S. *et al.* (2021) 'A complex gene regulatory architecture underlies the development and evolution of cuticle morphology in *Drosophila*', *Current Opinion in Genetics & Development*, 69, pp. 21–27. Available at: https://doi.org/https://doi.org/10.1016/j.gde.2021.01.003.

Koenecke, N. *et al.* (2017) '*Drosophila* poised enhancers are generated during tissue patterning with the help of repression', *Genome research*, 27(1), pp. 64–74.

Kojima, T., Tsuji, T. and Saigo, K. (2005) 'A concerted action of a paired-type homeobox gene, *aristaless*, and a homolog of Hox11/tlx homeobox gene, *clawless*, is essential for the distal tip development of the *Drosophila* leg', *Developmental Biology*, 279(2), pp. 434–445. Available at: https://doi.org/https://doi.org/10.1016/j.ydbio.2004.12.005.

Kopp, A. (2011) '*Drosophila* sex combs as a model of evolutionary innovations', *Evolution & Development*, 13(6), pp. 504–522. Available at: https://doi.org/https://doi.org/10.1111/j.1525-142X.2011.00507.x.

Kopp, A., Muskavitch, M.A.T. and Duncan, I. (1997) 'The roles of hedgehog and engrailed in patterning adult abdominal segments of *Drosophila*', *Development*, 124(19), pp. 3703–3714. Available at: https://doi.org/10.1242/dev.124.19.3703.

Kopp, A. and True, J.R. (2002) 'Evolution of male sexual characters in the Oriental *Drosophila melanogaster* species group', *Evolution & Development*, 4(4), pp. 278–291. Available at: https://doi.org/https://doi.org/10.1046/j.1525-142X.2002.02017.x.

Koshikawa, S. (2015) 'Enhancer modularity and the evolution of new traits', *Fly*, 9(4), pp. 155–159. Available at: https://doi.org/10.1080/19336934.2016.1151129.

Kudron, M.M. *et al.* (2018) 'The ModERN Resource: Genome-Wide Binding Profiles for Hundreds of *Drosophila* and *Caenorhabditis elegans* Transcription Factors', *Genetics*, 208(3), pp. 937–949. Available at: https://doi.org/10.1534/genetics.117.300657.

Kvon, E.Z. *et al.* (2016) 'Progressive Loss of Function in a Limb Enhancer during Snake Evolution', *Cell*, 167(3), pp. 633-642.e11. Available at: https://doi.org/10.1016/j.cell.2016.09.028.

Lamb, A.M. *et al.* (2020) '*ebony* Affects Pigmentation Divergence and Cuticular Hydrocarbons in *Drosophila americana* and *D. novamexicana*', *Frontiers in Ecology and Evolution*, 8. Available at: https://www.frontiersin.org/articles/10.3389/fevo.2020.00184.

Laurie, C.C. *et al.* (1997) *An Introgression Analysis of Quantitative Trait Loci That Contribute to a Morphological Difference Between Drosophila simulans and D. mauritiana*.

Leal, F. and Cohn, M.J. (2016) 'Loss and Re-emergence of Legs in Snakes by Modular Evolution of *Sonic hedgehog* and *HOXD* Enhancers', *Current Biology*, 26(21), pp. 2966–2973. Available at: https://doi.org/10.1016/j.cub.2016.09.020.

Lemon, B. and Tjian, R. (2000) 'Orchestrated response: a symphony of transcription factors for gene control', *Genes & development*, 14(20), pp. 2551–2569.

LeVasseur-Viens, H., Polak, M. and Moehring, A.J. (2015) 'No evidence for external genital morphology affecting cryptic female choice and reproductive isolation in *Drosophila*', *Evolution*, 69(7), pp. 1797–1807. Available at: https://doi.org/10.1111/evo.12685.

Li, X. *et al.* (1998) 'Generation of destabilized green fluorescent protein as a transcription reporter', *Journal of Biological Chemistry*, 273(52), pp. 34970–34975.

Lin, M. *et al.* (2006) 'Threshold response of *C15* to the Dpp gradient in *Drosophila* is established by the cumulative effect of Smad and Zen activators and negative cues', *Development*, 133(24), pp. 4805–4813. Available at: https://doi.org/10.1242/dev.02689.

Ling, L. *et al.* (2023) 'Increased chromatin accessibility promotes the evolution of a transcriptional silencer in *Drosophila*', *Science Advances*, 9(7), p. eade6529. Available at: https://doi.org/10.1126/sciadv.ade6529.

Liu, J. *et al.* (1996) 'Genetic Analysis of a Morphological Shape Difference in the Male Genitalia of *Drosophila simulans* and *D. mauritiana*', *Genetics*, 142(4), pp. 1129–1145. Available at: https://doi.org/10.1093/genetics/142.4.1129.

López-Varea, A. *et al.* (2021) 'Genome-wide phenotypic RNAi screen in the *Drosophila* wing: phenotypic description of functional classes', *G3 Genes|Genomes|Genetics*, 11(12), p. jkab349. Available at: https://doi.org/10.1093/g3journal/jkab349.

Luecke, D., Rice, G. and Kopp, A. (2022) 'Sex-specific evolution of a *Drosophila* sensory system via interacting cis- and trans-regulatory changes', *Evolution & Development*, 24(1–2), pp. 37–60. Available at: https://doi.org/https://doi.org/10.1111/ede.12398.

Ly, D. *et al.* (2017) 'Asymmetrically deployed actomyosin-based contractility generates a boundary between developing leg segments in *Drosophila*', *Developmental Biology*, 429(1), pp. 165–176. Available at: https://doi.org/https://doi.org/10.1016/j.ydbio.2017.06.031.

Lynch, V.J. and Wagner, G.P. (2008) 'Resurrecting the role of transcription factor change in developmental evolution', *Evolution*, 62(9), pp. 2131–2154. Available at: https://doi.org/10.1111/j.1558-5646.2008.00440.x.

Macdonald, S.J. *et al.* (2022) 'Powerful, efficient QTL mapping in *Drosophila melanogaster* using bulked phenotyping and pooled sequencing', *Genetics*, 220(3), p. iyab238. Available at: https://doi.org/10.1093/genetics/iyab238.

Macdonald, S.J. and Goldstein, D.B. (1999) *A Quantitative Genetic Analysis of Male Sexual Traits Distinguishing the Sibling Species Drosophila simulans and D. sechellia*.

Madeira, F. *et al.* (2022) 'Search and sequence analysis tools services from EMBL-EBI in 2022', *Nucleic acids research*, 50(W1), pp. W276–W279. Available at: https://doi.org/10.1093/nar/gkac240.

Marcellini, S. and Simpson, P. (2006) 'Two or Four Bristles: Functional Evolution of an Enhancer of *scute* in Drosophilidae', *PLOS Biology*, 4(12), pp. e386-. Available at: https://doi.org/10.1371/journal.pbio.0040386.

Martin, A. and Orgogozo, V. (2013) 'The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation', *Evolution*, 67(5), pp. 1235–1250.

Masly, J.P. *et al.* (2011) 'The Genetic Basis of Rapidly Evolving Male Genital Morphology in *Drosophila*', *Genetics*, 189(1), p. 357. Available at: https://doi.org/10.1534/genetics.111.130815.

Masly, J.P. and Kamimura, Y. (2014) 'Asymmetric mismatch in strain-specific genital morphology causes increased harm to *Drosophila* females', *Evolution*, 68(8), pp. 2401–2411. Available at: https://doi.org/10.1111/evo.12436.

Massey, J.H. *et al.* (2020) 'Distinct genetic architectures underlie divergent thorax, leg, and wing pigmentation between *Drosophila elegans* and *D. gunungcola*', *bioRxiv*, p. 2020.06.28.176735. Available at: https://doi.org/10.1101/2020.06.28.176735.

McGregor, A.P. *et al.* (2007) 'Morphological evolution through multiple cis-regulatory mutations at a single gene', *Nature*, 448(7153), pp. 587–590. Available at: https://doi.org/10.1038/nature05988.

McKimmie, C., Woerfel, G. and Russell, S. (2005) 'Conserved genomic organisation of Group B Sox genes in insects.', *BMC Genetics*, 6(1), p. 26. Available at: https://doi.org/10.1186/1471-2156-6-26.

McNeil, C.L., Bain, C.L. and Macdonald, S.J. (2011) 'Multiple Quantitative Trait Loci Influence the Shape of a Male-Specific Genital Structure in *Drosophila melanogaster*', *G3 (Bethesda, Md.)*. 2011/10/01, 1(5), pp. 343–351. Available at: https://doi.org/10.1534/g3.111.000661.

McQueen, E.W. *et al.* (2022) 'A standardized nomenclature and atlas of the female terminalia of *Drosophila melanogaster*', *Fly*, 16(1), pp. 128–151. Available at: https://doi.org/10.1080/19336934.2022.2058309.

Miller, S.A., Dykes, D.D. and Polesky, H.F. (1988) 'A simple salting out procedure for extracting DNA from human nucleated cells.', *Nucleic acids research*, 16(3), p. 1215. Available at: https://doi.org/10.1093/nar/16.3.1215.

Miyazono, K. *et al.* (2010) 'Cooperative DNA-binding and sequence-recognition mechanism of aristaless and clawless', *The EMBO Journal*, 29(9), pp. 1613–1623. Available at: https://doi.org/https://doi.org/10.1038/emboj.2010.53.

Moczek, A.P. (2009) 'On the origins of novelty and diversity in development and evolution: a case study on beetle horns', in *Cold Spring Harbor Symposia on Quantitative Biology*. Citeseer, p. 289.

Moczek, A.P. and Rose, D.J. (2009) 'Differential recruitment of limb patterning genes during development and diversification of beetle horns', *Proceedings of the National Academy of Sciences*, 106(22), pp. 8992–8997. Available at: https://doi.org/10.1073/pnas.0809668106.

Mojica-Vázquez, L.H. *et al.* (2017) 'Tissue-specific enhancer repression through molecular integration of cell signaling inputs', *PLoS genetics*, 13(4), pp. e1006718–e1006718. Available at: https://doi.org/10.1371/journal.pgen.1006718.

Molina-Gil, S. *et al.* (2023) 'Interlocking of co-opted developmental gene networks in *Drosophila* and the evolution of pre-adaptive novelty', *Nature Communications*, 14(1), p. 5730. Available at: https://doi.org/10.1038/s41467-023-41414-3.

Muller, G.B. and Wagner, G.P. (1991) 'Novelty in Evolution: Restructuring the Concept', *Annual Review of Ecology and Systematics*, 22(1), pp. 229–256. Available at: https://doi.org/10.1146/annurev.es.22.110191.001305.

Murugesan, S.N. *et al.* (2022) 'Butterfly eyespots evolved via cooption of an ancestral gene-regulatory network that also patterns antennae, legs, and wings', *Proceedings of the National*

Academy of Sciences, 119(8), p. e2108661119. Available at: https://doi.org/10.1073/pnas.2108661119.

Natori, K. *et al.* (2012) 'Progressive tarsal patterning in the *Drosophila* by temporally dynamic regulation of transcription factor genes', *Developmental Biology*, 361(2), pp. 450–462. Available at: https://doi.org/https://doi.org/10.1016/j.ydbio.2011.10.031.

Newcomb, S. *et al.* (2018) 'cis-regulatory architecture of a short-range EGFR organizing center in the *Drosophila melanogaster* leg', *PLOS Genetics*, 14(8), pp. e1007568-. Available at: https://doi.org/10.1371/journal.pgen.1007568.

Ohto, H. *et al.* (1999) 'Cooperation of Six and Eya in Activation of Their Target Genes through Nuclear Translocation of Eya', *Molecular and Cellular Biology*, 19(10), pp. 6815–6824. Available at: https://doi.org/10.1128/MCB.19.10.6815.

Ordoñez, R. *et al.* (2019) 'DNA Methylation of Enhancer Elements in Myeloid Neoplasms: Think Outside the Promoters?', *Cancers*, 11(10), p. 1424. Available at: https://doi.org/10.3390/cancers11101424.

Overton, P.M. *et al.* (2002) 'Evidence for differential and redundant function of the Sox genes *Dichaete* and *SoxN* during CNS development in *Drosophila*'.

Pagès, H., Aboyoun, P. and Gentleman, R. (2023) 'Biostrings: Efficient manipulation of biological strings', *R package version 2.70.1* [Preprint].

Paré, A.C. *et al.* (2019) 'An LRR Receptor-Teneurin System Directs Planar Polarity at Compartment Boundaries', *Developmental Cell*, 51(2), pp. 208-221.e6. Available at: https://doi.org/10.1016/J.DEVCEL.2019.08.003.

Peluffo, A.E. *et al.* (2015) 'A Major Locus Controls a Genital Shape Difference Involved in Reproductive Isolation Between *Drosophila yakuba* and *Drosophila santomea*', *G3: Genes|Genomes|Genetics*, 5(12), p. 2893. Available at: https://doi.org/10.1534/g3.115.023481.

Pfeiffer, B.D. *et al.* (2010) 'Refinement of tools for targeted gene expression in *Drosophila*', *Genetics*, 186(2), pp. 735–755. Available at: https://doi.org/10.1534/genetics.110.119917.

Phanstiel, D.H. *et al.* (2014) 'Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures', *Bioinformatics (Oxford, England)*, 30(19), pp. 2808–2810. Available at: https://doi.org/10.1093/bioinformatics/btu379.

Pignoni, F. *et al.* (1997) 'The eye-specification proteins So and Eya form a complex and regulate multiple steps in *Drosophila* eye development', *Cell*, 91(7), pp. 881–891.

Posnien, N. *et al.* (2012) 'Evolution of Eye Morphology and Rhodopsin Expression in the *Drosophila melanogaster* Species Subgroup', *PLOS ONE*, 7(5), pp. e37346-. Available at: https://doi.org/10.1371/journal.pone.0037346.

Preger-Ben Noon, E. *et al.* (2017) 'Pleiotropy in enhancer function is encoded through diverse genetic architectures', *bioRxiv*, p. 188532. Available at: https://doi.org/10.1101/188532.

Prud'homme, B. *et al.* (2006) 'Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene', *Nature*, 440(7087), pp. 1050–1053.

Pueyo, J.I. *et al.* (2000) 'Proximal-distal leg development in *Drosophila* requires the *apterous* gene and the Lim1 homologue *dlim1*', *Development*, 127(24), pp. 5391–5402. Available at: https://doi.org/10.1242/dev.127.24.5391.

Pueyo, J.I. and Couso, J.P. (2004) 'Chip-mediated partnerships of the homeodomain proteins Bar and Aristaless with the LIM-HOM proteins Apterous and Lim1 regulate distal leg development', *Development*, 131(13), pp. 3107–3120. Available at: https://doi.org/10.1242/dev.01161.

Quiquand, M. *et al.* (2021) 'New regulators of *Drosophila* eye development identified from temporal transcriptome changes', *Genetics*, 217(4). Available at: https://doi.org/10.1093/genetics/iyab007.

R Core Team (2020) 'R: A Language and Environment for Statistical Computing'. Vienna, Austria.

Rane, R. V *et al.* (2019) 'Genomic changes associated with adaptation to arid environments in cactophilic *Drosophila* species', *BMC Genomics*, 20(1), p. 52. Available at: https://doi.org/10.1186/s12864-018-5413-3.

Rebeiz, M. *et al.* (2009) 'Stepwise Modification of a Modular Enhancer Underlies Adaptation in a *Drosophila* Population', *Science*, 326(5960), pp. 1663–1667. Available at: https://doi.org/10.1126/science.1178357.

Rebeiz, M. and Tsiantis, M. (2017) 'Enhancer evolution and the origins of morphological novelty', *Current Opinion in Genetics & Development*, 45, pp. 115–123. Available at: https://doi.org/https://doi.org/10.1016/j.gde.2017.04.006.

Rebora, M. *et al.* (2021) 'Attachment devices and the tarsal gland of the bug *Coreus marginatus* (Hemiptera: Coreidae)', *Zoomorphology*, 140(1), pp. 85–102. Available at: https://doi.org/10.1007/s00435-020-00515-z.

Reinius, B. and Sandberg, R. (2015) 'Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation', *Nature Reviews Genetics*, 16(11), pp. 653–664.

Ren, X. *et al.* (2013) 'Optimized gene editing technology for *Drosophila melanogaster* using germ line-specific Cas9', *Proceedings of the National Academy of Sciences*, 110(47), pp. 19012–19017. Available at: https://doi.org/10.1073/pnas.1318481110.

Rice, G. *et al.* (2019) 'A standardized nomenclature and atlas of the male terminalia of *Drosophila melanogaster*', *Fly*, 13(1–4), pp. 51–64. Available at: https://doi.org/10.1080/19336934.2019.1653733.

Rice, G.R. *et al.* (2019) 'Modular tissue-specific regulation of *doublesex* underpins sexually dimorphic development in *Drosophila*', *Development*, 146(14), p. dev178285.

Richter, W.F. *et al.* (2022) 'The Mediator complex as a master regulator of transcription by RNA polymerase II', *Nature Reviews Molecular Cell Biology*, 23(11), pp. 732–749.

Riddle, R.D. *et al.* (1993) 'Sonic hedgehog mediates the polarizing activity of the ZPA', *Cell*, 75(7), pp. 1401–1416.

Ridgway, A.M. *et al.* (2024) '*Sox21b* underlies the rapid diversification of a novel male genital structure between *Drosophila* species', *Current Biology* [Preprint]. Available at: https://doi.org/10.1016/j.cub.2024.01.022.

Rizzo, N.P. and Bejsovec, A. (2017) 'SoxNeuro and Shavenbaby act cooperatively to shape denticles in the embryonic epidermis of *Drosophila*', *Development*, 144(12), pp. 2248–2258. Available at: https://doi.org/10.1242/dev.150169.

Robertson, H.M. (1988) 'Mating asymmetries and phylogeny in the *Drosophila melanogaster* species complex'.

Robinson, J.T. *et al.* (2011) 'Integrative genomics viewer', *Nature Biotechnology*, 29(1), pp. 24–26. Available at: https://doi.org/10.1038/nbt.1754.

Roignant, J.-Y. *et al.* (2010) 'The transcriptional co-factor Chip acts with LIM-homeodomain proteins to set the boundary of the eye field in *Drosophila*', *Development*, 137(2), pp. 273–281. Available at: https://doi.org/10.1242/dev.041244.

Rosales-Vega, M. *et al.* (2018) 'The role of the trithorax group *TnaA* isoforms in Hox gene expression, and in *Drosophila* late development', *PLOS ONE*, 13(10), pp. e0206587-. Available at: https://doi.org/10.1371/journal.pone.0206587.

Rothier, P.S. *et al.* (2023) 'Mammalian forelimb evolution is driven by uneven proximal-to-distal morphological diversity', *eLife*. Edited by M.E. Bronner, D.M. Grossnickle, and F. Gould, 12, p. e81492. Available at: https://doi.org/10.7554/eLife.81492.

Rowe, L. and Arnqvist, G. (2012) 'Sexual selection and the evolution of genital shape and complexity in water striders', *Evolution*, 66(1), pp. 40–54. Available at: https://doi.org/https://doi.org/10.1111/j.1558-5646.2011.01411.x.

Ruiz-Losada, M. *et al.* (2018) 'Specification and patterning of *Drosophila* appendages', *Journal of Developmental Biology*. MDPI Multidisciplinary Digital Publishing Institute. Available at: https://doi.org/10.3390/jdb6030017.

Sakurai, K.T. *et al.* (2007) 'Differential control of cell affinity required for progression and refinement of cell boundary during *Drosophila* leg segmentation', *Developmental Biology*, 309(1), pp. 126–136. Available at: https://doi.org/https://doi.org/10.1016/j.ydbio.2007.07.001.

Sánchez, L. *et al.* (1997) 'The genital disc of *Drosophila melanogaster*: II. Role of the genes *hedgehog*, *decapentaplegic* and *wingless*', *Development Genes and Evolution*, 207, pp. 229–241.

Sánchez, L. and Guerrero, I. (2001) 'The development of the *Drosophila* genital disc', *BioEssays*, 23(8), pp. 698–707. Available at: https://doi.org/https://doi.org/10.1002/bies.1099.

Sanchez-Herrero, E. *et al.* (1985) 'Anatomy and function of the bithorax complex of *Drosophila*', in *Cold Spring Harbor Symposia on Quantitative Biology*. Cold Spring Harbor Laboratory Press, pp. 165–172.

Schindelin, J. *et al.* (2012) 'Fiji: an open-source platform for biological-image analysis', *Nature Methods*, 9(7), pp. 676–682. Available at: https://doi.org/10.1038/nmeth.2019.

Schmitt, C. and Betz, O. (2017) 'Morphology and ultrastructure of the tarsal adhesive organs of the Madagascar hissing cockroach *Gromphadorhina portentosa*', *Cell and Tissue Research*, 370(2), pp. 243–265. Available at: https://doi.org/10.1007/s00441-017-2661-5.

Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012) 'NIH Image to ImageJ: 25 years of image analysis', *Nature Methods*, 9(7), pp. 671–675. Available at: https://doi.org/10.1038/nmeth.2089.

Schneitz, K., Spielmann, P. and Noll, M. (1993) 'Molecular genetics of *aristaless*, a prd-type homeo box gene involved in the morphogenesis of proximal and distal pattern elements in of appendages in *Drosophila*'.

Schwarzkopf, M. *et al.* (2021) 'Hybridization chain reaction enables a unified approach to multiplexed, quantitative, high-resolution immunohistochemistry and in situ hybridization', *Development*, 148(22), p. dev199847. Available at: https://doi.org/10.1242/dev.199847.

Serrano, N. and O'Farrell, P.H. (1997) 'Limb morphogenesis: connections between patterning and growth', *Current Biology*, 7(3), pp. R186–R195. Available at: https://doi.org/10.1016/S0960-9822(97)70085-X.

Shapiro, S.S. and Wilk, M.B. (1965) 'An analysis of variance test for normality (complete samples)†', *Biometrika*, 52(3–4), pp. 591–611. Available at: https://doi.org/10.1093/biomet/52.3-4.591.

Shellenbarger, D.L. and Mohler, J.D. (1978) 'Temperature-sensitive periods and autonomy of pleiotropic effects of *l(1)Nts1*, a conditional notch lethal in *Drosophila*', *Developmental Biology*, 62(2), pp. 432–446. Available at: https://doi.org/https://doi.org/10.1016/0012-1606(78)90226-9.

Shlyueva, D., Stampfel, G. and Stark, A. (2014) 'Transcriptional enhancers: from properties to genome-wide predictions', *Nature Reviews Genetics*, 15(4), pp. 272–286. Available at: https://doi.org/10.1038/nrg3682.

Shubin, N., Tabin, C. and Carroll, S. (2009) 'Deep homology and the origins of evolutionary novelty', *Nature*, 457(7231), pp. 818–823. Available at: https://doi.org/10.1038/nature07891.

Signor, S.A. *et al.* (2016) 'Genetic Convergence in the Evolution of Male-Specific Color Patterns in *Drosophila*', *Current Biology*, 26(18), pp. 2423–2433. Available at: https://doi.org/https://doi.org/10.1016/j.cub.2016.07.034.

Del Signore, S.J., Hayashi, T. and Hatini, V. (2012) 'odd-skipped genes and lines organize the notum anterior–posterior axis using autonomous and non-autonomous mechanisms', *Mechanisms of Development*, 129(5), pp. 147–161. Available at: https://doi.org/https://doi.org/10.1016/j.mod.2012.05.001.

Sinclair, A.H. *et al.* (1990) 'A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif', *Nature*, 346(6281), pp. 240–244.

Skene, P.J. and Henikoff, S. (2017) 'An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites', *eLife*. Edited by D. Reinberg, 6, p. e21856. Available at: https://doi.org/10.7554/eLife.21856.

Smith, J.M. and Sondhi, K.C. (1961) 'The Arrangement of Bristles in *Drosophila*', *Development*, 9(4), pp. 661–672. Available at: https://doi.org/10.1242/dev.9.4.661.

Smith, S.J., Davidson, L.A. and Rebeiz, M. (2020) 'Evolutionary expansion of apical extracellular matrix is required for the elongation of cells in a novel structure', *eLife*. Edited by A. Rokas and M.B. Eisen, 9, p. e55965. Available at: https://doi.org/10.7554/eLife.55965.

Soriano, N.S. and Russell, S. (1998) 'The *Drosophila* SOX-domain protein Dichaete is required for the development of the central nervous system midline', *Development*, 125(20), pp. 3989–3996. Available at: https://doi.org/10.1242/dev.125.20.3989.

Spencer, F.A., Hoffmann, F.M. and Gelbart, W.M. (1982) 'Decapentaplegic: A gene complex affecting morphogenesis in *Drosophila melanogaster*', *Cell*, 28(3), pp. 451–461. Available at: https://doi.org/https://doi.org/10.1016/0092-8674(82)90199-4.

Steiner, C.C., Weber, J.N. and Hoekstra, H.E. (2007) 'Adaptive Variation in Beach Mice Produced by Two Interacting Pigmentation Genes', *PLOS Biology*, 5(9), pp. e219-. Available at: https://doi.org/10.1371/journal.pbio.0050219.

Stern, D.L. (2011) *Evolution, Development, and the Predictable Genome*. W. H. Freeman.

Stern, D.L. (2014) 'Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizygosity test', *Trends in Genetics*, 30(12), pp. 547–554. Available at: https://doi.org/https://doi.org/10.1016/j.tig.2014.09.006.

Stern, D.L. *et al.* (2017) 'Genetic and Transgenic Reagents for *Drosophila simulans*, *D. mauritiana*, *D. yakuba*, *D. santomea*, and *D. virilis*', *G3 Genes|Genomes|Genetics*, 7(4), pp. 1339–1347. Available at: https://doi.org/10.1534/g3.116.038885.

Stern, D.L. and Frankel, N. (2013) 'The structure and evolution of cis-regulatory regions: the *shavenbaby* story', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1632), p. 20130028. Available at: https://doi.org/10.1098/rstb.2013.0028.

Stern, D.L. and Sucena, E. (2012) 'Rapid Mounting of Adult *Drosophila* Structures in Hoyer's Medium', *Cold Spring Harbor Protocols*, 2012(1), p. pdb.prot067371. Available at: http://cshprotocols.cshlp.org/content/2012/1/pdb.prot067371.abstract.

Stieper, B.C. *et al.* (2008) 'Imaginal discs regulate developmental timing in *Drosophila melanogaster*', *Developmental Biology*, 321(1), pp. 18–26. Available at: https://doi.org/10.1016/J.YDBIO.2008.05.556.

Sucena, E. *et al.* (2003) 'Regulatory evolution of *shavenbaby/ovo* underlies multiple cases of morphological parallelism', *Nature*, 424(6951), pp. 935–938. Available at: https://doi.org/10.1038/nature01768.

Sucena, É. and Stern, D.L. (2000) 'Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by cis-regulatory evolution of *ovo/shaven-baby*', *Proceedings of the National Academy of Sciences*, 97(9), pp. 4530–4534.

Tabin, C.J., Carroll, S.B. and Panganiban, G. (1999) 'Out on a Limb Parallels in Vertebrate and Invertebrate Limb Patterning and the Origin of Appendages', *American Zoologist*, 39(3), pp. 650–663. Available at: https://doi.org/10.1093/icb/39.3.650.

Tanaka, K. *et al.* (2011) 'Evolution of Sex-Specific Traits through Changes in HOX-Dependent *doublesex* Expression', *PLOS Biology*, 9(8), pp. e1001131-. Available at: https://doi.org/10.1371/journal.pbio.1001131.

Tanaka, K.M. *et al.* (2015) 'Genetic Architecture and Functional Characterization of Genes Underlying the Rapid Diversification of Male External Genitalia Between *Drosophila simulans* and *Drosophila mauritiana*', *Genetics*, 200(1), p. 357. Available at: https://doi.org/10.1534/genetics.114.174045.

Tanaka, K.M., Kamimura, Y. and Takahashi, A. (2018) 'Mechanical incompatibility caused by modifications of multiple male genital structures using genomic introgression in *Drosophila*', *Evolution*, 72(11), pp. 2406–2418.

Tavsanli, B.C. *et al.* (2004) 'Structure–function analysis of the *Drosophila* retinal determination protein Dachshund', *Developmental Biology*, 272(1), pp. 231–247. Available at: https://doi.org/https://doi.org/10.1016/j.ydbio.2004.05.005.

Thurmond, J. *et al.* (2018) 'FlyBase 2.0: the next generation', *Nucleic Acids Research*, 47(D1), pp. D759–D765. Available at: https://doi.org/10.1093/nar/gky1003.

True, J.R. *et al.* (1997) 'Quantitative Genetic Analysis of Divergence in Male Secondary Sexual Traits Between *Drosophila simulans* and *Drosophila mauritiana*', *Evolution*, 51(3), pp. 816–832. Available at: https://doi.org/10.2307/2411157.

True, J.R. and Haag, E.S. (2001) 'Developmental system drift and flexibility in evolutionary trajectories', *Evolution & development*, 3(2), pp. 109–119.

Tsai, A., Alves, M.R.P. and Crocker, J. (2019) 'Multi-enhancer transcriptional hubs confer phenotypic robustness', *eLife*, 8. Available at: https://doi.org/10.7554/eLife.45325.

Tse J *et al.* (2022) 'Single-Cell Atlas of the *Drosophila* Leg Disc Identifies a Long Non-Coding RNA in Late Development', *International Journal of Molecular Sciences*, 23(12).

Tsuji, T. *et al.* (2000) 'Requirements of *Lim1,* a *Drosophila* LIM-homeobox gene, for normal leg and antennal development', *Development*, 127(20), pp. 4315–4323. Available at: https://doi.org/10.1242/dev.127.20.4315.

Untergasser, A. (2006) *Cloning – Classic LR-Reaction II*, *Untergasser's Lab*.

Vedelek, V. *et al.* (2018) 'Analysis of *Drosophila melanogaster* testis transcriptome', *BMC Genomics*, 19(1), p. 697. Available at: https://doi.org/10.1186/s12864-018-5085-z.

Veling, M.W. *et al.* (2019) 'Identification of Neuronal Lineages in the *Drosophila* Peripheral Nervous System with a "Digital" Multi-spectral Lineage Tracing System', *Cell Reports*, 29(10), pp. 3303-3312.e3. Available at: https://doi.org/https://doi.org/10.1016/j.celrep.2019.10.124.

Vickrey, A.I. *et al.* (2018) 'Introgression of regulatory alleles and a missense coding mutation drive plumage pattern diversity in the rock pigeon', *eLife*. Edited by A. Rokas, 7, p. e34803. Available at: https://doi.org/10.7554/eLife.34803.

Villar, D. *et al.* (2015) 'Enhancer Evolution across 20 Mammalian Species', *Cell*, 160(3), pp. 554–566. Available at: https://doi.org/https://doi.org/10.1016/j.cell.2015.01.006.

Vincent, B.J. *et al.* (2019) 'An Atlas of Transcription Factors Expressed in Male Pupal Terminalia of *Drosophila melanogaster*', *G3 Genes|Genomes|Genetics*, 9(12), pp. 3961–3972. Available at: https://doi.org/10.1534/g3.119.400788.

Wasik, B.R., Rose, D.J. and Moczek, A.P. (2010) 'Beetle horns are regulated by the Hox gene, *Sex combs reduced*, in a species-and sex-specific manner', *Evolution & development*, 12(4), pp. 353–362.

Wasserman, W.W. and Sandelin, A. (2004) 'Applied bioinformatics for the identification of regulatory elements', *Nature Reviews Genetics*, 5(4), pp. 276–287.

Waymack, R., Gad, M. and Wunderlich, Z. (2021) 'Molecular competition can shape enhancer activity in the *Drosophila* embryo', *iScience*, 24(9). Available at: https://doi.org/10.1016/j.isci.2021.103034.

Weatherbee, S.D. *et al.* (1999) 'Ultrabithorax function in butterfly wings and the evolution of insect wing patterns', *Current Biology*, 9(3), pp. 109–115.

Weirauch, M.T. *et al.* (2014) 'Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity', *Cell*, 158(6), pp. 1431–1443. Available at: https://doi.org/10.1016/j.cell.2014.08.009.

Wickham, H. *et al.* (2022) 'svglite: An "SVG" Graphics Device'. R package version 2.1.0.

Williams, T.M. *et al.* (2008) 'The Regulation and Evolution of a Genetic Switch Controlling Sexually Dimorphic Traits in *Drosophila*', *Cell*, 134(4), pp. 610–623. Available at: https://doi.org/10.1016/j.cell.2008.06.052.

Wittkopp, P.J. *et al.* (2009) 'Intraspecific Polymorphism to Interspecific Divergence: Genetics of Pigmentation in *Drosophila*', *Science*, 326(5952), p. 540. Available at: https://doi.org/10.1126/science.1176980.

Wittkopp, P.J. and Beldade, P. (2009) 'Development and evolution of insect pigmentation: genetic mechanisms and the potential consequences of pleiotropy', in *Seminars in cell & developmental biology*. Elsevier, pp. 65–71.

Wooldridge, T.B. *et al.* (2022) 'An enhancer of *Agouti* contributes to parallel evolution of cryptically colored beach mice', *Proceedings of the National Academy of Sciences*, 119(27), p. e2202862119. Available at: https://doi.org/10.1073/pnas.2202862119.

Wray, G.A. *et al.* (2003) 'The Evolution of Transcriptional Regulation in Eukaryotes', *Molecular Biology and Evolution*, 20(9), pp. 1377–1419. Available at: https://doi.org/10.1093/MOLBEV/MSG140.

Xie, Y. (2022) 'knitr: A General-Purpose Package for Dynamic Report Generation in R'. R package version 1.39.

Yan, F., Powell, D.R. and Curtis, D.J. (2020) 'From reads to insight: a hitchhiker's guide to ATAC-seq data analysis', *Genome Biology*, 21(22). Available at: https://doi.org/https://doi.org/10.1186/s13059-020-1929-3.

Yassin, A. *et al.* (2016) 'The *pdm3* Locus Is a Hotspot for Recurrent Evolution of Female-Limited Color Dimorphism in *Drosophila*', *Current Biology*, 26(18), pp. 2412–2422. Available at: https://doi.org/10.1016/j.cub.2016.07.016.

Yassin, A. and Orgogozo, V. (2013) 'Coevolution between Male and Female Genitalia in the *Drosophila melanogaster* Species Subgroup', *PLoS ONE*. Edited by D.J. Colgan, 8(2), p. e57158. Available at: https://doi.org/10.1371/journal.pone.0057158.

Younger, M.A. *et al.* (2022) 'Non-Canonical Odor Coding in the Mosquito', *bioRxiv*, p. 2020.11.07.368720. Available at: https://doi.org/10.1101/2020.11.07.368720.

Zeitlinger, J. (2020) 'Seven myths of how transcription factors read the cis-regulatory code', *Current Opinion in Systems Biology*, 23, pp. 22–31. Available at: https://doi.org/https://doi.org/10.1016/j.coisb.2020.08.002.

Zeng, Z.B. *et al.* (2000) 'Genetic architecture of a morphological shape difference between two *Drosophila* species', *Genetics*, 154(1), pp. 299–310. Available at: https://pubmed.ncbi.nlm.nih.gov/10628989.

Zhang, L., Mazo-Vargas, A. and Reed, R.D. (2017) 'Single master regulatory gene coordinates the evolution and development of butterfly color and iridescence', *Proceedings of the National Academy of Sciences*, 114(40), pp. 10707–10712. Available at: https://doi.org/10.1073/pnas.1709058114.

Zhao, Y., Zheng, D. and Cvekl, A. (2019) 'Profiling of chromatin accessibility and identification of general cis-regulatory mechanisms that control two ocular lens differentiation pathways', *Epigenetics & Chromatin*, 12(1), p. 27. Available at: https://doi.org/10.1186/s13072-019-0272-y.

Zhong, L. *et al.* (2011) 'Parallel Expansions of Sox Transcription Factor Group B Predating the Diversifications of the Arthropods and Jawed Vertebrates', *PLOS ONE*, 6(1), pp. e16570-. Available at: https://doi.org/10.1371/journal.pone.0016570.

Zhu, H. (2021) 'kableExtra: Construct Complex Table with "kable" and Pipe Syntax'. R package version 1.3.4.

8. Supplementary Materials

8.1 Chapter Two Supplementary Materials

Experiment	Primary Antibody Dilution and Source	Secondary Antibody Dilution
D. mel male genital discs	Rabbit Anti-Drop 1:250, Chris Doe (University of Oregon, USA)	Goat Anti-Rabbit 488 1:600
D. mel male pupal terminalia	Rabbit Anti-Drop 1:250, Chris Doe (University of Oregon, USA)	Goat Anti-Rabbit 488, 1:400
D. mel embryos and male pupal terminalia	Mouse Anti-Eya 1:100, DSHB(eya10H6)	Donkey Anti-Mouse 647 1:400
D. mel male genital discs and pupal terminalia	Rabbit Anti-Dichaete* 1:400, Steve Russell (University of Cambridge, UK)	Goat Anti-Rabbit 488 1:600
D. mel male and female imaginal discs, and pupal terminalia.	Chicken Anti-GFP 1:200, Thermofisher	Goat Anti-chicken 488 1:400
D. mel male genital discs and pupal terminalia	Rabbit Anti-Opa* 1:50, John Gergen (Stony Brook University, USA)	Goat Anti-Rabbit 488 1:400
D. mel, D. mau, D. sim male and female imaginal discs, and pupal terminalia.	Rat Anti-C15 1:200, Gerard Campbell (University of Pittsburgh, USA)	Goat Anti-Rat 647 / Goat Anti-Rat 488
D. mel male and female imaginal discs, and pupal terminalia	Rat Anti-Al 1:500, Gerard Campbell (University of Pittsburgh, USA)	Goat Anti-Rat 647 / Goat Anti-Rat 488

D. mel male and female genital discs and pupal terminalia	Rabbit Anti-Lim1* 1:100, Juan Botas (Baylor College of Medicine, USA)	Goat Anti-Rabbit 488
D. mel male pupal terminalia	Guinea Pig Anti-Dll* 1:50, Richard Mann (Columbia University, USA)	Goat Anti-Guinea Pig 594
D. mel male pupal terminalia	Mouse Anti-AbdB* 1:50, DSHB (1A2E9)	Donkey Anti-Mouse 647 1:400
D. mel male pupal terminalia	Rat Anti-DCAD2 1:50 DSHB	Goat Anti-Rat 647 / Goat Anti-Rat 488 1:400

Supplementary Table 1: Antibody staining conditions used for staining. * primary antibody dilutions preabsorbed before use.

Oligonucleotide	**Sequence (5' – 3')**
Sox21b sense gRNA	GCAGCAGCAACAATCCGACCA
Sox21b antisense gRNA	TGGTCGGATTGTTGCTGCTGC
D. simulans and *IL108 Sox21b* Right HA F primer	CCAGGGACTCGGCCACTCG
D. simulans Sox21b Right HA R primer	TTTCGTGGCTTGCGTTACAC
D. simulans Sox21b Left HA F primer	CTTTCCATTATGCGACGGGG
D. simulans and *IL108 Sox21b* Left HA R primer	TCGGATTGTTGCTGCTGCTG
IL108 Sox21b Right HA R primer	CGAAAGAAACGTTGCCACCA
IL108 Sox21b Left HA F primer	GCTTGTTTTGGACACGCTGG
IL108 Sox21b exon 1-2 F primer	CTGAAACCGTGCTAAAGGCG
IL108 Sox21b exon 1-2 R primer	AGGCAAACACAATTCAACAGG
D. simulans Sox21b exon 1-2 F primer	TTGAAGGGGCAATTGAGGCA
D. simulans Sox21b exon 1-2 R primer	ACGTATCTGATCATTTCCTTTAAGC
D. simulans Enhancer 3 minimal promoter F primer	TTTCTCCCCACCGCCGGTACGATAAACGGCCGGCCGAG
D. melanogaster Enhancer 3 minimal promoter F primer	TGTCTCCCCACCTCCGGTACGATAAACGGCCGGCCGAG
D. mauritiana Enhancer 3 minimal promoter F primer	TGTCCCCCCACCGCCGGTACGATAAACGGCCGGCCGAG
D. melanogaster AbdB mutant Enhancer 3 minimal promoter F primer	TGTCTCCCCACCTCCGGTACGATAAACGGCCGGCCGAG
Universal minimal promoter R primer	CGGGATCCCGGATCTGGTACGTTTGGTATGCGTCTTGTGATTCAAAGT
HSP70 R primer	CGGTCGAGGGTTCGAAATCGATAAG

D. simulans Enhancer 3	CCCATTGTACCTCATGACCGCTTGTGCCTCGACTGTCTC TGATTGAATCCACTATTTATGCGCTTAAGGTTAATCAAAGC CCATTCCAGCAATGCCTTAACACTTGGACAGCGAGAGGG AGACGACAACAGAGGCGCGAGGGAGACAGGAGACAGG CAGACTAACAGACAGAACGGGAGGCCGGCAGAAACAAT TGGAATCTTTTTGGTCGTGTCATAAATGCCAATGAACATAT GGGGAAATATGAAAAGAACGACCGCAGGATATGCTGGCC CAGTGCCATCCATAAATAATCAGATTTTATGACTTTCGCG GGTTCTTCTGGGGATTAACCAAGAAAAGAATCCGGTGTA CACATATATATTTATATATGGGGCACCATGGGGCGCCAAC CGATTGTCCATACTCAACGGCCTGTTTGTTCTTATTAAAA GAGAAGGCGGCCGCTACACAAAAATCTAAAAAATGTTAAT GAAAATCCACACCAGCGAGATGAGAAGTTATGCAGCTTC GGACCATATCTCTAATACCTTTTTATTGCGACTGGCATATA ACTTTTTAATCTATATGAAACCTAATATAGTCTTAAAATATTA AATTACGATTGTATGTACTACAGAATTTATAAAATATAAAAA AGGGATAATGGAAATCTTTGTTTAAATAAATTTAAGATTTA CTTTACCTAGGAACAAGGTCTGCGCCAGTTTTAAAAACTA TTTTTCTTTAATTATAATCTTATATTAAACTGAAAGATTATTT GGGTATCAATTGTTACTTCCCGATATAAACCCTATATCTTT TCTCCCCACCGCC

D. mauritiana Enhancer 3	CCCATAGTACCTCATGACCGCTTGTGCCTCGACTGTCTC TGATTGAATCCACTATTTATGCGCTTAAGGTTAATCAAAGC CCATTCCAGCAATGCCTTAACACTTGGACAACGAGAGGG AGACGACAACAGAGGCGAGAGGGAGACAGTAGACAGG CAGACTAACAGACAGAACGGGAGGCCGGCAGAAACAAT TGGAATCTTTTTGGTCGTGTCATAAATGCCAATGAACATAT GGGGAAATATGAAAAGAACGGCCGCAGGATATGCTGGC CCAGTGCCATCCATAAATAATCAGATTTTATGACTTTCGC GGGTTCTTCTGGGGATTAACCAAGAAAAGAATCCGGTGT ACACATATATATTTATATATGGGGCACCATGGGGCGCCAA CCGATTGTCCATACTCAACGGCCTGTTTGTTCTTATTAAA AGAGAAGGCGGCCGCTACACAAAAATCTAAAAAATGTTA ATGAAAATCCACACCAGCGATATGAGAAGTTATGCTGCTT CTCTAATACCTTTTTATTGCGGCTGGCATATAACTTTTTTA ATCTATATGAAACCTAATATAGTCTGAAAATATTTAAATTAC GATTGTTTGTACTACAGAATTTATAAAATATAAAGAAGGGA TAATGGAAATCTTTGTTTAAATAAATTTAAGATTTAATTTGC CTAGGAACAAGGTTTGCGCCAGTTTTAAAAACTATTTTTT CTTTAATTATAATCTTATATTTAACTGAAAGGCTATTTGTGT ATCAATTGTTACTTCCCGATATAAATCCCTATATCTTTGTCC CCCCACCGCC

D. melanogaster Enhancer 3	CCCACTGTACCTCATGACCGCTTGTGCCTCGACTGTCTC TGATTGAATCCACTATTTATGCGCTTAAGGTTAATCAAAGC CCATTCCAGCAATGCCTTAACACTTGGACAACGAGAGGG AGACGACAACAGAGGCGAGAGGGAGACAGGAGACAGG CAGGAGACTAACAGACAGAACGGGAGGCCGGCAGAAA CAATTGGAATCTTTTTGGTCGTGTCATAAATGCCAATGAA CATACATATGGGGAAATATGAAAAGAACGGCCGCAGGAT ATGCTGGCCCAGTGCCATCCATAAATAATCAGATTTTATG ACTTTCGCGGGTTCTTCTGGGGATTAACCAAGAAAAGAA TCCGGTGTACACATATATATTTATATATGGGGCACCATGGG GCGCCAACCGATTGTCCATACTCAACGGCCTGTTTGTTC TTATTAAAAGAGAAGGCGGCCGCTACACAAAAATCTAAAA AATGTTAATGAAAATCCACACCAGCGAGATGAGAAGTTGT GCAGCTTCAGACCATATCTCTAATACCTTTTTATTGCGGC TGGCATATAACACTTTTAATCTATATGAAACCTAATATCGTA CTAAAATATTAAATAACGTTTGCACTATAGAATTTATAAAAT AAAAAAGGGATAATGGAAATCTTTGTTTAAATAAATTTAAG ATTTAATTTTCCCCAGAATAAGGTATGTGCCAGTCTTAAAA ACTATTTTTCTTTAATTATAATCATATATAAAACTGAAATGCT CTATGTAGGTATCAATTGTTATAAATCCCTATATCTTTGTCT CCCCACCTCC

D. melanogaster Enhancer 3 Abd-B mutant	CCCACTGTACCTCATGACCGCTTGTGCCTCGACTGTCTC TGATTGAATCCACGGGGGATGGGGGGGAGGTTAATCAA AGCCCATTCCAGCAATGCCTTAACACTTGGACAACGAGA GGGAGACGACAACAGAGGCGAGAGGGAGACAGGAGAC AGGCAGGAGACTAACAGACAGAACGGGAGGCCGGCAG AAACAATTGGAATCTTTTTGGTCGTGTCATAAATGCCAAT GAACATACATATGGGGAAATATGAAAAGAACGGCCGCAG GATATGCTGGCCCAGTGCCATCCATAAATAATCAGGGGG GATGGGGGGGGCGGGTTCTTCTGGGGATTAACCAAGAA AAGAATCCGGTGTACACATATATATTTATATATGGGGCACC ATGGGGCGCCAACCGATTGTCCATACTCAACGGCCTGTT TGTTCTTATTAAAAGAGAAGGCGGCCGCTACACAAAAATC TAAAAAATGTTAATGAAAATCCACACCAGCGAGATGAGAA GTTGTGCAGCTTCAGACCATATCTCTAATACCGGGGGAT GGGGGCTGGCATATAACACTTTTAATCTATATGAAACCTAA TATCGTACTAAAATATTAAATAACGTTTGCACTATAGAATTT ATAAAATAAAAAAGGGATAATGGAAATCTTTGTTTAAATAA ATTTAAGATTTAATTTTCCCCAGAATAAGGTATGTGCCAGT CTTAAAAACTATTTTTCTTTAATTATAATCATATATAAAACTG AAATGCTCTATGTAGGTATCAATTGTTATAAATCCCTATATC TTTGTCTCCCCACCTCC

Supplementary Table 2: Oligonucleotide sequences used to generate the *Sox21b-DsRed* reagents and the *Sox21b* Enhancer 3 GAL4 constructs.

See supplementary spreadsheet for details.

Supplementary Table 3: Statistical test results for Chapter 3 and 4 RNAi knockdowns phenotypic analysis.

See supplementary spreadsheet for details.

Supplementary Table 4: Statistical test results for Chapter 3 *Sox21b* reciprocal hemizygotes phenotypic analysis.

See supplementary spreadsheet for details.

Supplementary Table 5: EFA descriptors for *Sox21b* reciprocal hemizygotes posterior lobe shape analysis.

See supplementary spreadsheet for details.

Supplementary Table 6: EFA descriptors for *D. mauritiana* and *D. simulans* posterior lobe shape analysis.

See supplementary spreadsheet for details.

Supplementary Table 7: Statistical test results for Chapter 5 male RNAi knockdown and UAS-OE phenotypic analysis.

See supplementary spreadsheet for details.

Supplementary Table 8: Statistical test results for Chapter 5 female RNAi knockdown and UAS-OE phenotypic analysis.

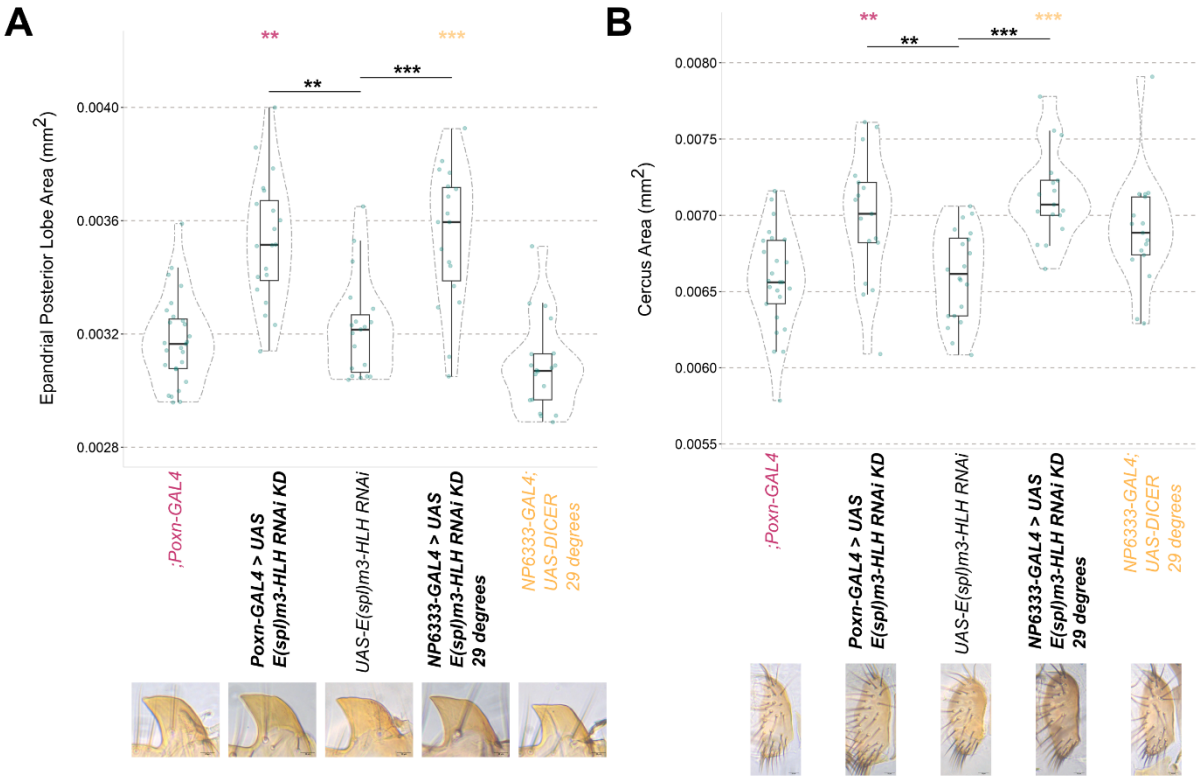
See supplementary spreadsheet for details.

Supplementary Table 9: TFs used to scan *C15* intron 2 Enhancer 1 sequence for binding sites

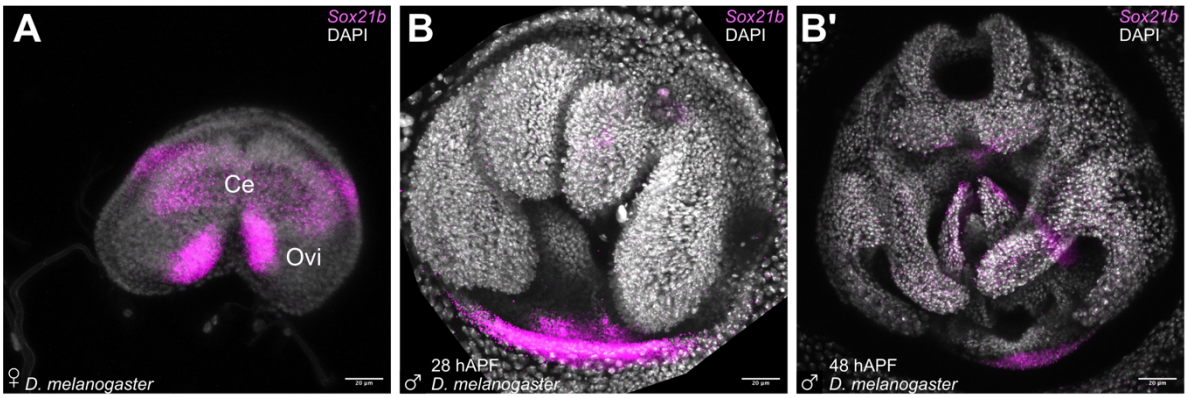
See supplementary spreadsheet for details.

Supplementary Table 10: TF binding sites predicted within *C15* intron 2 enhancer 1

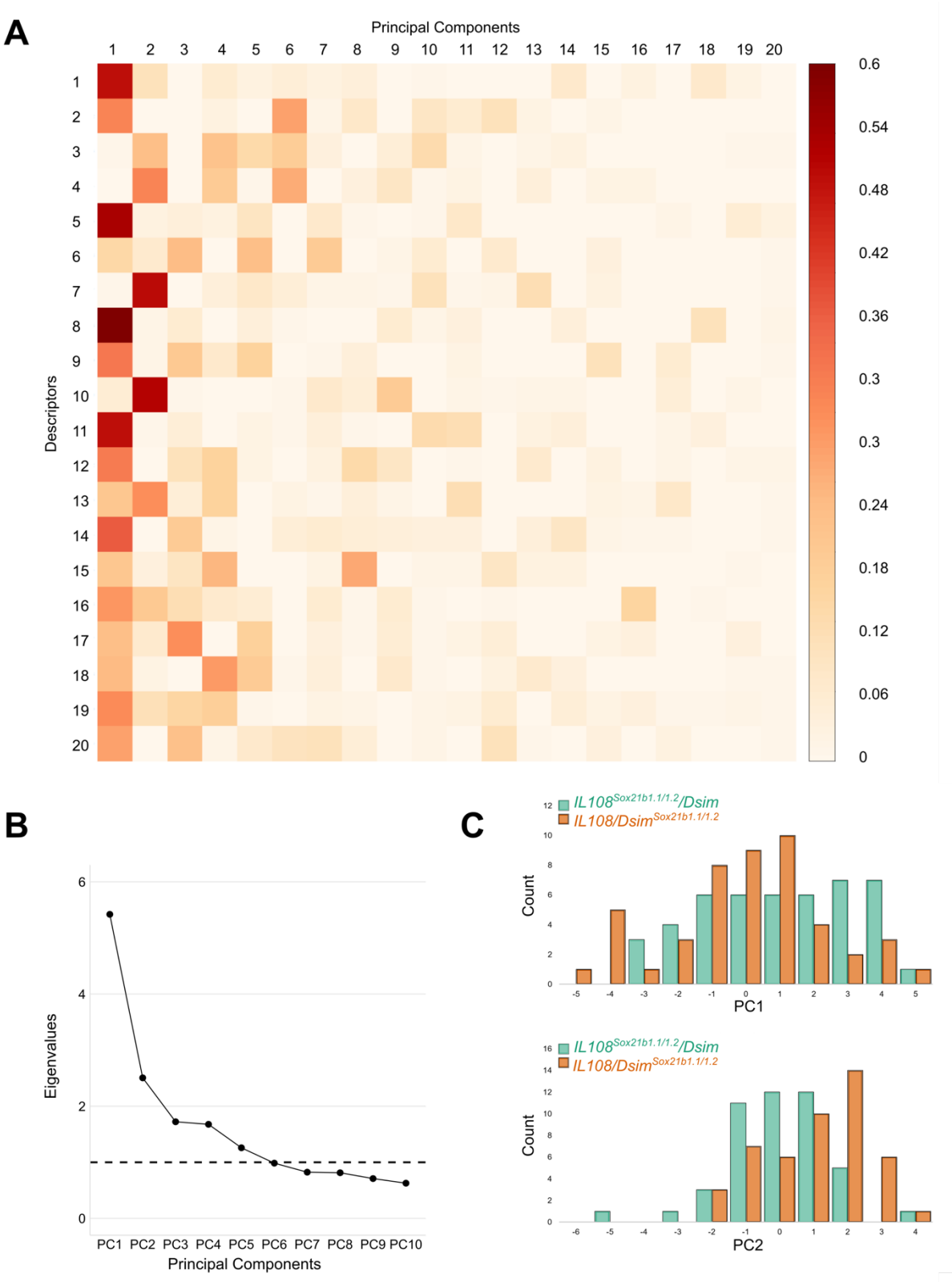
8.2 Chapter Three Supplementary Materials



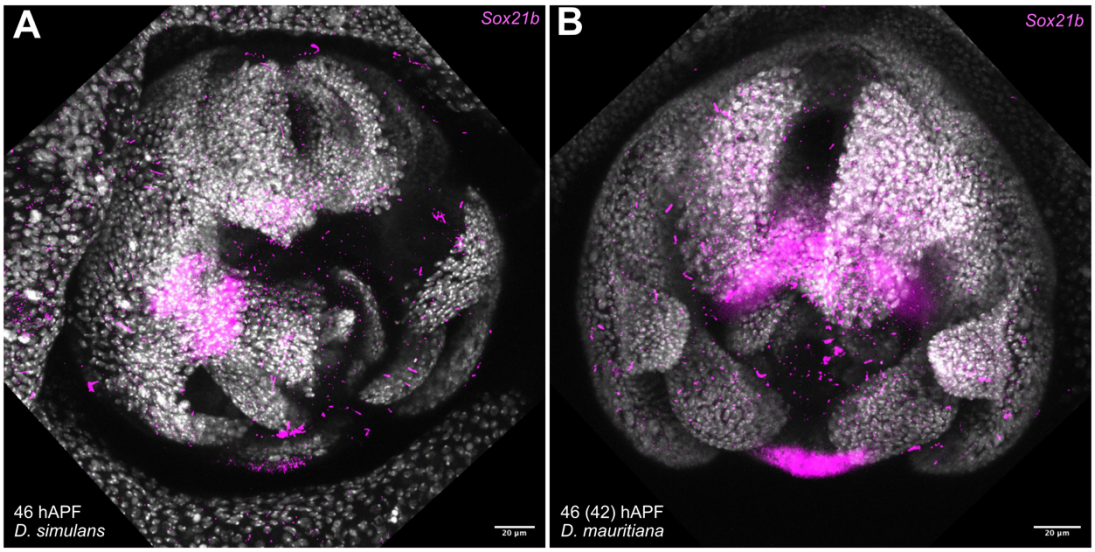
Supplementary Figure 1: An alternative GAL4 driver and increased temperature also generates an increase in cercus and posterior lobe area upon RNAi knockdown of *E(spl)m3-HLH*. (A,B) ** p<0.01, *** p<0.001. RNAi knockdowns are in bold. Pink-coloured asterisks mark statical significance compared to the *POXN-GAL4* parental control, and orange-coloured asterisks mark statistical significance compared to *NP6333-GAL4, UAS-DICER* parental control at 29 degrees (Supplementary Table 3). (A) n > 17, (B) n > 16.



Supplementary Figure 2: *Sox21b* HCR-IS in *D. melanogaster* female genital discs and developing male pupal terminalia. (A) *Sox21b* expression is localised to the female ovipositor region (dorsal expression) and the cercus primordium (ventral expression). (B,B') Male pupal terminalia shows *Sox21b* expression localised to the abdominal tissue surrounding the genital structures. Expression is strongest at the hypandrium. (A) n=3, (B) n=4, (B') n=3. Ce: Cercus and Ovi: Ovipositor primordia. All samples were stained with DAPI.


A
Principal Components
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Descriptors
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
0.6
0.54
0.48
0.42
0.36
0.3
0.3
0.24
0.18
0.12
0.06
0
B
Eigenvalues
0 2 4 6
PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8 PC9 PC10
Principal Components
C
IL108Sox21b1.1/1.2/Dsim
IL108/DsimSox21b1.1/1.2
Count
PC1
IL108Sox21b1.1/1.2/Dsim
IL108/DsimSox21b1.1/1.2
Count
PC2

Supplementary Figure 3: Shape variation between *Sox21b* reciprocal hemizygotes posterior lobes captured by principal components. (A) A correlation matrix illustrating the contribution of each descriptor used to summarise the lobe outlines against each principal component. Cos2 values were used to produce a heat map, where a higher cos2 value (red) reveals a good representation of the descriptor by the associated principal component. A lower cos2 value (white) reveals a poor representation of the descriptor by the associated principal component. (B) Screeplot representing the Eigenvalue for each principal component (PC). Only the first five PCs had an Eigenvalue >1 and were hence selected for analysis to explore any shape alterations arising from the *Sox21b* allelic origin. Each PC with an Eigenvalue > 1 described the following proportion of shape variation: PC1 27.1%, PC2 12.5%, PC3 8.6%, PC4 8.4% and PC5 6.3%. P values indicate the level of significance of PCs when compared between reciprocal hemizygotes using a MANOVA (F(1,277)=1.6426, p=0.1489, Pillai=0.29206): PC1 p = 0.07397, PC2 p = 0.05378, PC3 p = 0.3025, PC4 p = 0.7791, PC5 p = 0.9376. (C) Comparison of the shape changes caused by the presence of the species-specific working *Sox21b* allele on posterior lobe shape variation summarised by PC1 and PC2.



Supplementary Figure 4: *Sox21b* HCR-IS in *D. simulans* and *D. mauritiana* male pupal terminalia. (A,B) *Sox21b* shows negligible expression in the pupal male terminalia of both species. Staining seen in both samples is above the plane of the tissue and cannot be explained for presently. (A) n=2, (B) n=1. All samples were stained with DAPI.

```
Sox14       ------------------------------------------------------------      0
SoxN        -MLTM--ESDMKGSLLHATMPPHHTSAALHGHAASPYSAL-----------APLMNLGQ      45
Sox21a      ------------------------------------------------------------      0
Sox21b      MMERLSTHLSLTGLSLAHPLQTHLGSG----GGTHHLGQLAAAVHQQQQQQQQSDQGLGH      56
Dichaete    -MATLST---------HPNYGFHLGQA----QGLEDY---------APQSQLQLSPG---      34


Sox14       -----------------------------------MIAKPNQATTEPPLSLRPGTVPT      23
SoxN        SHLTHSQLSHHNHHHH-------------HMSAHIAASQSPNP---------------L      76
Sox21a      ------------------------------------------------------------      0
Sox21b      SHHLHGQQLSLNHQHHHPHQSPQHQQQHQHHPCSSNSSNGSPNGST-------GLGLHPH     109
Dichaete    ------MDMDI--------------KRVLHYSQSLAAMGGSPNGPA-------GQGVNGS      67


Sox14       VPATTPARPATITIQRRHPAPKADSTPHTLPPFSPSP---------SPASSPSPAPAQTP      74
SoxN        SSLQSSMANTLNGSQVGQQQQ-QQQQQQSS-PLHSSSELSPTQSSIGSHHMTSPVSHQQH     134
Sox21a      ------MTSISALLHRSHSH---------SN-GYTSSG--------SSNHSHSSSLSPQLP      37
Sox21b      SALH--LAHHHSQLHQHHPAG--QQQQHSQPPVSSSS--------TSHHSQSQALSHQTH     157
Dichaete    SGMGHHMSSHMTPHHMHQ-----------------------------------AVSAQQT      92
                         :  :                                        *


Sox14       GAQKTQS-QAAITHPA---------AVASPSAPVAA-ASTEDPQDPGPRSTHTHTHSQHFS     124
SoxN        TQQQQHGQQQHLGA-------------------------------------------------     148
Sox21a      GINLGLGMVGGLGMGMSVGVGSGSGNTTP-PPMPPADLAVPPAAPTPVAPKMHQHTHHHG      96
Sox21b      S--NG---SSQLGGGSASGAGSVAGGAANSHHSAPASSSV-------MAAAAAAHLHHNS     205
Dichaete    L--SP---NSSIGSAGSLGSQ-------------------------------------------     108
                    .
Sox14       PPPRESEMDGERSPSHSGHEMTLSMDGIDSSLVFGSARVPVNSSTPYSDATRTKKHSPGH     184
SoxN        ----------------------G-------SALSSLTGGSSN--NNNNSATANKNQQHADR     178
Sox21a      NSHHN------APTSHSN-----------------------------SNTGSHHNSHDH     120
Sox21b      ---QA------SPISNLHQNMG-------SLMNSGSSASDV--FFSLMIQNTTKRQNEEH     247
Dichaete    --------------SSLGSN-G-------SGLNSSS---GH--QSAGMHSLATSPGQEGH     141
                                                                .     :

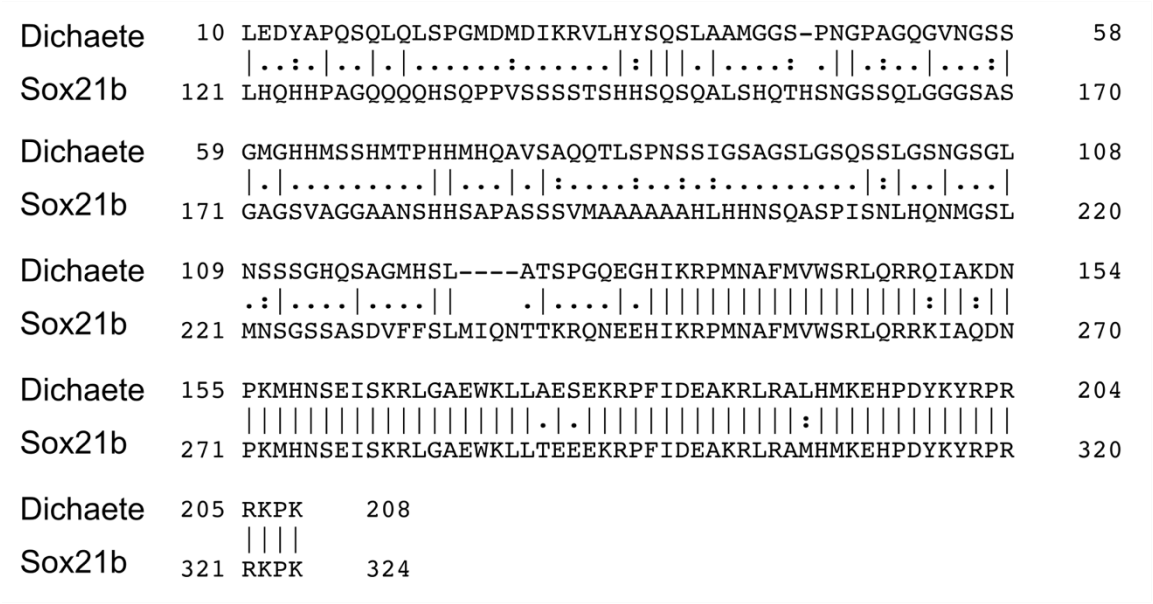

Sox14       IKRPMNAFMVWSQMERRKICERTPDLHNAEISKELGRRWQLLSKDDKQPYIIEAEKLRKL     244
SoxN        VKRPMNAFMVWSRGQRRKMASDNPKMHNSEISKRLGAQWKDLSESEKRPFIDEAKRLRAV     238
Sox21a      IKRPMNAFMVWSRGQRRKMAQDNPKMHNSEISKRLGAEWKLLTEGQKRPFIDEAKRLRAL     180
Sox21b      IKRPMNAFMVWSRLQRRKIAQDNPKMHNSEISKRLGAEWKLLTEEEKRPFIDEAKRLRAM     307
Dichaete    IKRPMNAFMVWSRLQRRQIAKDNPKMHNSEISKRLGAEWKLLAESEKRPFIDEAKRLRAL     201
            :***********: :**::.. .*.:**:****.** .*: *:: :*:*:* **::** :


Sox14       HMIEYPNYKYRPQKKQTRSPGSLKPN---------QDADGCEARNDTTNNNNSLTTLAING     296
SoxN        HMKEHPDYKYRPRRKTKTLTKTKEKYPMGGLMPGQTV-GGGAPGEP----------VTP-     286
Sox21a      HMKEHPDYKYRPRRKPKTLNKSPVPGGGGGG------GGGGANGGV----------NAGG     224
Sox21b      HMKEHPDYKYRPRRKPKALRRDGYPYPMPYPSVPVEALRAGITPGY----------FAPG     357
Dichaete    HMKEHPDYKYRPRRKPKNPLTAGPQGGLQMQAGGMGQQKLGAGPG---------------     246
            ** *:*:*****::* .


Sox14       TTTAGRKSKRSTSTCQSGSASKRLRNDSGDTS--------------SKPKYEVKMESAEQ     342
SoxN        -----TRVQGQPGQNQ-------SLNGSGGSAAAAAAAAAAAAQQARQDMYQMNA--PNG     332
Sox21a      AGNSGPS---GPGSVGSPKDMQPQLSPLGQSLPH-------LHGHPHQSPYQSHPHHPHP     274
Sox21b      PTAAAYHLGSHLGQTSTPTTTQATLSG--Q-MDK-------FA--LERSSYLSNSSQASA     405
Dichaete    AGAGGYN-------------------------P-------FH--Q--------------     257
```

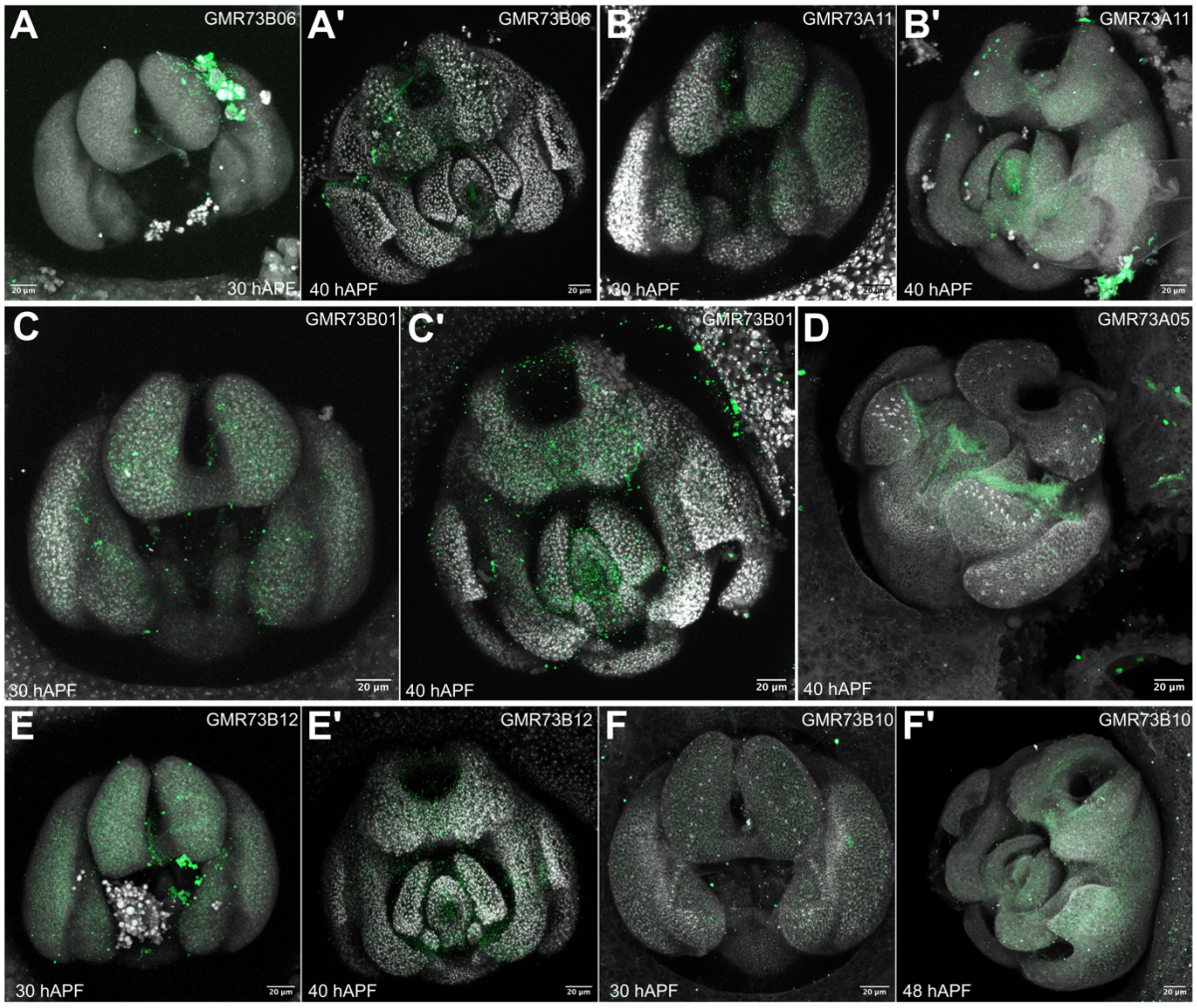


Supplementary Figure 5: Multiple sequence alignment of Group B Sox protein sequences from *D. melanogaster.* Sequence sourced from *D. melanogaster* R6.48. Alignment generated using CLUSTAL O (1.2.4). Outgroup is *Sox14*. * conserved amino acid. • amino acid differences where the charge is conserved. : amino acid difference where the charge is not conserved. Note, SoxN is considerably longer in length than the other Group B Sox proteins and the outgroup Sox14 protein.

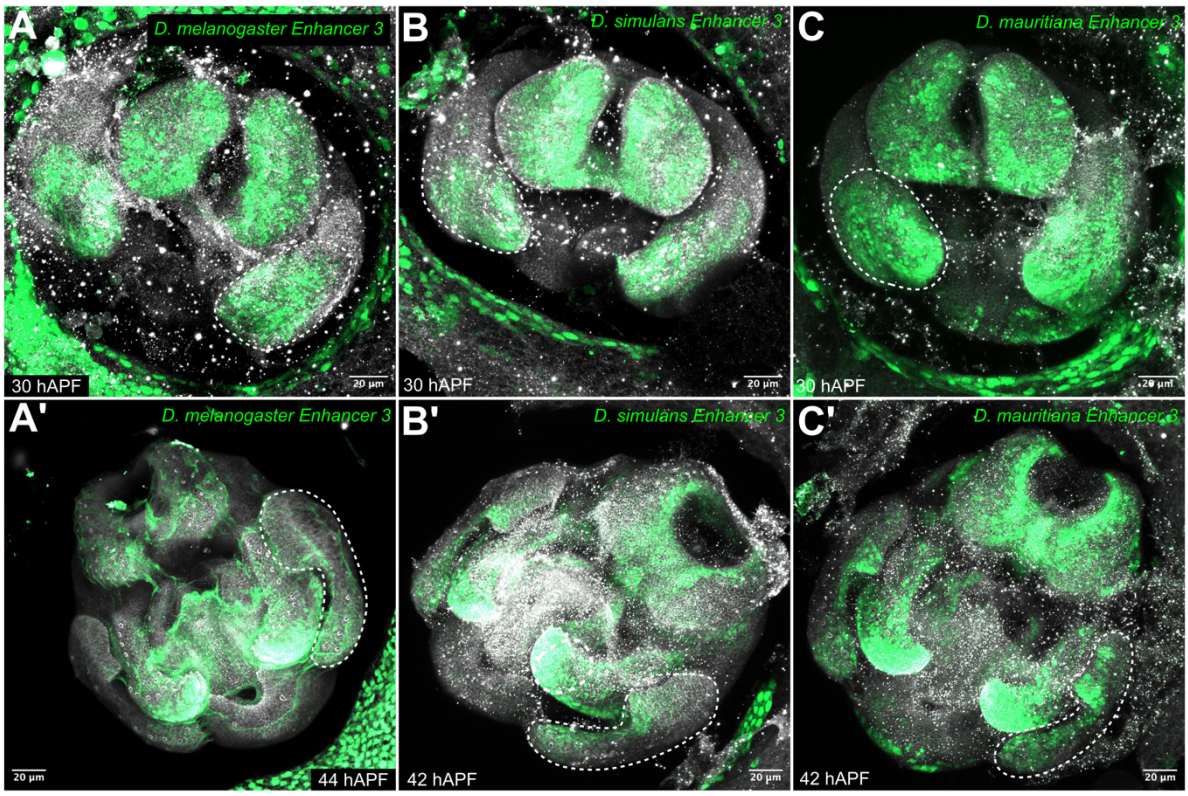


Supplementary Figure 6: Pairwise alignment of Dichaete and Sox21b *D. melanogaster* protein sequences. Both sequence were sourced from *D. melanogaster* R6.24. The Dichaete sequence shown in this figure is the region of the protein targeted by the Anti-Dichaete antibody. This sequence was compared to Sox21b to assess off-target effects. Alignment generated using EBLOSUM62 matrix.

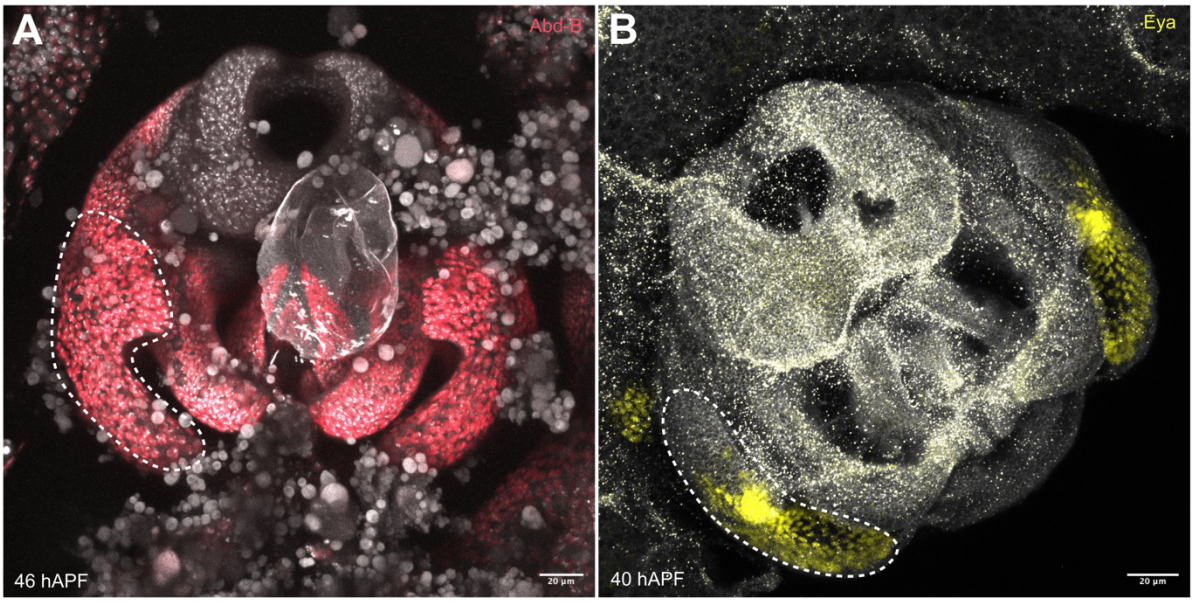
8.3 Chapter Four Supplementary Materials



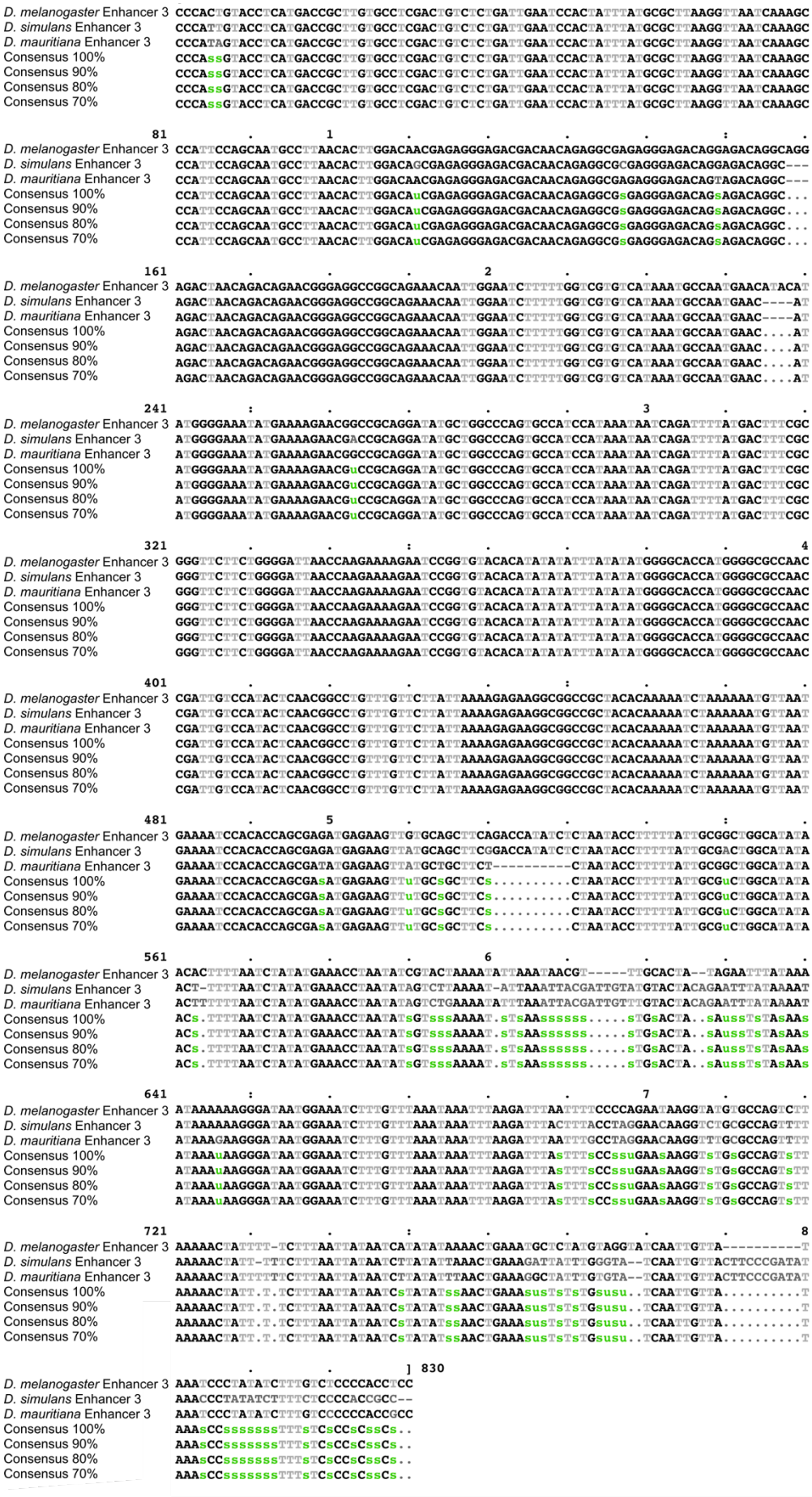
Supplementary Figure 7: *Sox21b* GAL4 enhancer driver lines not expressed during *D. melanogaster* male genitalia development. GAL4 enhancer lines were crossed to *UAS-GFP*, and resultant progeny were stained using anti-GFP to capture enhancer activity. (A) n=2, (A') n=2, (B) n=1, (B') n=2, (C) n=2, (C') n=3, (D) n=1, (E) n=2, (E') n=2, (F) n=1, (F') n=1. All samples were stained with DAPI apart from (D,F,F') where anti-DCAD2 staining is used to visualise cell membrane structures.



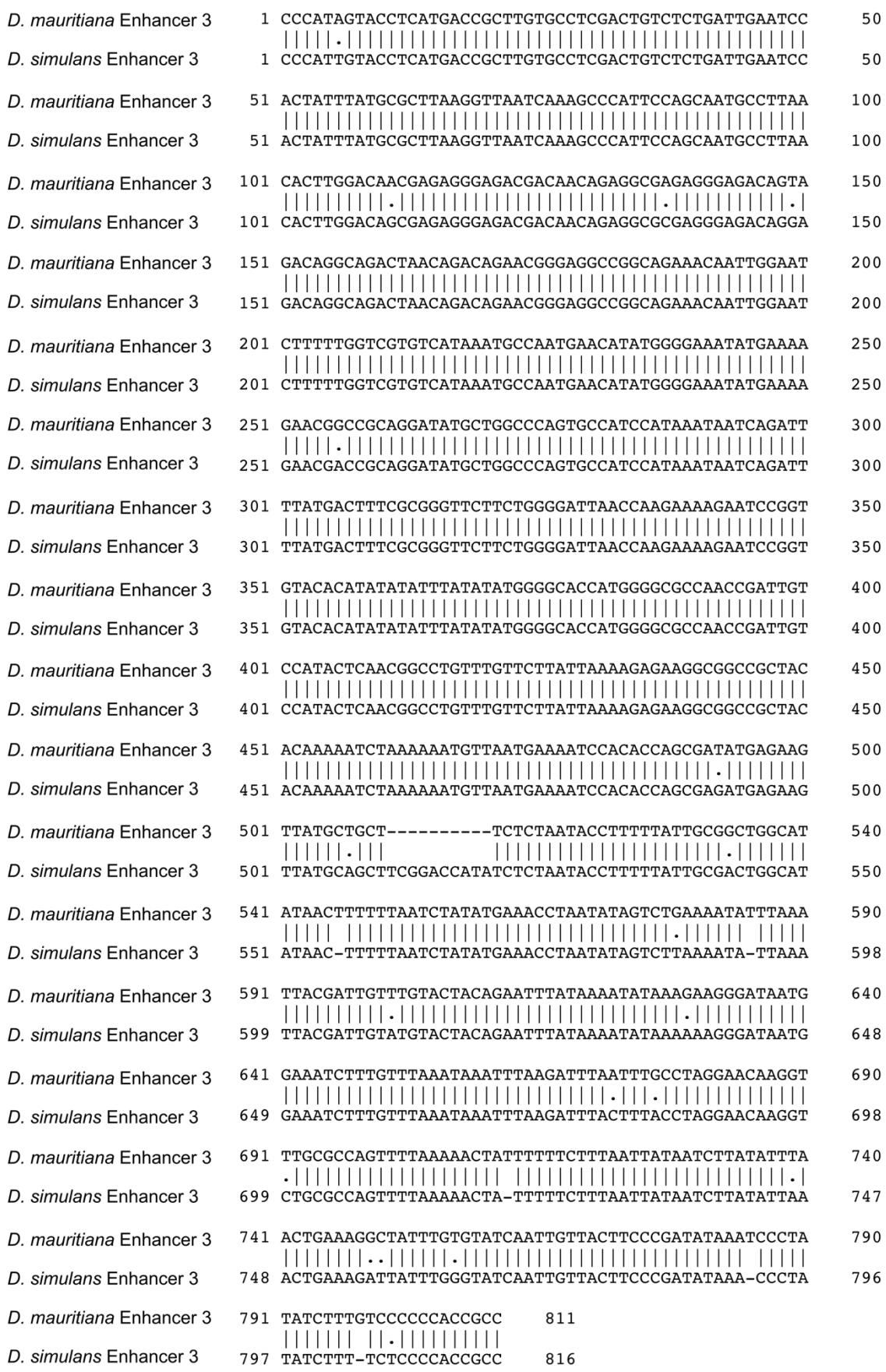
Supplementary Figure 8: Comparison of *D. melanogaster*, *D. simulans* and *D. mauritiana* Enhancer 3 GAL4 sequences driving *UAS-GFP* expression during male pupal genitalia development. All species-specific alleles for Enhancer 3 are screened for in *D. melanogaster*. All species-specific Enhancer 3 lines are active in the claspers, lateral plates, posterior lobes and cerci. Dashed white outlines show posterior lobe and lateral plate structures. All samples have been stained against DCAD2. (A) n=1, (A') n=1, (B) n=4, (B') n=2, (C) n=3, (C') n=4.



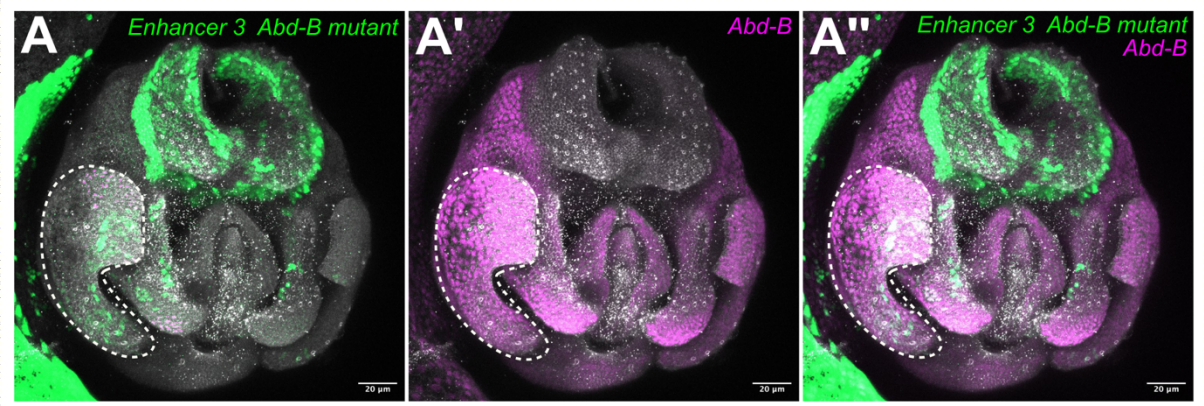
Supplementary Figure 9: Abd-B and Eya stains in *D. melanogaster* developing male genital structures. (A) Abd-B is present in all structures apart from the cerci. (B) Eya is expressed in the lateral plate and the hinge region between the posterior lobe and lateral plate. White dashed outlines marks the lateral plate and posterior lobe structures. (A) DAPI stained and (B) anti-DCAD2 staining. (A) n=3, (B) n=6.


D. melanogaster Enhancer 3 CCCACTGTACCTCATGACCGCTTGTGCCTCGACTGTCTCTGATTGAATCCACTATTTATGCGCTTAAGGTTAATCAAAGC
D. simulans Enhancer 3 CCCATTGTACCTCATGACCGCTTGTGCCTCGACTGTCTCTGATTGAATCCACTATTTATGCGCTTAAGGTTAATCAAAGC
D. mauritiana Enhancer 3 CCCATAGTACCTCATGACCGCTTGTGCCTCGACTGTCTCTGATTGAATCCACTATTTATGCGCTTAAGGTTAATCAAAGC
Consensus 100% CCCAssGTACCTCATGACCGCTTGTGCCTCGACTGTCTCTGATTGAATCCACTATTTATGCGCTTAAGGTTAATCAAAGC
Consensus 90% CCCAssGTACCTCATGACCGCTTGTGCCTCGACTGTCTCTGATTGAATCCACTATTTATGCGCTTAAGGTTAATCAAAGC
Consensus 80% CCCAssGTACCTCATGACCGCTTGTGCCTCGACTGTCTCTGATTGAATCCACTATTTATGCGCTTAAGGTTAATCAAAGC
Consensus 70% CCCAssGTACCTCATGACCGCTTGTGCCTCGACTGTCTCTGATTGAATCCACTATTTATGCGCTTAAGGTTAATCAAAGC
81 . 1 . . . : . 1
D. melanogaster Enhancer 3 CCATTCCAGCAATGCCTTAACACTTGGACAACGAGAGGGAGACGACAACAGAGGCGAGAGGGAGACAGGAGACAGGCAGG
D. simulans Enhancer 3 CCATTCCAGCAATGCCTTAACACTTGGACAGCGAGAGGGAGACGACAACAGAGGCGCGAGGGAGACAGGAGACAGGC---
D. mauritiana Enhancer 3 CCATTCCAGCAATGCCTTAACACTTGGACAACGAGAGGGAGACGACAACAGAGGCGAGAGGGAGACAGTAGACAGGC---
Consensus 100% CCATTCCAGCAATGCCTTAACACTTGGACAuCGAGAGGGAGACGACAACAGAGGCGsGAGGGAGACAGsAGACAGGC...
Consensus 90% CCATTCCAGCAATGCCTTAACACTTGGACAuCGAGAGGGAGACGACAACAGAGGCGsGAGGGAGACAGsAGACAGGC...
Consensus 80% CCATTCCAGCAATGCCTTAACACTTGGACAuCGAGAGGGAGACGACAACAGAGGCGsGAGGGAGACAGsAGACAGGC...
Consensus 70% CCATTCCAGCAATGCCTTAACACTTGGACAuCGAGAGGGAGACGACAACAGAGGCGsGAGGGAGACAGsAGACAGGC...
161 . . . 2 2
D. melanogaster Enhancer 3 AGACTAACAGACAGAACGGGAGGCCGGCAGAAACAATTGGAATCTTTTTGGTCGTGTCATAAATGCCAATGAACACATACAT
D. simulans Enhancer 3 AGACTAACAGACAGAACGGGAGGCCGGCAGAAACAATTGGAATCTTTTTGGTCGTGTCATAAATGCCAATGAAC----AT
D. mauritiana Enhancer 3 AGACTAACAGACAGAACGGGAGGCCGGCAGAAACAATTGGAATCTTTTTGGTCGTGTCATAAATGCCAATGAAC----AT
Consensus 100% AGACTAACAGACAGAACGGGAGGCCGGCAGAAACAATTGGAATCTTTTTGGTCGTGTCATAAATGCCAATGAAC....AT
Consensus 90% AGACTAACAGACAGAACGGGAGGCCGGCAGAAACAATTGGAATCTTTTTGGTCGTGTCATAAATGCCAATGAAC....AT
Consensus 80% AGACTAACAGACAGAACGGGAGGCCGGCAGAAACAATTGGAATCTTTTTGGTCGTGTCATAAATGCCAATGAAC....AT
Consensus 70% AGACTAACAGACAGAACGGGAGGCCGGCAGAAACAATTGGAATCTTTTTGGTCGTGTCATAAATGCCAATGAAC....AT
241 : . . . . 3 . . 3
D. melanogaster Enhancer 3 ATGGGGAAATATGAAAAGAACGGCCGCAGGATATGCTGGCCCAGTGCCATCCATAAATAATCAGATTTTATGACTTTCGC
D. simulans Enhancer 3 ATGGGGAAATATGAAAAGAACGACCGCAGGATATGCTGGCCCAGTGCCATCCATAAATAATCAGATTTTATGACTTTCGC
D. mauritiana Enhancer 3 ATGGGGAAATATGAAAAGAACGGCCGCAGGATATGCTGGCCCAGTGCCATCCATAAATAATCAGATTTTATGACTTTCGC
Consensus 100% ATGGGGAAATATGAAAAGAACGuCCGCAGGATATGCTGGCCCAGTGCCATCCATAAATAATCAGATTTTATGACTTTCGC
Consensus 90% ATGGGGAAATATGAAAAGAACGuCCGCAGGATATGCTGGCCCAGTGCCATCCATAAATAATCAGATTTTATGACTTTCGC
Consensus 80% ATGGGGAAATATGAAAAGAACGuCCGCAGGATATGCTGGCCCAGTGCCATCCATAAATAATCAGATTTTATGACTTTCGC
Consensus 70% ATGGGGAAATATGAAAAGAACGuCCGCAGGATATGCTGGCCCAGTGCCATCCATAAATAATCAGATTTTATGACTTTCGC
321 . . : 4 4
D. melanogaster Enhancer 3 GGGTTCTTCTGGGGATTAACCAAGAAAAGAATCCGGTGTACACATATATATTTATATATGGGGCACCATGGGGCGCCAAC
D. simulans Enhancer 3 GGGTTCTTCTGGGGATTAACCAAGAAAAGAATCCGGTGTACACATATATATTTATATATGGGGCACCATGGGGCGCCAAC
D. mauritiana Enhancer 3 GGGTTCTTCTGGGGATTAACCAAGAAAAGAATCCGGTGTACACATATATATTTATATATGGGGCACCATGGGGCGCCAAC
Consensus 100% GGGTTCTTCTGGGGATTAACCAAGAAAAGAATCCGGTGTACACATATATATTTATATATGGGGCACCATGGGGCGCCAAC
Consensus 90% GGGTTCTTCTGGGGATTAACCAAGAAAAGAATCCGGTGTACACATATATATTTATATATGGGGCACCATGGGGCGCCAAC
Consensus 80% GGGTTCTTCTGGGGATTAACCAAGAAAAGAATCCGGTGTACACATATATATTTATATATGGGGCACCATGGGGCGCCAAC
Consensus 70% GGGTTCTTCTGGGGATTAACCAAGAAAAGAATCCGGTGTACACATATATATTTATATATGGGGCACCATGGGGCGCCAAC
401 . . . . . : . . . 4
D. melanogaster Enhancer 3 CGATTGTCCATACTCAACGGCCTGTTTGTTCTTATTAAAAGAGAAGGCGGCCGCTACACAAAAATCTAAAAAATGTTAAT
D. simulans Enhancer 3 CGATTGTCCATACTCAACGGCCTGTTTGTTCTTATTAAAAGAGAAGGCGGCCGCTACACAAAAATCTAAAAAATGTTAAT
D. mauritiana Enhancer 3 CGATTGTCCATACTCAACGGCCTGTTTGTTCTTATTAAAAGAGAAGGCGGCCGCTACACAAAAATCTAAAAAATGTTAAT
Consensus 100% CGATTGTCCATACTCAACGGCCTGTTTGTTCTTATTAAAAGAGAAGGCGGCCGCTACACAAAAATCTAAAAAATGTTAAT
Consensus 90% CGATTGTCCATACTCAACGGCCTGTTTGTTCTTATTAAAAGAGAAGGCGGCCGCTACACAAAAATCTAAAAAATGTTAAT
Consensus 80% CGATTGTCCATACTCAACGGCCTGTTTGTTCTTATTAAAAGAGAAGGCGGCCGCTACACAAAAATCTAAAAAATGTTAAT
Consensus 70% CGATTGTCCATACTCAACGGCCTGTTTGTTCTTATTAAAAGAGAAGGCGGCCGCTACACAAAAATCTAAAAAATGTTAAT
481 . 5 . . . . : . 5
D. melanogaster Enhancer 3 GAAAATCCACACCAGCGAGATGAGAAGTTGTGCAGCTTCAGACCATATCTCTAATACCTTTTTATTGCGGCTGGCATATA
D. simulans Enhancer 3 GAAAATCCACACCAGCGAGATGAGAAGTTATGCAGCTTCGGACCATATCTCTAATACCTTTTTATTGCGACTGGCATATA
D. mauritiana Enhancer 3 GAAAATCCACACCAGCGATATGAGAAGTTATGCTGCTTCT----------CTAATACCTTTTTATTGCGGCTGGCATATA
Consensus 100% GAAAATCCACACCAGCGAsATGAGAAGTTuTGCsGCTTCs.....CTAATACCTTTTTATTGCGuCTGGCATATA
Consensus 90% GAAAATCCACACCAGCGAsATGAGAAGTTuTGCsGCTTCs.....CTAATACCTTTTTATTGCGuCTGGCATATA
Consensus 80% GAAAATCCACACCAGCGAsATGAGAAGTTuTGCsGCTTCs.....CTAATACCTTTTTATTGCGuCTGGCATATA
Consensus 70% GAAAATCCACACCAGCGAsATGAGAAGTTuTGCsGCTTCs.....CTAATACCTTTTTATTGCGuCTGGCATATA
561 . . . 6 6
D. melanogaster Enhancer 3 ACACTTTTAATCTATATGAAACCTAATATCGTACTAAAATATTAAATAACGT-----TTGCACTA--TAGAATTTATAAA
D. simulans Enhancer 3 ACT-TTTTAATCTATATGAAACCTAATATAGTCTTAAAAT-ATTAAATTACGATTGTATGTACTACAGAATTTATAAAAT
D. mauritiana Enhancer 3 ACTTTTTTAATCTATATGAAACCTAATATAGTCTGAAAATATTTAAATTACGATTGTTTGTACTACAGAATTTATAAAAT
Consensus 100% ACs.TTTTAATCTATATGAAACCTAATATsGTsssAAAAT.sTsAAssssss.....sTGsACTA..sAussTsTAsAAs
Consensus 90% ACs.TTTTAATCTATATGAAACCTAATATsGTsssAAAAT.sTsAAssssss.....sTGsACTA..sAussTsTAsAAs
Consensus 80% ACs.TTTTAATCTATATGAAACCTAATATsGTsssAAAAT.sTsAAssssss.....sTGsACTA..sAussTsTAsAAs
Consensus 70% ACs.TTTTAATCTATATGAAACCTAATATsGTsssAAAAT.sTsAAssssss.....sTGsACTA..sAussTsTAsAAs
641 : . . . . 7 . . 7
D. melanogaster Enhancer 3 ATAAAAAAGGGATAATGGAAATCTTTGTTTAAATAAATTTAAGATTTAATTTTCCCCAGAATAAGGTATGTGCCAGTCTT
D. simulans Enhancer 3 ATAAAAAAGGGATAATGGAAATCTTTGTTTAAATAAATTTAAGATTTACTTTACCTAGGAACAAGGTCTGCGCCAGTTTT
D. mauritiana Enhancer 3 ATAAAGAAGGGATAATGGAAATCTTTGTTTAAATAAATTTAAGATTTAATTTGCCTAGGAACAAGGTTTGCGCCAGTTTT
Consensus 100% ATAAAuAAGGGATAATGGAAATCTTTGTTTAAATAAATTTAAGATTTAsTTTsCCssuGAAsAAGGTsTGsGCCAGTsTT
Consensus 90% ATAAAuAAGGGATAATGGAAATCTTTGTTTAAATAAATTTAAGATTTAsTTTsCCssuGAAsAAGGTsTGsGCCAGTsTT
Consensus 80% ATAAAuAAGGGATAATGGAAATCTTTGTTTAAATAAATTTAAGATTTAsTTTsCCssuGAAsAAGGTsTGsGCCAGTsTT
Consensus 70% ATAAAuAAGGGATAATGGAAATCTTTGTTTAAATAAATTTAAGATTTAsTTTsCCssuGAAsAAGGTsTGsGCCAGTsTT
721 . . : 8 8
D. melanogaster Enhancer 3 AAAAACTATTTT-TCTTTAATTATAATCATATATAAAACTGAAATGCTCTATGTAGGTATCAATTGTTA----------T
D. simulans Enhancer 3 AAAAACTATT-TTTCTTTAATTATAATCTTATATTAAACTGAAAGATTATTTGGGTA--TCAATTGTTACTTCCCGATAT
D. mauritiana Enhancer 3 AAAAACTATTTTTTCTTTAATTATAATCTTATATTTAACTGAAAGGCTATTTGTGTA--TCAATTGTTACTTCCCGATAT
Consensus 100% AAAAACTATT.T.TCTTTAATTATAATCsTATATssAACTGAAAsusTsTsTGsusu..TCAATTGTTA.....T
Consensus 90% AAAAACTATT.T.TCTTTAATTATAATCsTATATssAACTGAAAsusTsTsTGsusu..TCAATTGTTA.....T
Consensus 80% AAAAACTATT.T.TCTTTAATTATAATCsTATATssAACTGAAAsusTsTsTGsusu..TCAATTGTTA.....T
Consensus 70% AAAAACTATT.T.TCTTTAATTATAATCsTATATssAACTGAAAsusTsTsTGsusu..TCAATTGTTA.....T
. .] 830
D. melanogaster Enhancer 3 AAATCCCTATATCTTTGTCTCCCCACCTCC
D. simulans Enhancer 3 AAACCCTATATCTTTTCTCCCCACCGCC--
D. mauritiana Enhancer 3 AAATCCCTATATCTTTGTCCCCCCACCGCC
Consensus 100% AAAsCCsssssssTTTsTCsCCsCssCs..
Consensus 90% AAAsCCsssssssTTTsTCsCCsCssCs..
Consensus 80% AAAsCCsssssssTTTsTCsCCsCssCs..
Consensus 70% AAAsCCsssssssTTTsTCsCCsCssCs..

Supplementary Figure 10: Multiple sequence alignment of *Sox21b Enhancer 3* between *D. melanogaster, D. simulans* and *D. mauritiana*. Consensus ‘s’ in green represents transversion SNP difference, i.e. a pyrimidine to purine mutation. Consensus ‘u’ in green represents transition SNP difference, i.e. purine to purine mutation. ‘-‘ represents missing nucleotides compared to other aligned sequences.


D. mauritiana Enhancer 3 1 CCCATAGTACCTCATGACCGCTTGTGCCTCGACTGTCTCTGATTGAATCC 50
D. simulans Enhancer 3 1 CCCATTGTACCTCATGACCGCTTGTGCCTCGACTGTCTCTGATTGAATCC 50
D. mauritiana Enhancer 3 51 ACTATTTATGCGCTTAAGGTTAATCAAAGCCCATTCCAGCAATGCCTTAA 100
D. simulans Enhancer 3 51 ACTATTTATGCGCTTAAGGTTAATCAAAGCCCATTCCAGCAATGCCTTAA 100
D. mauritiana Enhancer 3 101 CACTTGGACAACGAGAGGGAGACGACAACAGAGGCGAGAGGGAGACAGTA 150
D. simulans Enhancer 3 101 CACTTGGACAGCGAGAGGGAGACGACAACAGAGGCGCGAGGGAGACAGGA 150
D. mauritiana Enhancer 3 151 GACAGGCAGACTAACAGACAGAACGGGAGGCCGGCAGAAACAATTGGAAT 200
D. simulans Enhancer 3 151 GACAGGCAGACTAACAGACAGAACGGGAGGCCGGCAGAAACAATTGGAAT 200
D. mauritiana Enhancer 3 201 CTTTTTGGTCGTGTCATAAATGCCAATGAACATATGGGGAAATATGAAAA 250
D. simulans Enhancer 3 201 CTTTTTGGTCGTGTCATAAATGCCAATGAACATATGGGGAAATATGAAAA 250
D. mauritiana Enhancer 3 251 GAACGGCCGCAGGATATGCTGGCCCAGTGCCATCCATAAATAATCAGATT 300
D. simulans Enhancer 3 251 GAACGACCGCAGGATATGCTGGCCCAGTGCCATCCATAAATAATCAGATT 300
D. mauritiana Enhancer 3 301 TTATGACTTTCGCGGGTTCTTCTGGGGATTAACCAAGAAAAGAATCCGGT 350
D. simulans Enhancer 3 301 TTATGACTTTCGCGGGTTCTTCTGGGGATTAACCAAGAAAAGAATCCGGT 350
D. mauritiana Enhancer 3 351 GTACACATATATATTTATATATGGGGCACCATGGGGCGCCAACCGATTGT 400
D. simulans Enhancer 3 351 GTACACATATATATTTATATATGGGGCACCATGGGGCGCCAACCGATTGT 400
D. mauritiana Enhancer 3 401 CCATACTCAACGGCCTGTTTGTTCTTATTAAAAGAGAAGGCGGCCGCTAC 450
D. simulans Enhancer 3 401 CCATACTCAACGGCCTGTTTGTTCTTATTAAAAGAGAAGGCGGCCGCTAC 450
D. mauritiana Enhancer 3 451 ACAAAAATCTAAAAAATGTTAATGAAAATCCACACCAGCGATATGAGAAG 500
D. simulans Enhancer 3 451 ACAAAAATCTAAAAAATGTTAATGAAAATCCACACCAGCGAGATGAGAAG 500
D. mauritiana Enhancer 3 501 TTATGCTGCT----------TCTCTAATACCTTTTTATTGCGGCTGGCAT 540
D. simulans Enhancer 3 501 TTATGCAGCTTCGGACCATATCTCTAATACCTTTTTATTGCGACTGGCAT 550
D. mauritiana Enhancer 3 541 ATAACTTTTTTAATCTATATGAAACCTAATATAGTCTGAAAATATTTAAA 590
D. simulans Enhancer 3 551 ATAAC-TTTTTAATCTATATGAAACCTAATATAGTCTTAAAATA-TTAAA 598
D. mauritiana Enhancer 3 591 TTACGATTGTTTGTACTACAGAATTTATAAAATATAAAGAAGGGATAATG 640
D. simulans Enhancer 3 599 TTACGATTGTATGTACTACAGAATTTATAAAATATAAAAAAGGGATAATG 648
D. mauritiana Enhancer 3 641 GAAATCTTTGTTTAAATAAATTTAAGATTTAATTTGCCTAGGAACAAGGT 690
D. simulans Enhancer 3 649 GAAATCTTTGTTTAAATAAATTTAAGATTTACTTTACCTAGGAACAAGGT 698
D. mauritiana Enhancer 3 691 TTGCGCCAGTTTTAAAAACTATTTTTTCTTTAATTATAATCTTATATTTA 740
D. simulans Enhancer 3 699 CTGCGCCAGTTTTAAAAACTA-TTTTTCTTTAATTATAATCTTATATTAA 747
D. mauritiana Enhancer 3 741 ACTGAAAGGCTATTTGTGTATCAATTGTTACTTCCCGATATAAATCCCTA 790
D. simulans Enhancer 3 748 ACTGAAAGATTATTTGGGTATCAATTGTTACTTCCCGATATAAA-CCCTA 796
D. mauritiana Enhancer 3 791 TATCTTTGTCCCCCCACCGCC 811
D. simulans Enhancer 3 797 TATCTTT-TCTCCCCACCGCC 816

Supplementary Figure 11: Pairwise alignment of *D. mauritiana* and *D. simulans* Enhancer 3 sequences. Alignment generated using EDNAFULL. A line between nucleotides represents conserved sequences between *D. mauritiana* and *D. simulans*. A dot between nucleotides represents a mismatch in nucleotides between the two species.



Supplementary Figure 12: Abd-B staining in *D. melanogaster* Enhancer 3 Abd-B binding site mutant. Enhancer 3 Abd-B mutant GAL4 line is driving *UAS-GFP* and terminalia are stained against GFP to visualise enhancer activity. The mutated enhancer has prominent activity in the cerci, and comparatively less activity in the posterior lobes, lateral plates and claspers. White dashed outline marks the posterior lobe and lateral plate, (n=6).

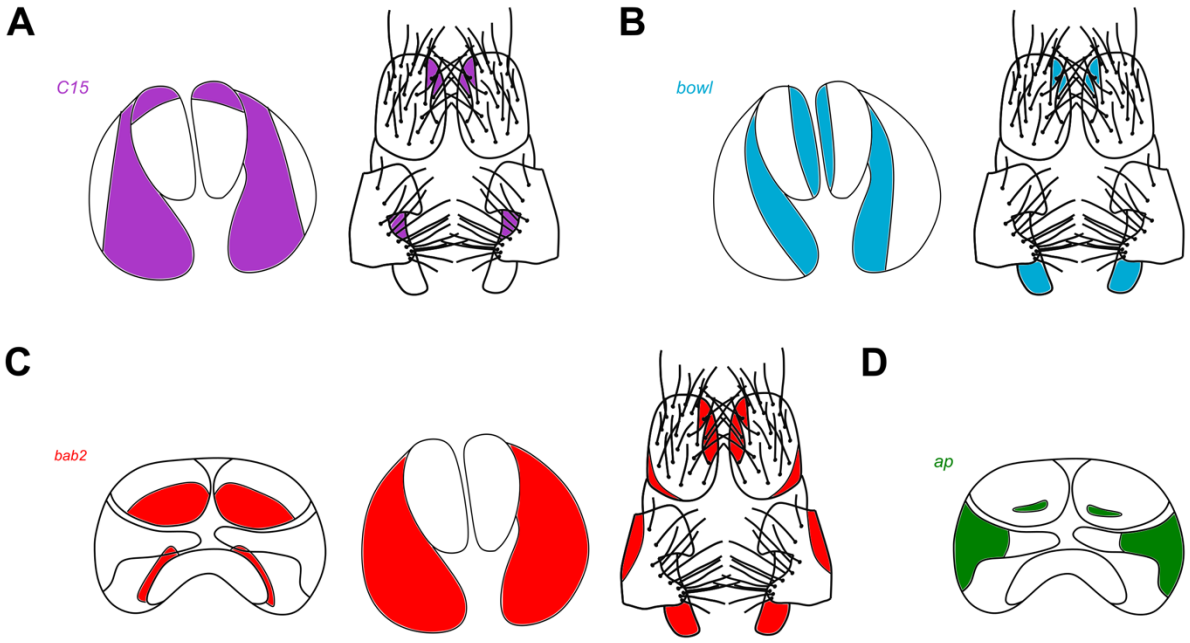
```
                                                
```

Supplementary Figure 13: Multiple sequence alignment of *Sox21b Enhancer 3* between *D. melanogaster*, *D. mauritiana* and *D. simulans*. The three Abd-B binding sites mapped in all three species are highlighted in green and labelled accordingly. '-' indicates indels between a species sequence to the other species sequences.

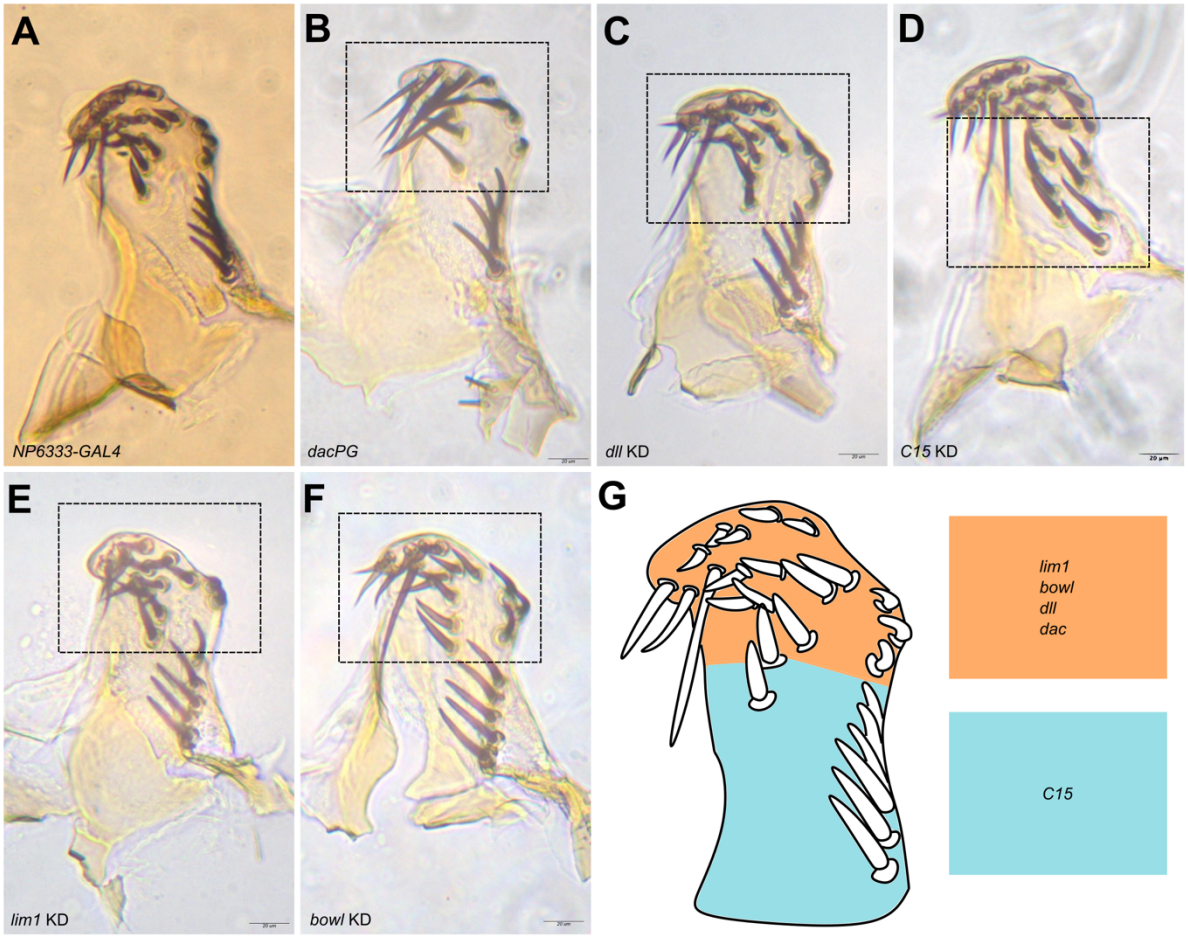


Supplementary Figure 14: Multiple sequence alignment of Enhancer 3 between lobed and non-lobed species. *D. melanogaster* (lobed), *D. yakuba* (non-lobed), *D. ananassae* (non-lobed), *D. biarmipes* (non-lobed). The three Abd-B binding sites are highlighted in green and labelled accordingly, and red boxes within these regions mark SNPs at the binding sites present between species.

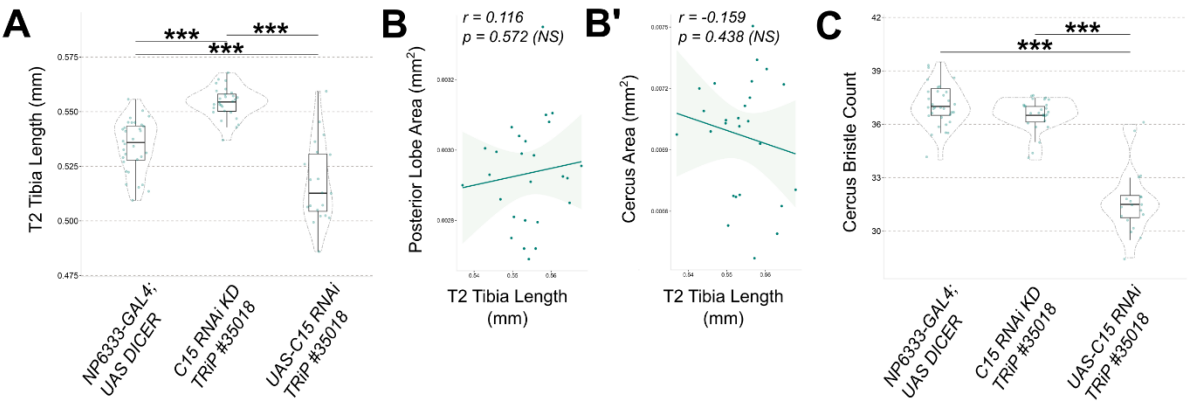
8.4 Chapter Five Supplementary Materials



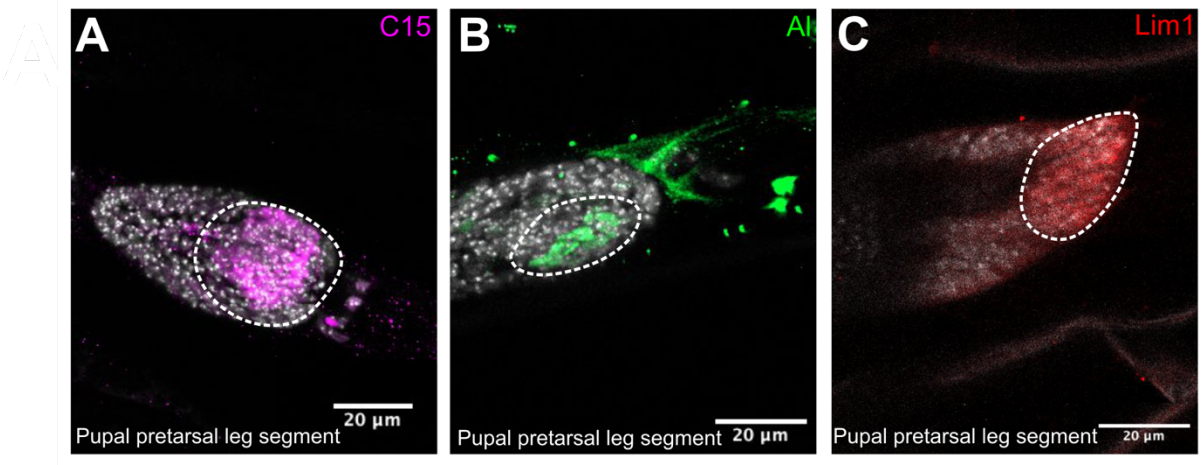
Supplementary Figure 15: Schematics of male genital expression patterns of known tarsal factors in larval discs and pupal tissues. (A) *ISH* of *C15* in male *D. melanogaster* shows clasper specific expression throughout pupal development, specifically in the proximal clasper domain during late development **(Vincent *et al.*, 2019)**. (B) *ISH* of *bowl* in male *D. melanogaster* is sustained at the throughout the clasper during pupal development and retracts to just the distal clasper domain in late development (Vincent et al., 2019). (C) Antibody staining against Bab2 in *D. melanogaster* male genital discs revealed expression surrounding the RFP and within the phallic primordium (Couderc et al., 2002). *ISH* of *bab2* during early male *D. melanogaster* pupal terminalia development showed expression in the claspers and lateral plates. During late development, expression can be seen in the distal clasper domain, posterior lobe base, and part of the cerci (Vincent et al., 2019). (D) Immunofluorescence of the *ap-lacZ* enhancer trap revealed expression in the clasper primordium of the genital disc, with some signal present in the phallic region **(Chatterjee *et al.*, 2011)**.



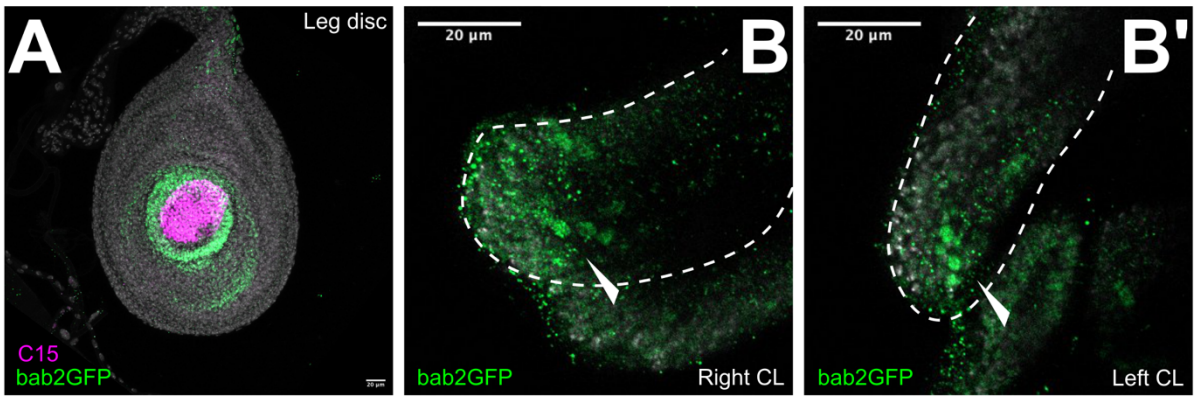
Supplementary Figure 16: TF expression in the proximal and distal clasper domains. *lim1, bowl, Dll* and *dac* have localised bristle disruption in the distal domain of the clasper upon RNAi knockdown, but only *C15* RNAi knockdown phenotypes accumulate in the proximal domain. (A) *NP6333-GAL4* is the parental control and driver used to generate RNAi knockdowns. (B-F) Dashed boxes highlight the regions of quantitative bristle differences generated from the RNAi knockdown in comparison to parental controls. (G) A schematic outlining the two TF expression domains within the clasper and those genes associated to each domain accordingly.



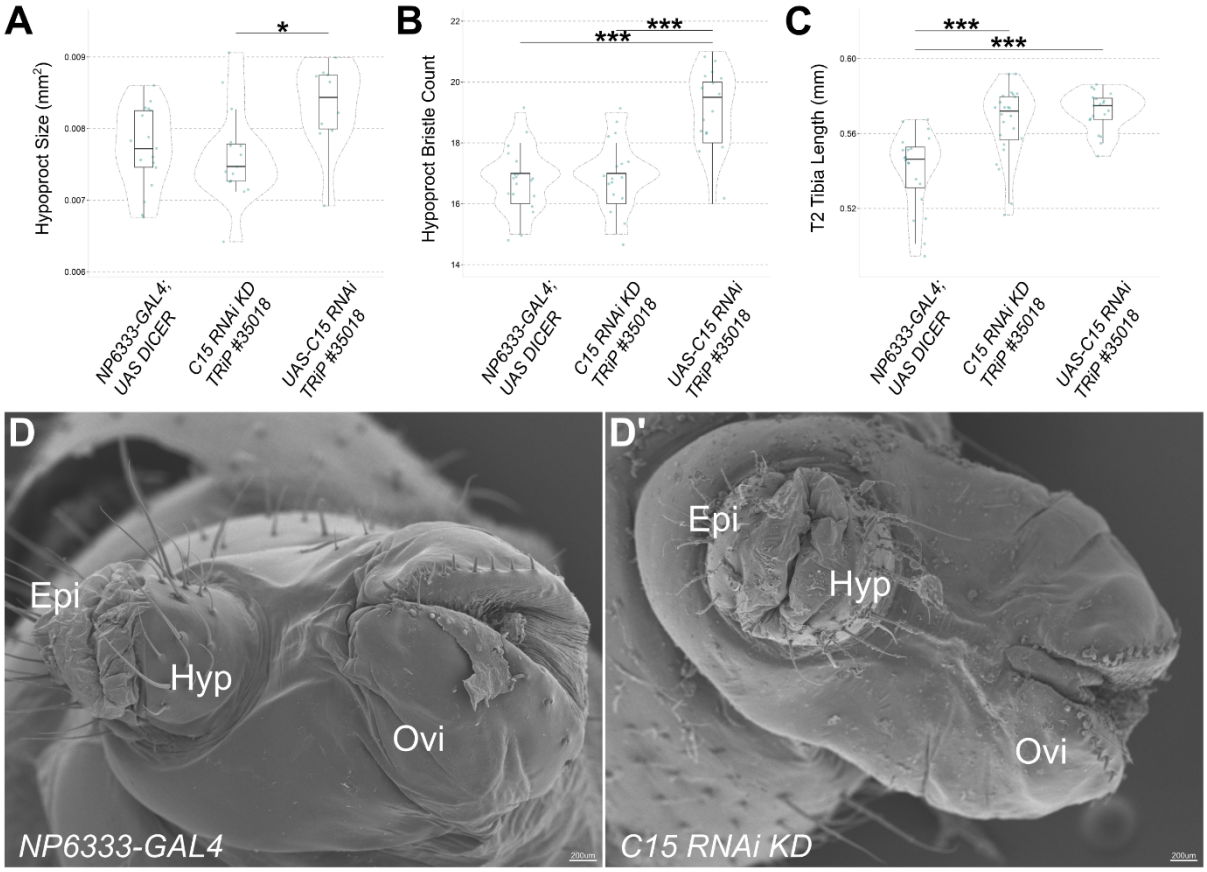
Supplementary Figure 17: *C15* RNAi knockdown effects on the cercus bristle count and T2 tibia length in *D. melanogaster* males. (A) *C15* RNAi knockdown significantly increased the T2 tibia length compared to both parental controls. (B,B') Pearson correlation of *C15* RNAi knockdown males for (B) posterior lobe area and (B') cercus area showed no significant relation with the T2 tibia length, concluding that the size phenotypes generated from RNAi knockdown are not due to overall body size effects. Shaded region represents 95% confidence interval of the mean using the Ordinary Least Squares linear regression. (C) Reduction of *C15* expression in the developing male terminalia led to no significant alterations in cercus bristle count compared to both parental controls. *** represents the statistical significance level of $p < 0.001$, $n>18$, (Supplementary Table 7).



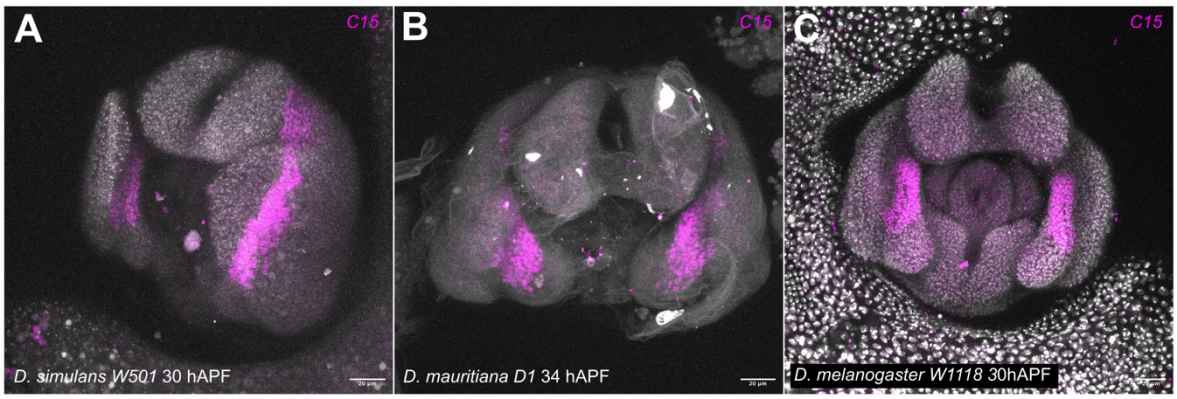
Supplementary Figure 18: Pupal pretarsal leg expression of C15, Al and Lim1. (A) C15, (B) Al and (C) Lim1 expression is at the pretarsal segment of late pupae legs, where the claw protrudes from. White dashed ovals marks regions of expression. All samples were stained with DAPI. (A) n=3, (B) n=3 and (C) n=3.



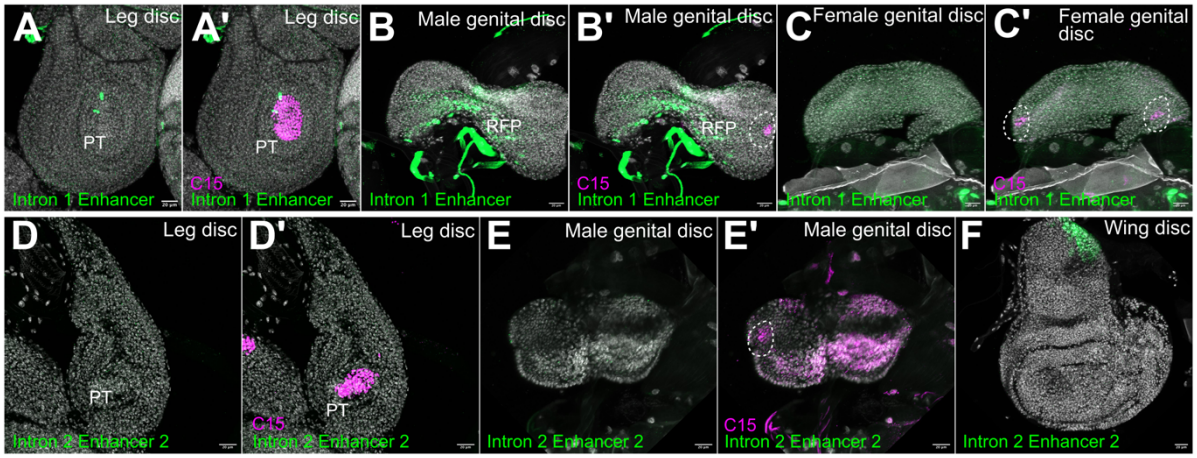
Supplementary Figure 19: Anti-GFP staining against bab2GFP in the larval leg disc and pupal male claspers. (A) Confirmation that the tagged *bab2* line is producing the endogenous expression known in the leg disc (de Celis Ibeas and Bray, 2003). Expression can be seen encircling C15 expression in the centre of the leg disc. (B,B') Individual z-stack slices of bab2GFP expression in the claspers at 30 hAPF. This was done to alleviate background noise generated during imaging of these stains. White arrows indicate expression in the distal region of the clasper, dotted white line outlines the clasper primordium. All samples were stained with DAPI. (A) n=2, (E,E') see Figure 43B.



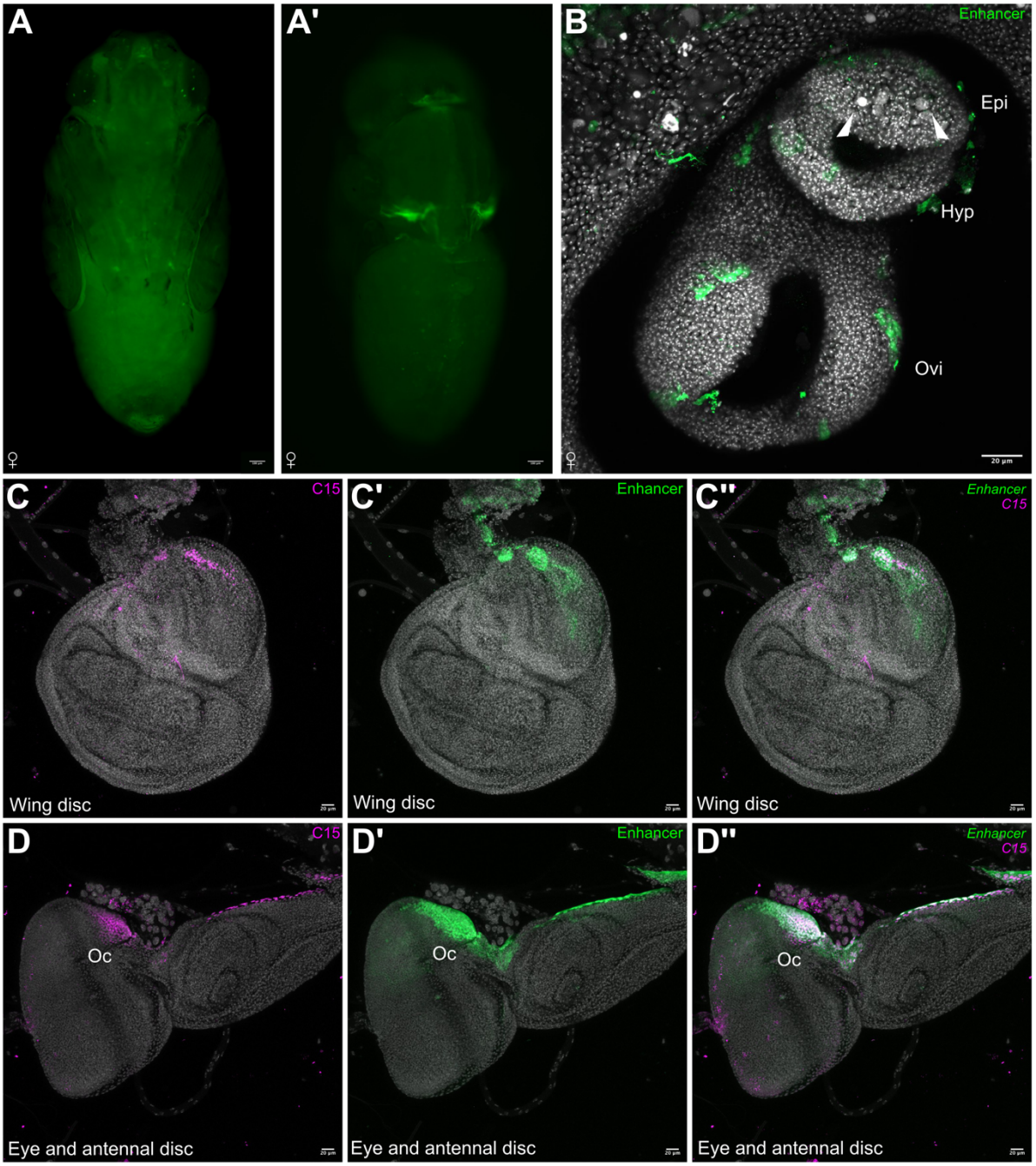
Supplementary Figure 20: *C15* RNAi knockdown has no significant effect on the hypoproct size and bristle count, or T2 tibia length in *D. melanogaster* females. (A) Reduction of *C15* expression in the developing female terminalia did not lead to a significant difference in the hypoproct size compared to both parental lines, (B) hypoproct bristle count and (C) T2 tibia length. *** represents the statistical significance level of $p < 0.001$, * $p<0.05$, n>9 (Supplementary Table 8). (D,D') SEM images of the entire female terminalia in (D) *NP6333-GAL4* parental control and (D') *C15* RNAi knockdown to mark the position of the analia and hypogynium relative to one another. Epi: Epiproct, Hyp: Hypoproct, Ovi: Ovipositor.



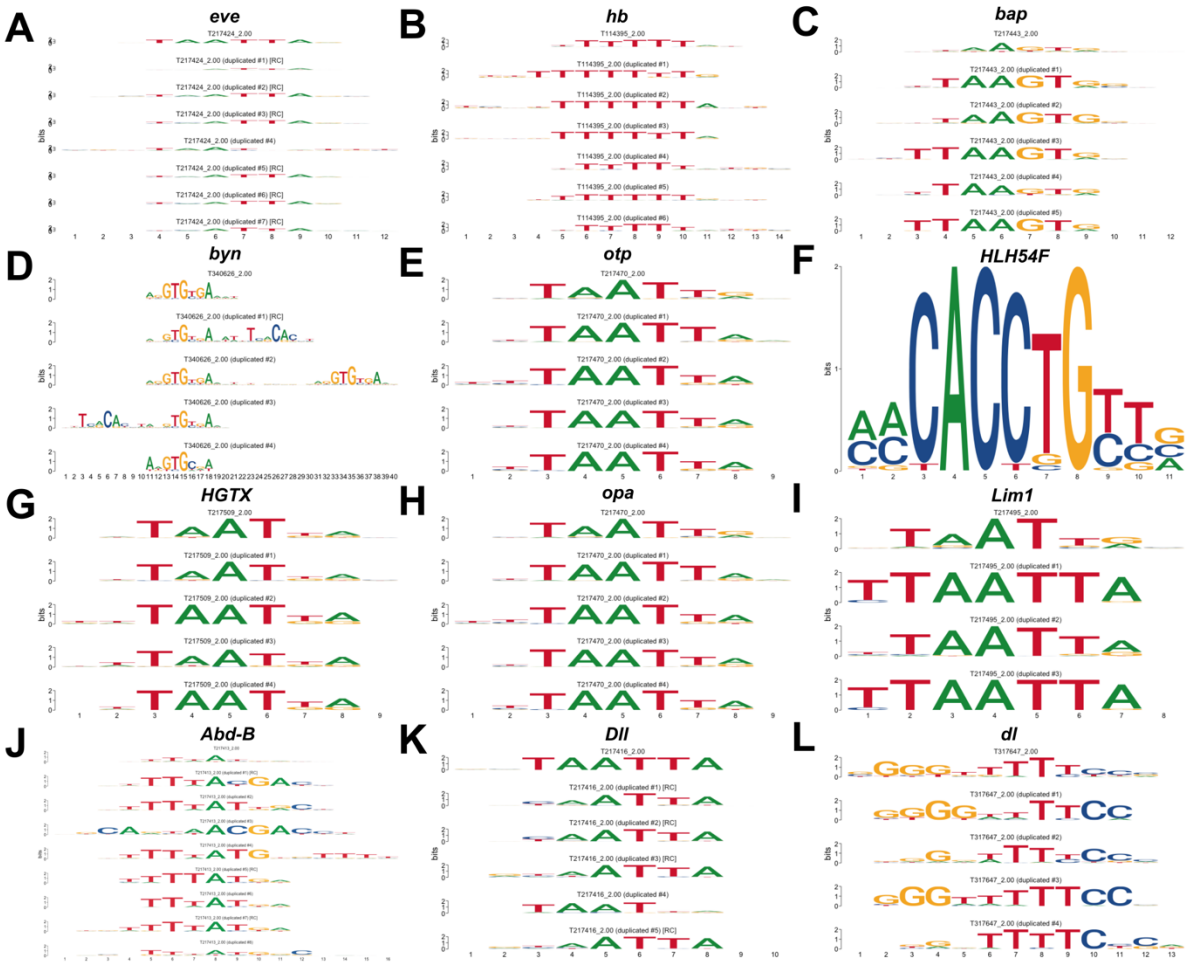
Supplementary Figure 21: C15 antibody staining in *D. simulans*, *D. mauritiana* and *D. melanogaster* developing male terminalia. *C15* is expressed in the proximal clasper domain in all three species. Differences in spatial expression reflect differences in clasper morphology between the three species. (A) n= 7 and (B) n = 3. All samples were stained with DAPI.



Supplementary Figure 22: *C15* enhancer screening in the imaginal discs. (A,A',D,D') Intron 1 enhancer and Intron 2 Enhancer 2 both show negligible expression in the leg disc, as well as no overlap with endogenous *C15* expression in either (B,B',E,E') male or (C,C') female genital discs. (F) The only enhancer activity to complement *C15* expression is Intron 2 enhancer 2 in the notum of the wing disc. (A,A') Intron 1 Enhancer showed activity in 3 cells that overlaps with C15 expression in the pretarsal central region of the leg disc. (B,B') Intron 1 Enhancer is only expressed at the RFP. (D,D,E,E') Leg discs and male genital discs show no expression of Intron 2 Enhancer 2. (A) n=2, (B) n=2, (C) n=1, (D) n=2, (E) n=1 and (F) n=1. All samples were stained with DAPI. PT: Pretarsus, RFP: repressed female primordium. Dashed white ovals mark C15 expression in male and female genital discs.



Supplementary Figure 23: *C15* Intron 2 Enhancer 1 screening in female pupa, wing discs and eye/antennal discs. Intron 2 enhancer 1 is active in the female hypogynium but absent from the epiproct. Activity is also seen to complement endogenous *C15* expression in the wing and eye/antennal discs. (A,A') Intron 2 Enhancer 1 GAL4 lines has been crossed to *UAS-GFP* and screened for in live female pupae. Activity can only be seen in the wing hinges and near the ocelli. (B) During female terminalia development, Intron 2 Enhancer 1 is active in the ovipositor but not where *C15* is localised (white arrows). (C-C'') Intron 2 Enhancer 1 overlapped with endogenous *C15* in the notum of the wing, (D-D'') and in the cuticle and ocelli primordium of eye/antennal discs in L3 larvae. (A) n=3, (B) n=4, (C) n=2, and (D) n=2. All samples were stained with DAPI. Epi: Epiproct, Hyp: Hypoproct, Ovi: Ovipositor, Oc: Ocellar.



Supplementary Figure 24: PWMs for TFs predicted to regulate *C15* expression during male genital development. Motifs were collected from data deposited on CIS-BP **(Weirauch *et al.*, 2014)**. This is relevant to *D. melanogaster C15* Intron 2 Enhancer 1 sequence.