UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

The genetic and genomic basis of crustacean body plan evolution

Permalink

https://escholarship.org/uc/item/4229t6t2

Author

Sun, Dennis August

Publication Date

2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

The genetic and genomic basis of crustacean body plan evolution

by

Dennis A. Sun

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

and the Designated Emphasis

in

Computational and Genomic Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:
Professor Nipam H. Patel, Chair
Professor Hernan G. Garcia, Co-Chair
Professor Craig T. Miller
Professor Robert Tjian
Professor Nir Yosef

Spring 2022

The genetic and genomic basis of crustacean body plan evolution

© 2022

by

Dennis A. Sun

Abstract

The genetic and genomic basis of crustacean body plan evolution

by Dennis A. Sun

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Nipam H. Patel, Chair
Professor Hernan G. Garcia, Co-Chair

Crustaceans are a group of organisms with tremendous diversity in the organization of limbs along their body, also known as the *bauplan*, or body plan. This diversity is hypothesized to arise from evolutionary changes to the expression domains of the *Hox* genes, an ancient gene family conserved to most animals. Previous researchers have used the amphipod crustacean *Parhyale hawaiensis* as a genetic model organism to examine the expression and function of *Hox* genes within crustaceans. This work revealed that each of the *Parhyale Hox* genes is expressed in a unique domain of expression that prefigures the ultimate arrangement of different limb types within the *Parhyale* body plan. Moreover, RNA interference, CRISPR-Cas9 mutagenesis, and heat shock overexpression experiments revealed that each of the *Hox* genes in *Parhyale* is required – or sufficient – for the development of different types of limbs. Taken in the context of changing *Hox* expression patterns, these results suggest that crustacean body plan evolution may be driven by changes to *Hox* gene regulation between different crustacean groups.

In this thesis, I present several avenues of investigation into the genetic and genomic basis of crustacean body plan evolution. First, I describe the generation of a developmental functional genomic database for the amphipod crustacean *Parhyale hawaiensis*. Using Omni-ATAC-Seq and short- and long-read RNA-Seq, I generate an improved genome annotation for *Parhyale*, discover hundreds of thousands of candidate *cis*-regulatory elements across the *Parhyale* genome, and infer additional features of the chromatin landscape of *Parhyale* development. Using this dataset, I demonstrate the potential to generate novel reporter genes. This dataset will serve as a platform for future researchers to investigate *cis*-regulation across the genome during *Parhyale* development, and will also enable the discovery of *Hox cis*-regulatory mechanisms.

Second, I describe experiments that reveal the role of two families of general transcriptional regulators, the Polycomb Group and Trithorax Group proteins, in regulation of the *Parhyale Hox* genes. Loss of function of Polycomb Group genes in particular results in homeotic transformations and changes to *Hox* expression. By examining the timing of these changes and synthesizing these data with other work,

I demonstrate that *Parhyale Hox* regulation occurs via at least three distinct mechanisms: early PcG-independent boundary establishment, PcG-dependent boundary maintenance, and *Hox* cross-regulatory interactions. I propose that crustacean *Hox* genes have evolved decoupled regulation at the level of developmental timing, tissue type, and specific PcG gene function. These results provide some of the first insights into the genetic mechanisms of *Parhyale Hox* regulation. Moreover, these data provide new mechanistic explanations for how crustaceans may have evolved their tremendous body plan diversity.

Finally, I describe efforts to establish animal husbandry, injection, and dissection techniques for the isopod crustacean *Asellus aquaticus* and the decapod crustacean *Neocaridina davidi*. I provide a developmental staging guide for each crustacean. I also describe the methods I have used to isolate early embryos for injection and the various types of troubleshooting I have performed in an attempt to establish these organisms as new genetic model systems. The work presented here will prove useful in future efforts to establish new crustacean genetic model systems.

This thesis is dedicated to my parents and grandparents,
who made great sacrifices to give me a life full of opportunity.

This work is especially dedicated to Shan Jiazhi, my *laoye*,
who passed away on Feb. 14, 2022.

Laoye was the most kind, resilient, and principled person I've ever met.
He spent his life in service to others as a physician.
His last wishes for me were to do good in the world and be kind to others.
I plan to honor his wishes for the rest of my life.

Table of Contents

Section	*Page*
Dedication	i
Table of Contents	ii
Figures and Tables	v
Acknowledgments	viii
Preface	xvii
Chapter 0. Introduction to Parhyale hawaiensis	
0 – The amphipod crustacean *Parhyale hawaiensis*: an emerging comparative model of arthropod development, evolution, and regeneration	1
0.1 – Introduction: Choosing a Model System	2
0.2 – Development, Anatomy and Life Cycle	5
0.3 – Tools for Genetic Manipulation	8
0.4 – Genomic and Bioinformatic Resources	9
0.5 – Case Studies	11
0.6 – References, Chapter 0	26
Chapter 1. Functional genomics of Parhyale hawaiensis	
1 – Identification and classification of cis-regulatory elements in the amphipod crustacean *Parhyale hawaiensis*	31
1.0 – Introduction	32
1.1 – Results	34
1.1.1 – Omni-ATAC-Seq identifies open chromatin across Parhyale developmental stages	34
1.1.2 – Improving the Parhyale genome annotation using short- and long-read RNA-Sequencing	36
1.1.3 – Inferring nucleosome positioning and transcription factor footprints from Omni-ATAC-Seq	40
1.1.4 – Identifying gene regulatory programs from Omni-ATAC-Seq using fuzzy clustering	41
1.1.5 – Concordant and discordant expression and accessibility dynamics appear across development	43
1.1.6 – Omni-ATAC-Seq re-identifies known Parhyale regulatory elements	45
1.1.7 – Minos transposase reporter assays can reveal novel promoters and distal enhancers	45
1.2 – Discussion	47
1.2.1 – The Parhyale genome contains many distant and dynamic regulatory elements	47
1.2.2 – Combined short- and long-read transcriptomes can improve genome annotations	48
1.2.3 – Deep analysis of ATAC-Seq data can reveal more than just dynamic accessibility	48
1.2.4 – Omni-ATAC-Seq identifies old and new regulatory elements, including enhancers	49
1.2.5 – Conclusion	50
1.3 – Methods, Chapter 1	64
1.4 – References, Chapter 1	69
1.5 – Supplementary Methods, Chapter 1	104
Chapter 2. Parhyale hawaiensis Polycomb and Trithorax	

2 – Distinct regulation of Hox genes by Polycomb Group genes in a crustacean 108
2.0 – Introduction 109
2.1 – Results 111
2.1.1 – The Parhyale genome contains all core PcG/TrxG components 111
2.1.2 – PcG KO induces homeotic transformations of anterior segments to T4/T5 walking leg identity 112
2.1.3 – PcG genes are essential for proper limb proximal-distal axis development and plate formation 113
2.1.4 – Parhyale Psc-2, but not Psc-1, is essential for proper embryonic development 114
2.1.5 – TrxG KO induces embryonic lethality, major developmental defects, or no phenotype; but not homeosis 114
2.1.6 – PcG gene knockout induces Hox misexpression during limb developmental stages 115
2.1.7 – PcG knockout-induced misexpression begins after wild-type Ubx boundary establishment 117
2.1.8 – Tissue- and PcG knockout gene-specific effects on Hox expression 118
2.2 – Discussion 119
2.2.1 – Proper Hox regulation in Parhyale requires at least three distinct mechanisms 119
2.2.2 – Each posterior Hox gene in Parhyale exhibits idiosyncratic regulation 122
2.2.3 – PcG genes are crucial for diverse developmental processes in Parhyale 123
2.2.4 – TrxG proteins show no obvious role in Parhyale Hox regulation 124
2.3 – Ideas and Speculation 124
2.3.1 – PcG knockout phenotypes suggest a limb-specific repression mechanism for Abd-B 124
2.3.2 – PcG knockouts reveal gene-specific roles in abd-A regulation and an unknown posterior boundary maintenance mechanism 125
2.3.3 – Implications for the cis-regulatory architecture of the Parhyale Hox cluster 125
2.3.4 – Implications for the ancestral roles of PcG in Hox regulation 126
2.4 – Conclusion 127
2.5 – Methods 142
2.6 – References, Chapter 2 144

Chapter 3. Embryonic development of Asellus aquaticus

3 – Developmental staging and embryological techniques for the isopod crustacean Asellus aquaticus 182
3.0 – Introduction 183
3.1 – Asellus animal husbandry and reproductive biology 184
3.2 – Strategies for the isolation of early stage Asellus embryos 185
3.3 – Stages of Asellus aquaticus development 186
3.4 – Injection trials for Asellus embryos 191
3.5 – Discussion 192
3.5.1 – The potential for dual growth zones during Asellus germband elongation 192
3.5.2 – The developmental and evolutionary origin of the Asellus lateral organ 193
3.5.3 – Multiple false “hatching” events during Asellus development and similarities to nauplii 194
3.5.4 – Next steps and future directions 195
3.5.5 – Conclusion 195

3.6 – Materials and Methods 196
3.7 – References, Chapter 3 198
Chapter 4. Transcriptomics of Asellus aquaticus
4 – Developmental transcriptomic analysis of the cave-dwelling crustacean, Asellus aquaticus 215
4.0 – Introduction 216
4.1 – Results 217
4.1.1 – Characterization of Surface, Cave, and Hybrid Transcriptomes 217
4.1.2 – Differential RNA-Seq Analysis Between Cave and Surface Morphs 218
4.1.3 – Allele-Specific Expression Analysis in Hybrid Individuals 218
4.2 – Discussion 220
4.2.1 – Candidate Genes 220
4.2.2 – Involvement of Regulatory Mutation Versus Coding Mutation in Evolution of Cave Traits 221
4.2.3 – Cis-Versus Trans-Regulation 221
4.2.4 – Parent-of-Origin Effects Versus Cis-Regulation 222
4.2.5 – Comparison to Adult Asellus Transcriptome 223
4.2.6 – Comparison to Other Cave-Dwelling Animal Transcriptomes 223
4.2.7 – Future Steps 224
4.3 – Materials and Methods 224
4.4 – References, Chapter 4 228
Chapter 5. Embryonic development of Neocaridina davidi
5 – Developmental staging and embryological techniques for the decapod crustacean Neocaridina davidi 237
5.0 – Introduction 238
5.1 – Neocaridina animal husbandry and reproductive biology 239
5.2 – Stages of Neocaridina davidi development 239
5.3 – Antibody staining reveals unexpected Ubx + abd-A expression 242
5.4 – Challenges of establishing microinjection in Neocaridina 244
5.5 – Discussion 248
5.6 – Materials and Methods 250
5.7 – References, Chapter 5 251

Figures and Tables

Section	*Page*
Chapter 0. Introduction to Parhyale hawaiensis	
0 – The amphipod crustacean *Parhyale hawaiensis*: an emerging comparative model of arthropod development, evolution, and regeneration	
Figure 0.1: Phylogeny and development of Parhyale hawaiensis.	17
Figure 0.2: Parhyale cell lineage and development.	18
Figure 0.3: Germ layer compensation and germ layer interactions in Parhyale.	20
Figure 0.4: Transgenesis in Parhyale using the Minos transposase system.	22
Figure 0.5: CRISPR-Cas9 mutagenesis and transgenesis in Parhyale.	23
Figure 0.6: Body plan organization and Hox gene expression of Parhyale and other representative arthropods.	24
Figure 0.7: Homeotic transformations of segment identity in Parhyale through Hox gene knockout.	25
Chapter 1. Functional genomics of Parhyale hawaiensis	
1 – Identification and classification of cis-regulatory elements in the amphipod crustacean *Parhyale hawaiensis*	
Figure 1.1: Time-course ATAC-Seq in Parhyale hawaiensis embryos.	51
Figure 1.2: Whole-genome analysis of ATAC-Seq reveals enrichment of promoter signals, dynamic accessibility over developmental time	52
Figure 1.3: Time-course RNA-Seq of Parhyale embryos	53
Figure 1.4: Inference of nucleosome positioning and transcription factor binding from Omni-ATAC-Seq data	55
Figure 1.5: Identification and classification of regulatory element clusters	56
Figure 1.6: Functional annotation of clusters using GO enrichment and transcription factor binding prediction	58
Figure 1.7: Correlation between RNA-Seq and Omni-ATAC-Seq	60
Figure 1.8: Minos transposase reporter assay identification of novel regulatory elements	62
Table 1.1: Plasmids tested using Minos transgenesis	63
Supplementary Figure 1.1: Embryo images for time course Omni-ATAC-Seq and RNA-Seq	75
Supplementary Figure 1.2.1: Omni-ATAC-Seq analysis pipeline and quality control metrics	76
Supplementary Figure 1.2.2: Examining outlier effects of S21/S22 samples	78
Supplementary Figure 1.2.3: Peak category analyses and data visualization tools	79
Supplementary Figure 1.3.1: Additional transcriptome and gene model evaluation metrics	81
Supplementary Figure 1.3.2: Additional transcriptome and gene model evaluation metrics	83
Supplementary Figure 1.3.3: Additional transcriptome and gene model evaluation metrics	84
Supplementary Figure 1.4: NucleoATAC quality control and signal over time	86
Supplementary Figure 1.5.1: Identification and classification of regulatory element clusters	87
Supplementary Figure 1.5.2: Mfuzz clustering assessment	88
Supplementary Figure 1.6: Alignment of motifs in fuzzy clusters	90
Supplementary Figure 1.7.1: Correlation between accessibility/gene expression at individual timepoints	92

Supplementary Figure 1.7.2: Visualization of accessibility versus expression for arbitrary peak-gene pairs 94
Supplementary Figure 1.8.1: Omni-ATAC-Seq recovers previously identified regulatory elements 96
Supplementary Figure 1.8.2: VISTA homology identification between Parhyale and Hyalella 97
Supplementary Video 1.1: Yolk cells labeled by the Hsp70-p2 reporter 98
Supplementary Figure 1.9: Spontaneous, non-reproducible expression patterns observed in Minos assays 99
Supplementary Figure 1.10: Expression observed from Sp69-p3 reporter 100
Supplementary Table 1.1: Illumina sequencing indices for Omni-ATAC libraries 101
Supplementary Table 1.2: Manual annotation coordinates of 49 selected RACE genes 102
Supplementary Table 1.3: Software version numbers 103

Chapter 2. Parhyale hawaiensis Polycomb and Trithorax

2 – Distinct regulation of Hox genes by Polycomb Group genes in a crustacean
Figure 2.1: The Parhyale hawaiensis genome contains all core PcG and TrxG genes. 128
Figure 2.2. PcG knockout hatchlings exhibit homeotic transformations 129
Figure 2.3: Proximal-distal and plate defects of PcG mutants 131
Figure 2.4. CRISPR mutagenesis of PcG induces Hox misexpression 133
Figure 2.5. Misexpression of Scr and Antp in PcG mutants 134
Figure 2.6: Differential timing of PcG knockout-mediated misexpression for Ubx and Abd-B 135
Figure 2.7: Differential tissue-specific misexpression of Hox genes between PcG mutants 137
Figure 2.8: Summary of findings and implications for Hox regulatory evolution 139
Table 2.1: Summary of PcG and TrxG CRISPR phenotypes 141
Supplementary Figure 2.1.1: Background 150
Supplementary Figure 2.1.2: Identification and expression of PcG/TrxG genes in Parhyale hawaiensis 151
Supplementary Figure 2.2.1: Psc family genes in Parhyale 152
Supplementary Figure 2.2.2: Sce CRISPR phenotype data 154
Supplementary Figure 2.2.3: Pcl CRISPR phenotype data 156
Supplementary Figure 2.2.4: Su(Z)12 CRISPR phenotype data 158
Supplementary Figure 2.2.5: Phenotype data for additional PcG genes 160
Supplementary Figure 2.2.6: Example phenotypes for TrxG CRISPR knockouts 162
Supplementary Figure 2.2.7: Lack of obvious phenotypes for Trx and Kis knockout experiments 164
Supplementary Figure 2.3.1: Asymmetric fusion of plates in Pc mutant hatchlings 166
Supplementary Figure 2.4.1: Expression of posterior Hox genes in PcG mutant embryos 167
Supplementary Figure 2.5.1: Misexpression of Scr, Ubx, and Abd-B in Pcl mutant embryos 168
Supplementary Figure 2.6.1: Differences in PcG dependence timing in Pcl mutant embryos 169
Supplementary Figure 2.6.2: Expression of posterior Hox genes in S21 embryos 171
Supplementary Figure 2.7.1: abd-A, Abd-B, and elav expression 173
Supplementary Figure 2.8.1: Hypothesized Hox regulatory mechanisms 175
Supplementary Table 2.1: Select Drosophila PcG/TrxG homeotic phenotypes 177
Supplementary Table 2.2: Select Mus PcG/TrxG homeotic phenotypes 178
Supplementary Table 2.3: Parhyale PcG genes 179
Supplementary Table 2.4: Parhyale TrxG genes 180

Supplementary Table 2.5: CRISPR guide sequences 181

Chapter 3. Embryonic development of Asellus aquaticus

3 – Developmental staging and embryological techniques for the isopod crustacean Asellus aquaticus

Figure 3.1: Animal husbandry approach for Asellus aquaticus 201

Figure 3.2: Earliest stages of Asellus embryogenesis 203

Figure 3.3: “Soccerball” and gastrulation stages 204

Figure 3.4: Early germband and lateral organ bud stages 205

Figure 3.5: Mid germband to late germband and limb bud elongation stages. 206

Figure 3.6: Late limb elongation stages to “angel” stage 208

Figure 3.7: Schematic timeline of Asellus development 210

Figure 3.8: Comparison of developmental timelines to previous studies. 211

Figure 3.9: Trial injections of DsRed mRNA in Asellus embryos. 213

Chapter 4. Transcriptomics of Asellus aquaticus

4 – Developmental transcriptomic analysis of the cave-dwelling crustacean, Asellus aquaticus

Figure 4.1: Top 50 overexpressed genes and top 50 underexpressed genes in the cave samples as compared to the surface samples. 233

Figure 4.2: Experimental design of allele-specific expression analysis. 234

Figure 4.3: Allele-specific expression suggests cis-regulatory contribution to population difference. 235

Table 4.1: Comparison of transcriptome assemblies of Asellus cave morphs, surface morphs and hybrid individuals. 236

Table 4.2: Comparison of transcriptome assemblies of Asellus cave morphs, surface morphs and hybrid individuals. 236

Chapter 5. Embryonic development of Neocaridina davidi

5 – Developmental staging and embryological techniques for the decapod crustacean Neocaridina davidi

Figure 5.1: Adult morphology and animal husbandry strategy for Neocaridina davidi 254

Figure 5.2: Early cell divisions of Neocaridina davidi embryos 255

Figure 5.3: Live imaging reveals timing of early Neocaridina developmental stages 256

Figure 5.4: Soccerball to early germband stages (2-6dpf) 257

Figure 5.5: Head development and posterior segmentation stages (7-10dpf) 258

Figure 5.6: Anterior expansion stages (11-12dpf) 260

Figure 5.7: Late developmental stages (14dpf to hatching) 261

Figure 5.8: Unexpected Ubx + abd-A expression patterns in Neocaridina embryos 263

Figure 5.9: Z-slice analysis confirms unexpected register of Ubx + abd-A staining 265

Figure 5.10: Thoracic limb buds show Ubx + abd-A staining 266

Figure 5.11: Schematic summary of Neocaridina development 267

Figure 5.12: High intraembryonic pressure as a barrier to successful Neocaridina injections 269

Figure 5.13: Effects of various osmotic solutions on early embryonic pressure and survival 270

Figure 5.14: Methods to generate synthetic “membranes” for Neocaridina injection 271

Figure 5.15: Proposed future strategies for embryonic injection in Neocaridina 272

Acknowledgments

This work could not have been completed without the advice and hard work of the many supportive members of my scientific and social communities. My Ph.D. journey had some unique challenges, including the transition of my thesis lab to another university and geographic location and the onset of a global pandemic. I could not have navigated this challenging, but ultimately rewarding experience without their support.

Patel Lab

I'd like to pay thanks to Nipam Patel, my advisor. He gave me a great deal of freedom to explore, fail, grow, and learn. He has cultivated a lab environment that brings in kind and supportive people and provides a positive research environment for budding scientists of many career stages. He provided many special opportunities, including having the chance to teach Embryology and to do work at the MBL, giving me the space to explore my interests in quantitative and genomic biology, allowing me to mentor a variety of undergraduates and lab visitors, letting me stay at Berkeley and navigating a long-distance mentorship, and connecting me with numerous interesting scientists with diverse interests over the years. I've appreciated how Nipam has kept me in the fold despite the geographic distance and turned to me for advice in my areas of expertise. He's helped me use my scientific strengths for the benefit of others in the lab and guided me to grow in fields where I have room to learn. I'm grateful to have received his mentorship for the last seven years, during which I've been able to grow tremendously as a scientist.

During my time in Nipam's lab, I've had the pleasure of interacting with a variety of folks who gave me great mentorship and support in my research. I've also enjoyed their company and friendship. Heather Bruce and Erin Jarvis were the most senior graduate students when I joined, and I am very thankful for the time they spent teaching me all about working with *Parhyale*. Heather gave me a lot of guidance in my early days of learning bioinformatics, as well as excellent instruction on embryological techniques over the years. Heather also provided invaluable transcriptome data for the functional genomics project in Chapter 1. Erin gave me great advice on a variety of molecular biology techniques and introduced me to Benchling, which I've relied on during the entirety of my PhD. She also generously gave me stewardship over her *Hox* cross-regulation paper. I'm glad we were able to put that work out to the world before I graduated! Ryan Null provided constant moral support and encouragement, as well as helpful tips about cloning. His belief in my potential has been a guiding star in some of the darkest times in my career. Rachel Thayer and I shared a bay for many years, through my rotation and into the first year of the pandemic, and I'm grateful for her constant companionship, mentorship, advocacy, and advice. The camaraderie between baymates is very special, and I'm grateful for all of the numerous chats we had about science and life over the years. Aaron Pomerantz inspired me with his tremendous productivity and courage in trying new things. I greatly appreciated his help with Nanopore sequencing and career advice. The heart-to-hearts we had about the challenges of grad school helped keep me going. Kyle DeMarr and I bonded over various shared interests, and I enjoyed learning about his very ambitious experiments. It's been very cool to see all of the different organisms Kyle has brought into the lab and to feel the infectious enthusiasm and learn from the expertise he has

in entomology. Jenny McCarthy and I taught developmental biology together, which was a somewhat exhausting but also rewarding experience. I'm glad we've been able to help each other with a variety of protocols and analyses.

While I only knew most of the folks in Patel Lab East in the brief times I visited the MBL, I'm very grateful for their technical support and companionship during the intense days of Embryology and through the vagaries of long-distance collaboration. Sophia Kelly has been an incredibly reliable, organized, resilient, and enthusiastic labmate. I've relied on her for help shipping reagents and troubleshooting protocols. I had a lot of fun working with her whenever I've come to the MBL and am super excited to see where her career goes from here. Gabby Jerz has provided great experimental support and friendship when I've needed her help both at the MBL and Berkeley. It's been very fun to see her projects come together to find those elusive butterfly compartment boundaries. Carsten Wolff has taught me a great deal about crustacean embryology as one of the world's experts on this topic. It's been very interesting to work with someone whose work I had read for several years before we actually met! I greatly appreciated all of his very helpful and insightful edits on this dissertation. Though I only briefly overlapped with them, I'm also grateful to have met and worked with Ana McMenamin and Eric Chen.

I have to give a particularly special thanks to Sarah Mileikis, who has been the executive assistant for the lab since Nipam moved to the MBL. Sarah has been the bridge between me and Nipam, as well as the central hub for the entire lab. Without her support, navigating a long-distance mentorship could not have gone as smoothly as it has. I hope my endless onslaught of emails wasn't too much trouble! I'm also grateful to Briana Bertochi, who helped with reimbursements and other financial issues in the lab, and Scott Koerner and Brian Cortright, who helped me connect with the Unimatrix One servers of the lab.

I had the honor to mentor a number of undergraduates during my time in the Patel Lab. My mentorship experiences were the single most rewarding aspect of my Ph.D. years, and I'm grateful to all of the students I worked with for giving me the opportunity to work with them. It has been incredible to see each of them grow as a scientist and go on to accomplish great things. Honghao Song was an exchange student I only worked with for a few months, but she made a tremendous impact on me. She is one of the smartest people I've ever met, and I'm grateful that she constantly challenged me to explain the exact rationale behind all of our experiments. I'm glad we were able to find a way to work together well and to learn from each other. Jeremy Chen worked on the function of *Kruppel* in *Parhyale*. I'm grateful I had the opportunity to provide him experimental support in a very independent project and thesis. Priyanka Pradhan worked with Nipam on antibody staining in a variety of different crustaceans and established the protocols for anesthetizing and staining *Neocaridina*. Her enthusiasm and independence were inspiring. Taewoo Jean Kim worked with Erin Jarvis on some of her projects and switched to helping me with *Neocaridina* work. He put many hours into building the developmental staging guide for *Neocaridina*, and I'm grateful for his diligence and hard work. Gideon Harianja helped me with experiments to try to identify cis-regulatory elements from ATAC-Seq data. He optimized and ran probably thousands of PCR reactions and was one of the most diligent and organized people I had the pleasure of working with. Lauren Zane helped with experiments to understand the development of *Asellus*. Her hard work and excitement about marine

biology were infectious. Thien Kim Ho trained on *Parhyale* with both Gideon and Lauren and helped with experiments to evaluate CRISPR-Cas9 guide efficiency in *Parhyale*. I'm inspired by her willingness to go outside of her comfort zone and to try new things. Rebecca Chang and Yuri Takahashi worked with me as a powerhouse Crustacean Team to understand Polycomb and Trithorax genes in *Parhyale*. It's been an honor to work with each of them over the last three years of my Ph.D. Rebecca inspired me with her kindness and diligence, and I'm eternally grateful for the many hours of hard work she put in – during a global pandemic – to work on our shared project. I think she is exactly the kind of person you would want to be a physician. Yuri challenged herself to learn many new skills and to take on intellectually challenging classes for our project, working with Rebecca and I all the way from Japan. Her rigorous and insightful observations were instrumental to our work, and I am incredibly excited to see what she accomplishes during her Ph.D. at Cambridge University.

Garcia Lab

I'd like to thank Hernan Garcia, my "adoptive" advisor, who has taken on the administrative obligations of a thesis chair and mentor for the latter parts of my career. I'm incredibly grateful that he allowed me to embed with his lab socially over the course of my entire Ph.D. and gave me lots of great scientific and career advice over the years. It's been amazing to watch how he fosters collaboration and community between his lab members and to be exposed to a very different way of approaching biology.

The Garcia Lab has been a second home for me during my Ph.D., ever since the "joint" rotation I did with Nipam and Hernan at the end of my first year. Jacques Bothma gave me so much great scientific and personal advice over the years. I've been incredibly grateful for his mentorship, both in the challenging scientific experiments I endeavored in my Ph.D. and in navigating the world as a queer person. Though I only briefly overlapped with her, Kasia Oktaba told me about ATAC-Seq, which ended up being one of the major techniques I used during my dissertation work. I greatly appreciated her scientific advice and direction at the early stages of my project. Matty Norstad and Myron Barber Child VI provided both experimental advice on challenges with cloning and neighborly friendship during their time in the lab. Matty helped me feel connected to everyone despite the substantial difference in field. Armando Reimer and Simon Alamos provided helpful scientific advice and mentorship. I appreciated their friendship and the various chats we had in the break room over the years and for tolerating my constant complaints and questions about cloning. Elizabeth Eck and Yang Joon Kim were in the same cohort (albeit in Biophysics) and we took MCB200 together as first-years. I appreciated Liz's level-headed approach to challenges and Yang Joon's easy-going attitude. Both of them provided helpful experimental advice and career mentorship. Gabriella Martini and I chatted a great deal about politics, departmental drama, and various other topics whenever we'd see each other. I'm grateful for our friendship and for her emotional support during experimental nadirs. Meghan Turner helped streamline some of the more chaotic situations in the labs (may we never see another fly mite again). I found her perspective on life refreshing and informative, and I certainly learned a lot about what life is like in Wisconsin! Nick Lammers provided a lot of helpful brainstorming on statistics and probability for the allele-specific expression analyses in Chapter 4. I'm grateful to have learned a tiny bit more about math from him. Jonathan Liu provided friendship, encouragement, and

career advice during my time in the lab. I greatly appreciated his help and advice on all manner of topics, and his belief and acknowledgment of my scientific growth helped keep me going in the most challenging moments. Brandon Schlomann and Julia Falo Sanjuan joined the lab late in my Ph.D. but provided tremendously helpful mentorship and friendship. Brandon gave me a great deal of his time in taking a look at my manuscripts and offering career advice. I'm grateful we had the opportunity to learn from each other. Julia and I shared many interesting conversations about our interests, both scientific and otherwise, and I appreciated her friendly demeanor and open-minded perspective to science. I'm also grateful to the other folks I overlapped with in the Garcia Lab: Jiaxi "Jake" Zhao, Niranjan Srinivas, Augusto Berrocal, Bruno Moretti, and the many undergraduate researchers and visiting scientists who I had conversations with over the years.

Hariharan Lab

I'm grateful to Iswar Hariharan, who served as a tertiary advisor during the latter stages of my Ph.D. Iswar provided me with advice and support through some rather difficult events during my time at Berkeley and supported my efforts to make change in the MCB department through my advocacy work. He also gave me a great deal of honest career and scientific advice, which gave me additional perspective and insight about academia. He also demonstrated by example a kind of steady and compassionate leadership that inspires me. Iswar encouraged me to attend Hariharan Lab events and functions and to consider myself a part of the lab, for which I am very grateful.

I ended up also embedding in the Hariharan Lab in part because I asked to have a bench in the lab when the rest of Patel Lab West was graduating and moving on. Unfortunately, the timing of this transition came right before the start of the COVID-19 pandemic, and so I did not end up spending much time at that bench. However, attending and presenting at Hariharan Lab meetings over the last several years has given me insight into diverse fields and topics and helped me make many new friends and colleagues, from whom I have learned a great deal. Maya Emmons-Bell was a great classmate and confidant, and very brief baymate. I enjoyed learning about her science and also chatting with her about all manner of topics: scientific, departmental, social, and otherwise. Melanie Worley gave me a great deal of helpful scientific and career advice. She is one of the sharpest scientists I know, and I learned to more carefully work through and explain my science from observing her and listening to her guidance. We also had a great time teaching at Embryology 2021! I was inspired by Jamie Lahvic's careful science on tumorigenesis and leadership in departmental advocacy. She gave me helpful advice for navigating departmental politics and making effective change. Nick Everetts, a fellow member of the Computational Biology community, gave me great advice on statistics, probability, and general quantitative biology skills. I appreciated chatting about our shared interests and learning about programming and data analysis from him. Danielle Spitzer and I taught together during Woods Hole Embryology in 2021, the first Embryology cohort after the start of the pandemic. I appreciated our many chats about science education, academic culture, and numerous other topics. Her cheerleading of my work was greatly appreciated and made me feel like my work was valuable and interesting to people outside of my immediate field. Sophia Friesen's unique perspective, artistic talent, and quirky humor helped make the Hariharan Lab an enjoyable social environment. Their enthusiasm for science communication helped improve my communication skills. Joyner Cruz was a great confidant and

supporter of my work. I was grateful for our chats about various topics, from quantitative training to issues of equity and inclusion in the department. I'm also grateful for other folks I overlapped with in the Hariharan Lab, including Riku Yasutomi and Andreana Gomez, for their friendship and camaraderie.

Weisblat Lab

After the onset of the COVID-19 pandemic, I ended up moving my lab bench down to a space in David Weisblat's lab. I am very grateful to David for allowing me to set up shop in the Weisblat lab for the last two years. David and I worked together as members of the first equity and inclusion section on the admissions committee for the MCB department. During my time in the Weisblat Lab, I appreciated his scientific advice, the warm and welcoming environment he created in the lab, and our many long conversations and collaborations on improving equity and inclusion in the department.

I am also very grateful to Christopher Winchell, a Patel Lab alumna working as a postdoc in the Weisblat Lab. Chris provided emotional support as well as scientific and career advice throughout our shared time in the Weisblat Lab. He was incredibly accommodating of the large amounts of reagents and equipment I ended up bringing into the lab space. I appreciated his patience, warm demeanor, scientific curiosity, and support of my science.

Miller Lab

My first rotation at Berkeley was with Craig Miller, who gave me my first scientific home in the department and brought me under his wing. Craig advised and mentored me in many different capacities over the years, including on a successful NSF fellowship application and serving on my qualifying exam and dissertation committees. I greatly appreciated his honest advice and insights into the inner workings of the department and its history. His scientific advice also helped me consider additional perspectives, directions, and outcomes of my work. I'm also grateful that he gave me the opportunity to serve on a faculty search during my time at Berkeley, which gave me a great deal of insight into how faculty hiring decisions in academia are made.

Various members of the Miller Lab gave me great scientific training and advice over the years. I'm grateful to James Hart, who mentored me during my rotation in the Miller Lab and taught me how to use the command line. Learning from his quantitative and computational skills was one of the major reasons I decided to do a Designated Emphasis in Computational Biology. Nicholas Ellis and Priscilla Erickson mentored me on stickleback biology and gave me a great deal of helpful career advice as I went about my rotations. Tyler Square and I had many great conversations over the years and I was grateful for his scientific perspective, camaraderie, and experimental advice. I was also grateful to share conversations and scientific ideas with Alyssa Rowley, whose enthusiasm for science inspired me.

Eisen Lab

I'm grateful to Michael Eisen, who gave me the opportunity to present in his lab meetings several times on my work and who gave me experimental, analytical, and career advice. Much of my understanding and inspiration for studying gene regulation came from evolutionary work that his lab

and the Garcia Lab have undertaken over the years. I've been inspired by Mike's fearless advocacy for improvements to the way we communicate science and for his efforts and support in improving the culture of the MCB department.

I received a great deal of experimental and intellectual advice from members of the Eisen Lab during my work. Jenna Haines taught me much of what I needed to know to perform ATAC-Seq, including sharing protocols and code that I adapted for my work. I appreciated her friendship, advice, and upbeat spirit. Colleen Hannon gave me great advice about analyzing ATAC-Seq data and thinking about how I could use my data to learn more about the chromatin state of the genome. I appreciate her pointing me to software such as HINT-ATAC and NucleoATAC, which ended up being major portions of Chapter 2. Michael Stadler also gave me intellectual, analytical, and experimental advice, for which I am very grateful. I also appreciated feedback, friendship, and support from other members of the Eisen Lab, including Victoria Blake and Laura Kovacs.

King Lab

I'm grateful to Nicole King, who gave me the opportunity to rotate in her lab during my first year, and who gave me scientific and career advice during my Ph.D. I appreciated the precision and rigor of the science in her lab, as well as her insights and perspectives on the MCB faculty search that we worked on together.

During my rotation in the King Lab, Laura Wetzel was my primary rotation mentor. I appreciated her willingness to entertain my numerous questions and requests and for teaching me about working with choanoflagellates. I also am grateful for others in the King Lab who mentored me and supported my science and advocacy efforts throughout the years, including Tess Linden, Thibaut Brunet, Flora Rutaganira, and Max Coyle.

Molecular and Cell Biology Community

I'm grateful to various other members of the MCB community, who supported me through the years. I'm thankful to Robert Tjian for his advice as the chair of my qualifying exam committee and member of my thesis committee. I'm grateful to have had the opportunity to teach with Craig Miller, Xavier Darzacq, and Roberto Zonco in MCB 104; and Richard Harland, John Gerhart, and Nipam Patel in MCB 141. I'm grateful to Jasper Rine, who gave me a platform to speak to the GGD faculty during my time as a divisional representative. I'm grateful to David Bilder and Elçin Ünal, who led the Graduate Affairs Committee and supported me in my advocacy efforts. Craig Miller, Elçin Ünal, Nicole King, Fyodor Urnov, Noah Whiteman, and I worked together on a faculty search committee that taught me a great deal about academia and for which I am very grateful. I'm grateful to Dan Rokhsar and Jessen Bredeson for support and advice on genomic analyses for Chapter 1. I'm particularly grateful to Jessen for helping me learn how to write scripts in bash when I was just starting out with quantitative analyses. I'm also grateful to Jase Gehring, who provided me with the enzymes and helped me troubleshoot protocols for performing ATAC-Seq.

Many of the experiments I performed could not have been accomplished without excellent technical support from professional scientists in the MCB community. I'm grateful to the members of the QB3 Functional Genomics lab, including Karen Lundy, Shana McDevitt, and Y. Justin Choi for

advice and experimental support on ATAC-Seq and RNA-Seq. I'm grateful to Brooks Bond-Watts of the QB3 Macrolab for cloning support and molecular biology reagents, including Cas9 enzyme.

I could not have made it through my Ph.D. without the amazing support of the members of the MCB Graduate Affairs Office and other departmental administrators. I'm grateful to Sarah Bottger and Sarah Hamilton, who served as the primary contact for graduate students in my first few years. I'm grateful to Berta Parra and Eric Buhlis, veterans of the GAO who provided steady and clear guidance and support. I'm grateful to Hannah Bloom, who has most recently taken the lead in the GAO. I'm grateful to Carina Galicia, whose great efforts and collaboration with students have improved graduate and undergraduate student life in MCB. I'm also grateful to Carlos Garcia and David Apilado Sr., who have provided additional GAO support, and to Sue Burchfield and others working in requisitions at Berkeley, who helped me navigate purchasing supplies for the lab to function.

Computational Biology Community

As a Designated Emphasis student, I was able to meet and interact with many members of the Computational Biology community over the years. I'm grateful to Haiyan Huang, who served on my qualifying exam committee and gave me advice about how to develop statistical training. I'm grateful to Nir Yosef for serving on my dissertation committee and providing me feedback on quantitative analyses, and for teaching the Comp Bio Doctoral Seminar, where I learned a great deal about genomics and bioinformatics. I'm grateful to Shaked Afik, James Kaminski, and Tal Ashuach of the Yosef Lab for advice on analyzing ATAC-Seq data.

Integrative Biology Community

The Patel Lab, during its time at Berkeley, sat at the juncture between the Molecular and Cell Biology and Integrative Biology communities. As a result, I had the opportunity to work with and learn from many folks in IB. I'm grateful to Lydia Smith of the Evolutionary Genetics Lab for experimental support and reagents. I'm grateful to Kamalankar Chatla of the Bachtrog Lab, who along with Aaron Pomerantz taught me how to use Oxford Nanopore sequencing and spent many long days troubleshooting experiments from several defective flow cells in a row. I'm grateful to fellow Comp Bio Designated Emphasis students in IB, Nicolas Alexandre and Ana Lyons, for being constant friends and dissertation-writing teammates and for our collaborative efforts with Aaron Pomerantz to organize the Endless Forms symposium.

Broader Scientific Community

Thank you to Meredith Protas, who has been a collaborator, mentor, and confidant over the last seven years. I've greatly enjoyed having the chance to work with her on *Asellus*. I appreciate her giving me the opportunity to get my hands into RNA-Seq data and to learn how to analyze this data (very slowly) through lots of troubleshooting and experimentation on the side. The work we did together gave me the space to grow as a computational biologist even when I didn't yet have my own data and allowed me to improve my analytical skills through independent exploration.

I'm grateful to the faculty and scientists I had the chance to work with during my teaching experiences at Woods Hole Embryology in 2018 and 2021, including Team Arthropod members Matthew Ronshaugen, Tom Pettini, Rachel Smith-Bolton, Sally Horne-Badovinac, and Seth Donoughe. I appreciated their scientific guidance, friendship, and career advice, as well as having the opportunity to work well together despite the intensity and chaos of teaching Embryology. I'm also grateful to the students in the Embryology courses that I had the chance to work with and learn from.

I'm grateful to members of the *Parhyale* community who gave me advice, including Leo Blondel of Cassandra Extavour's lab for advice on genome annotation and Damian Kao of Aziz Aboobaker's lab for advice on transcriptomics and genomic analyses. I'm also grateful to Tassos Pavlopoulos and Arnaud Martin for experimental advice and feedback. I'm grateful to have had the opportunity to work with Henrique Marques-Souza and David Wilcockson when they visited the Patel Lab. I'm grateful to Jonathon Stillman and his lab, as well as Anthony Petalunas, for advice on working with *Neocaridina*. I'm grateful to Helen Poynton, James Lazorchak, William Thoeny, and Donald Weston, who gave me access to *Hyalella azteca* for some projects that didn't make it into this thesis. I'm grateful to Heather Bracken-Grissom and Joanna Wolfe for inviting me to speak at the International Crustacean Congress 9 in Washington, D.C. in 2019 and also for experimental and career advice.

Thank you to Ryoji Amamoto and Stephanie Tsai, who mentored me when I was an undergraduate and continue to give me great scientific and career advice to this day. Thanks as well to Paola Arlotta, Jessica Whited, and Douglas Melton, who gave me the opportunity to research during my undergraduate years that prepared me for the work presented in this dissertation.

Classmates

I'm grateful to the MCB class of 2015 for being an incredibly fun, kind, collaborative, smart, and sociable group of people. Thanks to Shally Margolis for being my partner and collaborator in the injection room and for letting me observe the adorable *Nematostella* adults and embryos you studied. Thanks to Ashley Albright for being an amazing friend, MCB GSO Co-Chair, and confidant and for many great impromptu outings to grab boba or other beverages. Thanks to Helen Vander Wende for being a friend and collaborator in many departmental initiatives and for all of our efforts on shared initiatives on the GAC. Thanks to Mathew Summers for supporting me through some rough times and for our long and winding chats about science and life. Thanks to my "co-tons" Dylan Carter and Mark Stepaniak for being great comrades and labmates in our first year.

I'm also grateful to other folks in MCB in both younger and older cohorts who served as amazing friends and mentors over the years. Thanks to Bria Castellano, Diya Das, Nora Kostow, Natalie Dall, Andrea Higdon, and many others for your leadership and advice on issues ranging from departmental advocacy to quantitative training. Thanks to Tess Linden and Arik Shams for being amazing friends and for inviting me to many enjoyable hangouts and for our great Follies collaborations.

I'm grateful to Mathew Summers, Ryan Muller, Samantha Fernandez, Alec Heckert, Thomas Laughlin, Sara Popham, and Joseph Aman. Without the camaraderie and silliness of the Goldheart Companions and our other subsequent misadventures, I don't think graduate school would have been

nearly as tolerable. I'm grateful to Arik Shams, Tess Linden, Phil Frankino, Raymond Blackwell, and Noah Hermalin for going on various adventures through the Sword Coast and the Feywild. It's been very fun to introduce most of you all to D&D and to go on a big adventure together.

Friends and Family

Thanks to my partner Eric Wilbanks, without whom I could not have found the balance and stability I needed to make it to the end of this journey. You surprise me every day with your humor, charm, intellect, and creativity. Your love and support got me through some extremely dark, sad, and even terrifying moments. I am eternally grateful to you.

I'm also grateful to the many friends I had outside of Berkeley who kept me sane and grounded on this long and difficult journey. Thanks to Jordan Ugalde, whom I have known for nearly 15 years, for being a voice of both reason and chaos throughout my teen and adult life. I'm so grateful we've been able to grow into our full selves and to help each other improve and develop. Thanks to Victoria Zhuang, who believed in me throughout my college years and continues to be an amazing friend and confidant. I'm hoping to start writing prose again when this dissertation is finally done! Thanks to Daniel Windham, who has introduced me to many interesting people and has gone out of his way many times to visit me from the distant reaches of the Peninsula whenever he's in the Bay. I've appreciated our intellectual exchanges and watching each other grow in our careers and our lives over the years.

Thank you to the Manzanilla Company: Alex Dizon, Lih Seng Goh, Zach Hallberg, Anthony Le, Jon Titus, and Eric Wilbanks. Through over 50 sessions and hundreds of hours, we've been able to tell silly, moving, and thrilling stories together. I'm grateful to have a space to engage in uninhibited creativity with you all, which has been a fantastic creative outlet to keep me sane. I'm very excited to see how things continue to evolve in our ever-expanding narrative.

Thank you to Edwin Ko, Alice Shen, Eric Wilbanks, Paxton Oberg Wilbanks, and Justin Wilbanks for letting me explore my spookier side. It's been very fun exploring Barovia with you all, and I hope to continue to explore many different worlds with you all in the future.

Thank you to my coven: Kevin Sun, Jon Titus, and Eric Park. Your friendship was a beacon for me through the most brutal parts of this process. Though we're now geographically much more scattered, I'll never forget the many nights we spent cooking food, watching silly movies or playing games, or going clubbing and sometimes drinking a little too much. I'm grateful to have spent my early to mid twenties with all of you and for the chosen family we became. Reunion tour soon?

Thank you to Claire Sun, my sister, who has tolerated me being the "third parent" in the household. I'm grateful we're able to share many interests and to regularly keep in touch, and to lean on each other through some very painful and difficult moments.

Finally, thank you to my parents, Min Shan and Siquan Sun. I'm grateful for the sacrifices and hard work you had to put in to give me the opportunity to make it this far in life. I hope you know that I'm always trying to make the most of what you've given me. Thank you for your continued support and advice. I'm looking forward to celebrating with you soon!

Preface

The genetic and genomic basis of crustacean body plan evolution

Morphological diversity has been a topic of fascination from the earliest days of biological research. The classic work of scientists such as Gregor Mendel and Charles Darwin was focused on morphological variation in diverse organisms, including traits such as the color of flowers or the dimensions of finch beaks. Their work provided an understanding that diverse traits are inherited and may change over time. Others, such as Carl Linnaeus and Ernst Haeckel, cataloged and categorized diverse organisms based on their morphology, allowing us to better understand the breadth and the connections between diverse species. Much of the work of early biologists thus provided crucial theoretical insights into morphological diversity. However, the precise physical basis of these inherited differences – nucleotide-based genetic material and amino-acid based protein function – would not be discovered until centuries later.

In the contemporary study of biology, we now have a clear understanding of the material basis for many biological phenomena. The work of many scientists has revealed that a central dogma controls all life on our planet: DNA is transcribed into RNA, which is translated into protein. We have the tools to explore the composition and function of the "genes" that underpin this central dogma using genomics and genome editing. We also have technologies developed in molecular biology, biochemistry, microscopy, embryology, cell biology, and innumerable other perspectives that allow us to more deeply understand gene function within an organism. As such, it is now possible to investigate the physical basis behind the morphological diversity that has long fascinated biologists.

In this thesis, I examine the genetic and genomic basis of morphological diversity in crustaceans, with a focus on the crustacean body plan. Diverse groups of crustaceans show differences in the identity and organization of limbs along the anterior-posterior axis of the body. Researchers have begun to leverage technological advances from diverse biological fields to uncover the molecular basis of these morphological differences. In this thesis, I apply functional genomics, molecular biology, and developmental biology techniques to understand the development and evolution of three crustacean species: the sandhopper *Parhyale hawaiensis*, the pond louse *Asellus aquaticus*, and the cherry shrimp *Neocaridina davidi*.

Overview by Chapter

The first portion of the thesis focuses on the amphipod crustacean *Parhyale hawaiensis*, a tractable genetic model system for studying crustacean development. Work in *Parhyale* has benefited from technological advances ranging from advances in live imaging to the discovery of RNA-guided genome editing. I advance the study of *Parhyale* by generating new functional genomic resources. I also apply functional genomic resources and genome-editing technology to identify the function of Polycomb and Trithorax group genes in *Parhyale* development.

Chapter 0

In Chapter 0, we describe our understanding of *Parhyale* biology (circa 2019) and how work in this organism has allowed biologists to understand diverse evolutionary processes. I provide a detailed description of the known cell biological and genetic processes that regulate *Parhyale* development; comprehensively describe the genetic and genomic resources currently available; and describe two case studies for how *Parhyale* has expanded our understanding of body plan evolution and the evolution of regeneration. This chapter was previously published as a review in *Wiley Interdisciplinary Reviews: Developmental Biology*.

Chapter 1

The *Parhyale* community has developed substantial tools and resources to identify coding genes in the genome and understand their function. However, the study of gene regulation in *Parhyale* has been limited. In Chapter 1, we adapt protocols and advances in the field of functional genomics to *Parhyale* in order to identify gene regulatory elements and infer their function. We improve the genome annotation using newer RNA-Seq and genome annotation approaches and identify numerous regulatory elements across the genome using Omni-ATAC-Seq. By integrating gene expression and accessibility information, we identify distinct groups of regulatory elements with shared function and reveal concordant and discordant relationships between regulatory element accessibility and gene expression. Moreover, we identify two new functional regulatory elements in *Parhyale* using a reporter gene approach. This work will provide a substantial genomic resource for the *Parhyale* community. This chapter was released on *bioRXiv* in 2021 and is accepted for publication at *Development*.

Chapter 2

To understand the regulation of the body plan in *Parhyale*, in Chapter 2, we identify and perturb the function of Polycomb and Trithorax group genes using CRISPR-Cas9 mutagenesis. Previous work has demonstrated the role of *Hox* genes, a group of genes crucial for body regionalization across animals, in *Parhyale* development. This work demonstrated that these genes work in combinatorial and cross-regulatory fashion to pattern the *Parhyale* body plan, a network which we refer to as the *Parhyale* "*Hox* code." However, the upstream regulators of *Hox* expression in *Parhyale* remain undiscovered. By leveraging knowledge from two organisms for which *Hox* regulatory mechanisms have been thoroughly studied – the vinegar fly *Drosophila melanogaster* and the common mouse *Mus musculus* – we identify two groups of genes for further study in *Parhyale*: Polycomb Group and Trithorax Group genes. These general transcriptional activators and repressors are crucial for *Hox* regulation in both *Drosophila* and *Mus*, and perturbation of the function of these genes results in homeotic transformations. As such, we wanted to understand the function of these genes during *Parhyale* development.

Using CRISPR-Cas9 mutagenesis, we ablate a number of *Parhyale* Polycomb and Trithorax Group genes to reveal that Polycomb Group genes appear to play an essential role in *Hox* regulation, whereas Trithorax Group genes do not. Then, using *in situ* hybridization chain reaction (HCR), we

reveal that each *Hox* gene in *Parhyale* appears to have distinct regulation by Polycomb Group genes in developmental timing, tissue location, and interaction with specific Polycomb Group genes. By combining our understanding of the *Drosophila* regulatory system with the current literature in *Parhyale*, we determine that crustaceans may have additional layers of *Hox* regulation than those found in insects. Variations in these additional layers of regulation may be directly responsible for crustacean body plan evolution. This chapter was released on *bioRXiv* in 2022.

In addition to examining *Parhyale*, I also aimed to establish additional genetic model systems to understand the evolution of the crustacean body plan. To demonstrate that the "*Hox* code" that we have established for *Parhyale* can be analogized across diverse crustaceans, we must perturb the function of these genes in different species. However, at the start of this work, genetic perturbation was not possible in other Malacostracan crustaceans. As such, it was necessary to establish additional crustacean genetic model systems. I focused my efforts on two species: the isopod crustacean *Asellus aquaticus* and the decapod crustacean *Neocaridina davidi*, each of which has a different body plan from *Parhyale*. While I was unable to establish genetic manipulation for these organisms, by characterizing the development of these organisms and describing embryological techniques, I provide a framework for future researchers to establish genetic perturbation.

Chapter 3

The isopod crustacean *Asellus aquaticus* is an emerging ecological model system for understanding diverse evolutionary processes. *Asellus* has been studied for the relationship between its microbiome and environment, the evolution of distinct cave and surface ecotypes, and numerous other evolutionary processes. However, our understanding of *Asellus* embryonic development is limited and genetic perturbation has not yet been established in this organism. We describe *Asellus* embryonic development through a series of morphological stages. We also provide suggestions for improvements to animal husbandry protocols that enable the regular collection of embryos for embryonic injection, and identify the latest stages at which embryo injection would be possible. Finally, we describe preliminary results which suggest that genetic perturbation may be possible.

Chapter 4

Different populations of *Asellus* have independently migrated into caves in Southeastern Europe and evolved distinct cave populations. These cave animals display troglobitic traits including loss of pigmentation and loss of eyes. We examine the genetic basis of this trait evolution using differential RNA-Seq expression analysis to identify genes that are more highly expressed in either surface or cave embryos. We also explore the genomic basis of this evolution by using allele-specific expression analysis in hybrid embryos, which allows us to determine whether differential expression between populations might be caused by cis- or trans-regulatory mechanisms. We reveal that many genes that are differentially expressed between populations are likely to have cis-regulatory evolutionary differences. The candidate genes we identify will serve as an interesting set of target genes for future perturbation studies. Moreover, the transcriptomic resources will enable other researchers in

Asellus to identify embryonic genes – such as the *Hox* genes – for future mutagenesis studies. This chapter was previously published as a manuscript in *Genes (Basel)*.

Chapter 5

The decapod crustacean *Neocaridina davidi* is a popular aquarium pet. Previous researchers have described basic genome sequence data and proposed to establish *Neocaridina* as a decapod genetic model system. We describe the development of *Neocaridina* and challenges with establishing injection in *Neocaridina* embryos, which develop under tremendous intra-embryonic pressure. We provide a guide that enables future researchers to stage embryos based on key morphological landmarks and use antibody staining to investigate protein expression. Our data suggest that *Neocaridina* may exhibit unexpected expression of the *Hox* genes *Ubx* and *abd-A* compared to other decapods. While we were unable to establish injection in this organism, other researchers report being able to successfully use CRISPR-Cas9 mutagenesis, albeit at a very low success rate. By combining our troubleshooting approaches with the techniques described in their manuscript, it may be possible to establish *Neocaridina* as a tractable decapod genetic model system.

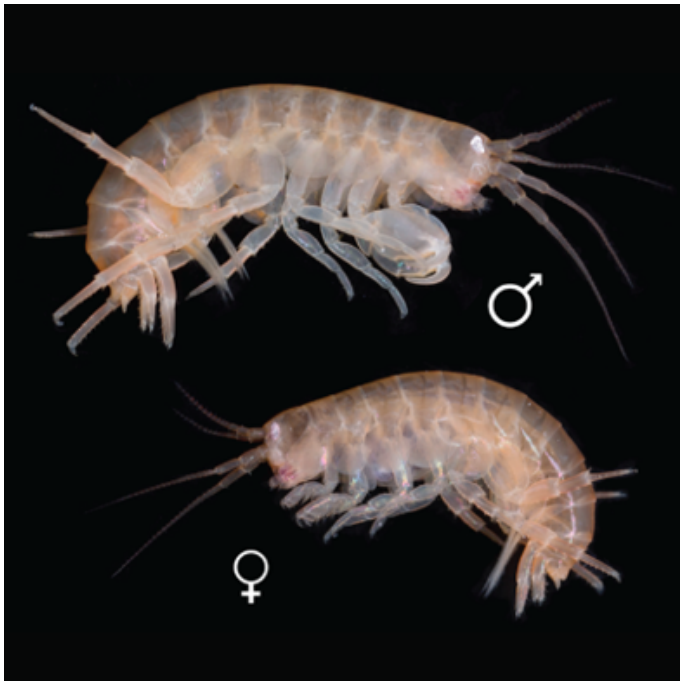
0 – The amphipod crustacean *Parhyale hawaiensis*: an emerging comparative model of arthropod development, evolution, and regeneration

Dennis A Sun[1] and Nipam H Patel[2,3*]
1. Department of Molecular and Cell Biology, University of California, Berkeley
2. Marine Biological Laboratory, Woods Hole, MA
3. Department of Organismal Biology and Anatomy, University of Chicago, Chicago, USA
* corresponding author

Previously published in Wiley Interdisciplinary Reviews: Developmental Biology

Abstract

Recent advances in genetic manipulation and genome sequencing have paved the way for a new generation of research organisms. The amphipod crustacean Parhyale hawaiensis is one such system. Parhyale are easy to rear and offer large broods of embryos amenable to injection, dissection, and live imaging. Foundational work has described Parhyale embryonic development, while advancements in genetic manipulation using CRISPR-Cas9 and other techniques, combined with genome and transcriptome sequencing, have enabled its use in studies of arthropod development, evolution, and regeneration. This article introduces Parhyale development and life history, a catalog of techniques and resources for Parhyale research, and two case studies illustrating its power as a comparative research system.



Caption

The amphipod crustacean Parhyale hawaiensis is an emerging research system for the study of arthropod development, evolution, and regeneration. Image in graphical abstract generated by Aaron Pomerantz.

0.1 – Introduction: Choosing a Model System

Research organisms have long been essential to the study of genetics. Early animal systems, such as Drosophila melanogaster and Mus musculus, were used to uncover basic genetic principles through the study of spontaneous and induced mutant alleles (Cuenot, 1905; Sturtevant, 1965). The ease of husbandry and short generation time of Drosophila in particular have enabled genetic screens that have revealed basic principles of animal development, such as axis formation, segmentation, body regionalization, cell fate specification, and tissue morphogenesis. As genetics became more specialized, researchers began to choose additional organisms for unique traits that made them ideal for studying particular topics. For example, the roundworm Caenorhibitis elegans was first established to study the development of the nervous system (Ankeny, 2001; Brenner, 1974). Its simple behavior, short generation time, and hermaphroditic lifestyle allowed for genetic mutant analysis of traits otherwise difficult to study in other systems. Around the same time, the zebrafish Danio rerio was established to understand vertebrate genetics and development (Streisinger et al., 1981). Its small, transparent embryos proved useful for identifying developmental mutants.

New discoveries in each of these systems expanded the breadth of their research potential, while a number of technological breakthroughs greatly enhanced their usefulness and provided new ways to address long-standing biological questions. Examples include creating transgenic animals using transposable elements or homologous recombination, imaging molecular events in living animals using fluorescently labeled proteins, and carrying out reverse genetic screens using RNAi.

Work in such so-called "model" systems has uncovered basic principles of molecular, cell, and developmental biology. However, these classical systems are separated by vast phylogenetic distances and represent a tiny fraction of the planet's biological diversity. To understand the evolution of animals, it has become increasingly important for biologists to develop novel research systems. These systems are sometimes selected to bridge phylogenetic gaps between current systems, and enable investigations into the evolution of fundamental biological processes. In other cases, novel systems possess interesting properties not found in classical systems, and are thus able to reveal previously unknown biological processes. Finally, comparisons between novel and well-established systems can enable deeper understanding of animal development and diversity.

One of the outstanding questions in evolutionary biology is how animals have diversified over hundreds of millions of years. The most successful group of animals on the planet, measured by biomass (Bar-On et al., 2018) and species diversity, is the arthropods, and their diversity has been the subject of scientific curiosity and investigation for centuries. Arthropods have many shared features, including their namesake jointed appendages (arthro – joint; pod – foot) and their hard exoskeletons, which they molt repeatedly over their lifetimes. Along with insects, other arthropod lineages include chelicerates (spiders, mites, and scorpions), myriapods (centipedes and millipedes), and non-insect crustaceans (copepods, crabs, barnacles, and many other organisms)(Fig. 1a). These organisms contain tremendous diversity in their development, body plan, and life history. However, much of the research on arthropod diversity has relied on a handful of genetically tractable research systems.

In the field of arthropod development, Drosophila melanogaster remains the predominant research system. Drosophila has been the subject of many foundational studies, such as the discovery of the hierarchy of genes involved in segmentation (Nüsslein-Volhard and Wieschaus, 1980), body

regionalization (Lewis, 1978), and embryonic axis formation (Driever and Nüsslein-Volhard, 1988a, 1988b), and many other developmental processes. Drosophila are easy and inexpensive to rear, tools for genetic manipulation are well established, and animal stock centers maintain numerous mutant lines for the large Drosophila research community. Over the years, various databases have been built to store genetic, genomic and bioinformatic data for Drosophila, such as Flybase.org. With the discovery of genome-editing technology such as CRISPR-Cas9, mutant and transgenic alleles are even easier to generate. Thus, Drosophila remains a key system for the study of many biological phenomena.

Though Drosophila offers many resources, it represents but one type of organism within the tremendous morphological diversity of arthropods. For example, Drosophila's two-winged, six-legged body plan stands in contrast to other insects, which may have two pairs of wings, or no wings at all. Arthropods outside of the insects have even greater body plan diversity, with extreme examples in the myriapods, which can have over one hundred pairs of limbs. Comparative studies have also demonstrated that some aspects of Drosophila development are derived relative to those of other arthropods. During embryogenesis, Drosophila undergoes long-germ development, wherein all body segments form simultaneously towards the end of a syncytial stage (Davis and Patel, 2002). In contrast, most insects undergo short-germ or intermediate-germ development, wherein only anterior segments are specified during the early syncytial stage, while more posterior segments are successively established in a cellular environment; in contrast, other arthropod groups lack a syncytial stage entirely (Davis and Patel, 2002). Moreover, some key Drosophila segmentation genes, such as the homeobox-containing genes fushi tarazu, even-skipped, zen, and bicoid, have changed dramatically in expression and function when compared to homologs in species within early-branching arthropod lineages (Hughes and Kaufman, 2002). Thus, while Drosophila has provided an excellent platform for studying general developmental principles, some aspects of its development are not representative of more general, and apparently ancestral, arthropod development.

Other models have expanded upon the foundational work in Drosophila to inform our understanding of arthropod development and evolution. For example, the flour beetle Tribolium castaneum has been offered as a more representative insect research system. Its intermediate-germ development and leg-bearing larval stages are features shared with a broader diversity of insect clades (Davis and Patel, 2002; Lynch et al., 2012). Tribolium has been used as a comparative model to examine how pair-rule genes (Brown et al., 1994; Choe et al., 2006; El-Sherif et al., 2012; Sarrazin et al., 2012) Hox genes (Beeman, 1987; Tomoyasu et al., 2005), and other developmental programs that have evolved within Insecta, particularly by comparing its developmental processes to those of Drosophila. Such studies have provided mechanistic insight into how genes and developmental processes change over evolutionary time. In addition to gene expression data, genetic manipulations in Tribolium, such as parental RNAi (Bucher et al., 2002), have provided methods to efficiently probe gene function, and even carry out forward genetic screens (Christian et al., 2015; Knorr et al., 2013). The development of such techniques has made Tribolium an appealing system for testing hypotheses about the course of evolution at long timescales. Beyond insects, gene expression data and some functional data from crustaceans, chelicerates, myriapods, and onychophorans have provided insights into the evolution of body plan diversity within the arthropods and other closely related clades (Hughes and Kaufman, 2002; Janssen et al., 2014; Pace et al., 2016; Paese et al., 2018; Schwager et al., 2007, 2009). By

synthesizing the observations made in these disparate systems, evolutionary biologists can develop a more comprehensive understanding of the genetic and developmental processes that lead to organismal diversification.

To enable broader comparative studies of arthropod development, Browne et al. (2005) established the amphipod crustacean Parhyale hawaiensis (Dana, 1853) as a novel research system (Gerberding et al., 2002). P. hawaiensis (hereafter referred to in short as Parhyale) is a phylogenetically strategic choice that serves as an outgroup for insects (Oakley et al., 2013; Schwentner et al., 2017). As a crustacean, it serves as a representative member from an ancient group of organisms that traces its origin to the Cambrian era (Schwentner et al., 2017), and which have remained tremendously successful in diversification since their emergence. As a malacostracan crustacean, it is also closely related to commercially important shrimp, lobsters, and crabs. Moreover, its life history makes it amenable to laboratory culture and molecular genetic manipulation. Parhyale are shallow-water, circumtropical detritivores with a high tolerance for environmental osmotic changes (Clarence R Shoemaker, 1956; J Laurens Barnard, 1965; Poovachiranon et al., 1986). Their large embryos are easy to collect and image; though the yolk is opaque, the superficially-positioned embryo and surrounding chorion are transparent, and a thorough staging scheme enables researchers to identify the relative developmental stage of any embryo (Browne et al., 2005). In contrast to many other crustaceans, which emerge as planktonic nauplii larvae that subsequently metamorphose into adults and can thus be difficult to culture, Parhyale hatchlings emerge as miniature versions of adults, and can easily be raised to adulthood to establish new genetic lines. These features have made Parhyale an ideal candidate to become a new research system, and indeed, continued technique development over the last two decades has advanced its tractability. Published protocols describe techniques for embryo injection (Rehm et al., 2009a), dissection (Rehm et al., 2009b), antibody staining (Rehm et al., 2009c), and in situ hybridization (Rehm et al., 2009d), enabling new researchers to explore Parhyale with ease.

An explosion in nucleotide sequencing technologies has facilitated gene discovery in Parhyale, while advances in genetic manipulation have enabled studies of gene function. Such technological advancements have enabled Parhyale researchers to examine several aspects of Parhyale development and address some important issues in the evolution of arthropods. Initial genetic approaches used the Minos transposase system to generate transgenic animals for misexpression (Pavlopoulos et al., 2009), while morpholinos, injected siRNA, and expressed RNA hairpins have enabled gene knockdown (Liubicich et al., 2009; Martin et al., 2015; Özhan-Kizil et al., 2009). Since the start of the genome-editing revolution, CRISPR-Cas9 programmable nucleases have been used to generate knock-out and knock-in animals (Kao et al., 2016; Martin et al., 2015; Serano et al., 2016). Meanwhile, early genomics efforts included the isolation of conserved genes using PCR and RACE, as well as the generation of a BAC library, from which early insights into genome structure were achieved (Parchem et al., 2010). As sequencing costs have plummeted, several different groups have published transcriptome sequencing data of both the adult and embryonic animals (Blythe et al., 2012; Nestorov et al., 2013; Zeng and Extavour, 2012; Zeng et al., 2011). A recent paper describes the Parhyale genome, positioning Parhyale for a new age of genome-scale investigations (Kao et al., 2016). Moreover, an active and expanding group of Parhyale researchers has begun to curate protocols and other resources to facilitate the expansion of the community (see “Further Reading” below.)

Since its introduction as a research system almost two decades ago, Parhyale has enabled investigations that address key questions in developmental and evolutionary biology. For example, its short-germ development has allowed investigation into mechanisms of embryonic axis formation and segmentation in the absence of a Drosophila-like embryonic syncytium (Browne et al., 2005; Vargas-Vila et al., 2010)(Extavour, 2005; Gupta and Extavour, 2013; Nestorov et al., 2013; Özhan-Kizil et al., 2009). These studies have offered clues to the ancestral function of genes critical to dorsal-ventral patterning and neural specification in arthropods. In another set of studies, researchers discovered correlations between germ layer identity and the differential distribution of mRNAs and other cytoplasmic components at the 8-cell stage of development (Extavour, 2005; Gupta and Extavour, 2013; Nestorov et al., 2013; Özhan-Kizil et al., 2009). Parhyale has also been used to explore how germ layers communicate during development, and how the embryo compensates within germ layers to recover lost tissue (Hannibal et al., 2012; Price et al., 2010).

In this review, we summarize much of the work that has been accomplished to date. We begin with a comprehensive summary of Parhyale life history and embryonic development, and then describe techniques for genetic manipulation and resources for genomic analysis. Finally, we describe two case studies that illustrate the power of Parhyale as a comparative system for studying body plan evolution and regeneration. We review how work in Parhyale has provided insights into the evolution of arthropod body plan diversity, through studies examining the expression and function of Hox genes in this outgroup for Drosophila and all other insects. Then, we discuss how a striking discovery about the mechanisms of Parhyale limb regeneration mirrors findings from vertebrate systems, and how the optical properties of this small crustacean make it an ideal platform for future regenerative studies.

0.2 – Development, Anatomy and Life Cycle

Within Crustacea, Parhyale belong to the class Malacostraca (Fig. 0.1a). The Malacostracan clade is composed of familiar shrimp, krill, crabs, and lobsters, in addition to crustaceans with other unique body forms. Amphipoda, the order to which Parhyale belongs, are commonly referred to as scuds, side swimmers, or beach hoppers, and are a ubiquitous sight on beaches. Parhyale are detritivores with a circumtropical, intertidal, and shallow-water marine distribution (Clarence R Shoemaker, 1956; J Laurens Barnard, 1965). The animals are easy and inexpensive to rear and reproduce year-round, making them ideal for laboratory work.

Parhyale embryos can be collected from the ventral brood pouch of adult female animals without harm to either the mother or the embryos. Adult females can grow up to 20mm in length, and can produce large broods of embryos (50+). Development has been thoroughly characterized into a series of morphologically identifiable stages (Browne et al., 2005)(Fig. 0.1b), and is consistent with observations in other amphipod crustaceans, such as Orchestia cavimana (Wolff and Scholtz, 2002). Embryonic development lasts about 11 days, beginning with a mediolateral holoblastic (total) cleavage at 4 hours that generates a 2-cell embryo (Fig. 0.2a). After two additional stereotyped divisions, the embryo at the 8-cell stage contains four micromeres (small cells, abbreviated in lowercase) and four macromeres (large cells, abbreviated in Sentence Case) (Fig. 2b). At 12 hours, the embryo contains ~100 cells of about equal size. This is termed the ‘soccerball’ stage for its appearance. Shortly after, cell cytoplasm separates from yolk, and cells migrate along the surface of the embryo into two predominant

clusters of cells, one at the presumptive ventral end of the embryo, which will produce the ectoderm anlagen, and the other into a small circular cluster of visceral mesoderm and germ cells, referred to as the 'rosette' (Fig. 2b). After this stage, the majority of Parhyale morphogenesis remains at the surface of the embryo, allowing for easy live imaging of development.

Gastrulation occurs at around 20hrs, when the rosette migrates beneath the ectoderm anlagen (Alwes et al., 2011; Browne et al., 2005; Chaw and Patel, 2012)(Fig. 2c). This generates the 'germ disc' stage. Germ disc cells then migrate and proliferate in the transition to the 'germband' stage (Browne et al., 2005)(Fig. 2d, e). During the initial phase of germband elongation, cells of the posterior ectoderm organize into transverse rows of cuboidal cells referred to as "parasegment precursor rows" (PSPRs), which together create a grid-like germband (Fig. 2f). The first of these cells to become morphologically distinct will align along the midline of the embryo and express orthodentical (Browne et al., 2006). These midline precursor cells will later express single-minded, which is critical for specifying the DV axis of the Parhyale embryo (Vargas-Vila et al., 2010). Two waves of mitotic division sweep across the PSPRs, beginning adjacent to the midline and spreading laterally, generate two cells (a/b and c/d), then four cells (a, b, c, and d) from each original row of cells. These waves of division are highly organized; more anterior PSPRs form first and begin dividing as more posterior PSPRs are beginning to organize. The anterior-most cell row (a) expresses engrailed; each original PSPR row is thus equivalent to a parasegment in Drosophila (Fig. 2h). The sequential addition of engrailed stripes provides a benchmark for staging Parhyale embryos (Browne et al., 2005).

A well-documented feature of Parhyale development is the presence of precise lineage restriction, particularly between ectoderm and mesoderm (Gerberding et al., 2002; Price et al., 2010). Lineage tracing has revealed that Parhyale restricts germ layers at the 8-cell stage (Fig. 2b). Two of the macromeres, 'El' and 'Er', produce the anterior left and anterior right ectoderm respectively, while the macromere 'Ep' produces the posterior ectoderm and the midline of the germband. During germband elongation, there is some mixing between descendants of 'El' and 'Er' in the head segments, but the gnathal and thoracic segments appear to have a distinct left-right division. Additionally, the boundary between 'El'/'Er' contribution and 'Ep' contribution to the posterior ectoderm appears to be determined stochastically.

Cell ablation experiments have demonstrated that 'El', 'Er', and 'Ep' represent an "ectoderm group" that is able to undergo intra-germ layer compensation as late as gastrulation, but not after initial germband formation (Price et al., 2010)(Fig. 3a). 'Ep' lineage cells, when ablated, can be replaced by 'El' and 'Er' cells together, while 'Ep' cells can also replace either an ablated 'El' or 'Er'. Notably, 'El' and 'Er' cells retain their left/right distinctions when compensating for the ablated 'Ep' cell.

Two micromeres and one macromere, 'ml', 'mr' and 'Mav', together produce the mesoderm, including the head mesoderm, the visceral mesoderm, and a set of teloblasts that produce the segmented mesoderm of the body (Gerberding et al., 2002; Price et al., 2010). The anterior mesoderm, primarily composed of 'Mav' descendants with posterior contribution from 'ml and 'mr', forms two circular populations of cells at the germband stage that eventually migrate ventrally to form the head and visceral mesoderm (Fig. 2e). The posterior teloblastic mesoderm, which arises from 'ml' and 'mr' descendants, produces the left and right mesoderm respectively (Fig. 2e). Lineage mixing between 'ml' and 'mr' descendants is rare, but occasionally observed. During germband elongation, eight

mesodermal stem cells (four on each side of the embryo; descended from mr and ml; called mesoteloblasts) migrate posteriorly, depositing rows of daughter cells, termed mesoblasts, beneath the ectoderm (Gerberding et al., 2002) (Fig. 2g). (For a detailed analysis of the early mesoteloblast divisions, see Price and Patel, 2008). The mesoblast rows eventually align with each ectodermal parasegment, and go on to divide and differentiate into the muscle cells of each body segment (Price and Patel, 2008).

Cell ablation experiments have also demonstrated that the 'ml', 'mr', and 'Mav' cells represent a "mesoderm group" (Price et al., 2010)(Fig. 3d). Ablation of any one of the cells can be compensated for by other members of the group. As with the ectoderm group, 'ml' and 'mr' compensate for an ablated 'Mav'. Moreover, ablation of two of the cells in the mesoderm group can be compensated for entirely by the single remaining cell. Ablating all three mesoderm group cells results in a mesoderm-free embryo, indicating that there is no inter-layer compensation between mesoderm and ectoderm lineages (Price et al., 2010).

Additional ablation experiments have demonstrated that proper mesodermal segmentation is dependent on signals from the ectoderm, while the ectoderm is capable of segmentation and morphogenesis independent of mesoderm (Hannibal et al., 2012). Ablation of all three mesoderm group cells results in an embryo with properly patterned ectoderm, but no muscle tissue, as visualized by Mef-2 in situ hybridization. Meanwhile, ablation of either the El or Er cell after the initiation of germband elongation, after which intra-germ layer compensation no longer occurs, causes defects in mesenteloblast division and migration (Hannibal et al., 2012).

The micromere that sits above the smallest macromere is named 'g' and gives rise to the germline (Gerberding et al., 2002). This cell inherits a specialized germ plasm that is sequestered as early as the 1-cell stage and is visible in living embryos (Gupta and Extavour, 2013). During early germband elongation, 'g' descendants sit in a single cluster beneath the ectodermal row that will give rise to the future mandible. This cluster divides and migrates in two separate clusters until each eventually settles in its final location on the dorsal side of the hatchling.

After limb development and organogenesis, animals hatch as miniature versions of adults (Browne et al., 2005). The body plan of Parhyale contains 20 segments, 19 of which have appendages. There are six head segments: an ocular segment (which does not contain an appendage), two antennal segments (An1, An2), and three segments with feeding appendages: mandible (Mn), maxillule (Mx1), and maxilla (Mx2). The thorax consists of eight segments, beginning with a segment bearing a feeding-like appendage called a maxilliped (T1/Mxp). T2 and T3 segments have chelipeds, or claws; in sexually mature males, the T3 cheliped is enlarged and used to grasp onto females before copulation. T4-T8 segments have pereopods, or locomotary legs; T4-T5 legs face forward and are used for walking, while T6-T8 legs are oriented in opposition to T4-T5, and are used when the animal backs up or "jumps" rapidly. Finally, the abdomen consists of six segments with biramous appendages, with A1-A3 segments bearing pleopods, or swimming appendages, and A4-A6 bearing uropods, or anchoring appendages.

About two months after hatching, animals become sexually mature. Adults are sexually dimorphic: females are smaller and have prominent ovaries, which are visible through the body wall, while males are larger, and have a pair of larger chelipeds on T3. Animals ready to mate can be found in

pairs, with males clasping females before copulation and releasing females after sperm has been deposited. This pre-copulatory amplexus allows for simple identification of animals ready to produce new embryos, allowing for easy collection of early embryos for injections to perform genetic manipulation and analysis.

0.3 – Tools for Genetic Manipulation

To investigate gene function in any research organism, it is critical to have tools to manipulate gene activity, as well as to generate mutant or transgenic alleles. In the field of evolutionary developmental biology, researchers are especially dependent on technology development, and often make use of techniques first developed in other research organisms. Over the past decade, techniques for gene knockdown, knockout, and transgenesis have been adapted for Parhyale. These techniques take advantage of the ease of collecting and injecting early-stage embryos, and the long time-window of the first several cell divisions.

Transgenesis in Parhyale was first achieved using the Minos transposon from Drosophila hydei (Pavlopoulos and Averof, 2005). This transposon has been used for transgenesis in a variety of animal systems, and consists of two terminal inverted repeats flanking a desired insert sequence (example in Fig. 4a). Transgenesis can be achieved by injecting a recombinant construct containing the inverted repeat sequences along with Minos transposase mRNA into a 1- or 2-cell Parhyale embryo. The Minos construct will be randomly inserted into the genome, integrating transgenic alleles in up to 30% of surviving hatchlings (Pavlopoulos and Averof, 2005). Minos was first tested by generating animals that expressed DsRed in cells around the eye using a synthetic 3xP3 enhancer, and was subsequently used to identify a muscle-specific enhancer upstream of the Parhyale hsp70 protein. This approach has allowed for the generation of many transgenic lines (examples in Fig. 4b, c)(Pavlopoulos and Averof, 2005).

More sophisticated transgenic approaches have also been developed for Parhyale. Temporal control of gene expression was first achieved in Parhyale through the identification of a heat shock element from the Parhyale hsp70 gene (Pavlopoulos et al., 2009). Placing transgenic animals bearing the hsp70 heat shock element at 37°C for 30min is sufficient to transiently induce high levels of transgene expression within a few hours. This technique has been used to misexpress Ubx in mosaic animals to study the effect of this Hox gene on segment identity across the Parhyale body plan (Pavlopoulos et al., 2009). Another method, called “iTRAC” for gene-trapping and trap conversion in Parhyale, used the Minos system to deliver a splice acceptor isolated from upstream of the Hsp70 gene to a random genomic location (Kontarakis et al., 2011). Minos donor plasmid containing Hsp70a-DsRed or EGFP was injected into 1-cell embryos to generate a diverse catalog of different transgenic lines, in which each established line labels a different portion of the animal, depending on available splice donor exons from genes in the neighborhood of the Minos insertion. Another feature of this construct is the presence of an attB or attP sequence, which allows for insertion of additional alleles or the replacement of alleles using the phi-C31 integrase system (Groth et al., 2004). Using such transgenic tools, one can label and clone developmental and tissue-specific genes.

Genetic knockdown techniques in Parhyale have been established using morpholino and RNAi technology. Morpholinos have been used to knock down the vasa gene, which is a well-studied germ cell determinant conserved across many arthropods (Özhan-Kizil et al., 2009). This experiment

suggested that Parhyale vasa is not necessary for the establishment of the 'g' cell, but it is critical for the maintenance of germline identity. At about the same time, small interfering RNAs (siRNAs) were used to knock down the leg patterning gene Distalless and the Hox gene Ultrabithorax (Liubicich et al., 2009), producing a loss of distal limb elements and a transformation of the T2 and T3 segment towards a more anterior T1 fate, respectively. In addition to siRNA, engineered RNA hairpins have also been used to achieve gene knockdown via RNAi. By expressing a heat-shock controlled abd-A sense/ antisense fusion transcript linked with a Drosophila white intron, Martin et al. were able to generate knockdown animals that helped uncover the function of the abd-A gene (Martin et al., 2015). These technologies often produce a range of phenotypes due to variation in the extent of knockdown, which can be useful for understanding how gene dosage contributes to specific phenotypes.

The field of genome engineering has recently been revolutionized by the discovery of the CRISPR-Cas9 programmable nuclease system (Doudna and Charpentier, 2014; Hsu et al., 2014; Jinek et al., 2012). Following double-stranded cleavage at a target site specified by a guide RNA, native DNA repair machinery can either induce indel mutations, or can be coopted to introduced exogenous DNA through homologous recombination, among many more advanced techniques. This transformative technology has enabled researchers of many organisms to perform genome editing and other genetic manipulations.

In Parhyale, the CRISPR-Cas9 system was first used to systematically knock out a large repertoire of Hox genes in order to examine their function in specifying segment and appendage identity in G0 embryos (Martin et al., 2015)(example in Fig. 5a). It was also used for homologous recombination with double-stranded DNA plasmids to generate animals expressing EGFP-T2A-Antp (Serano et al., 2016)(Fig. 5b). Recent work in Parhyale has demonstrated that even in phenotypically wild-type G0 animals, targeted alleles have a high rate of mutations within the germline, and loss-of-function mutations can be revealed by crossing these phenotypically wild-type G0 animals to each other and analyzing their G1 progeny (Kao et al., 2016). Moreover, NHEJ-mediated repair has also been used to generate transgenic animals (Kao et al., 2016). By using a double-stranded donor template without homology arms, one can generate fused transcripts at a higher rate than via a homologous recombination pathway. More recent protocols in other systems have shown that an alternate single-stranded DNA repair pathway can generate transgenic alleles at much higher efficiencies than homologous recombination (Chen et al., 2015; Quadros et al., 2016). Moreover, the identification and engineering of other CRISPR-Cas systems, including those that target RNA, or modify specific bases, has created a diverse toolkit for targeted genetic manipulations (Abudayyeh et al., 2017; Cox et al., 2017; Gaudelli et al., 2017; Komor et al., 2016). Future work in Parhyale will benefit from adapting additional advances in CRISPR-Cas targeting technologies to this emerging research system.

0.4 – Genomic and Bioinformatic Resources

Gene sequence and its analysis are instrumental to many molecular genetic approaches. For example, in situ hybridization requires the generation of labeled antisense probes from specific sequences of interest, while transgenesis using CRISPR-Cas9 requires sequences for specific targets in the genome. In newer research systems such as Parhyale, identifying conserved genes is critical for

comparative developmental genetic studies. Moreover, examining genome content and organization in Parhyale can provide insights into arthropod and metazoan genome evolution.

The initial identification of gene sequences in Parhyale relied on low-throughput approaches, such as degenerate PCR and RACE (Browne et al., 2006; Price and Patel, 2008; Prpic and Telford, 2008). The extensive efforts of early investigators generated sequences for the Hox genes and many other developmentally important Parhyale genes for performing expression and genetic analysis.

Analysis of large genomic sequences was spurred on by the construction of a BAC library consisting of 129,024 clones with an average insert size of 140kb (Parchem et al., 2010). This library amounted to five genome equivalents of the ~3.6Gb genome. Analysis of specific genes and gene families, including sequencing of several selected BAC clones, revealed several features of Parhyale, which could be compared to those in other arthropods specifically and metazoans more broadly. For example, the intergenic distance in the Parhyale genome is much greater than those of previously sequenced arthropods. Furthermore, analysis of coding regions in Pax group III and other genes revealed significantly greater intron size. These results are consistent with the very large genome size of Parhyale, and suggest that Parhyale has undergone genome expansion.

Transcriptome data has also been generated from various stages of embryogenesis and adult tissues (Blythe et al., 2012; Zeng et al., 2011). GO-term analysis of embryonic transcripts revealed no significant enrichment in genes of a particular function relative to Daphnia and Drosophila, while BLAST revealed an enrichment in transcripts containing a C2H2 zinc finger domain (Zeng et al., 2011). (Blythe et al., 2012) generated a great deal of additional transcriptome data, including two additional small RNA libraries from early embryos (S1-S4) and slightly later embryos (S4-germband) to identify 55 conserved miRNA sequences. Two miRNAs conserved to Eumetazoa (miR-10 and miR-100) and 9 out of 30 miRNA families conserved to Bilateria were identified in Parhyale using this approach.

A recently published paper on a Parhyale genome assembly describes in deeper detail the genomic features of this crustacean (Kao et al., 2016). Genomic DNA from a single adult male was used to generate a genome assembly containing 133,015 scaffolds (N50 of 81,190bp), 259,343 unplaced contigs, and 584,293 shorter contigs that could represent polymorphic regions. Annotation by Augustus, which uses high-confidence gene models generated using assembled transcriptomes, gene homology, and ab initio predictions, generated 28,155 gene models. Combined genomic and second-generation Illumina transcriptomic data accounted for 247 of the 248 CEGMA (Core Eukaryotic Genes Mapping Approach) orthology groups, which are a measure of assembled genome completeness (Parra et al., 2007). Moreover, 96% of new transcriptome reads map to the genome. Taken together, this suggests that the current genome assembly captures nearly all of the protein-coding information for Parhyale. This genome sequence is publically available through NCBI (https://www.ncbi.nlm.nih.gov/genome/15533).

Further transcriptomic and genomic analysis was carried out to identify noncoding RNAs, such as piRNAs, miRNAs, and lncRNAs, as well as their associated proteins. Parhyale has key components of all major small RNA pathways, including 4 argonaute family members, 2 PIWI family members, and orthologs of Dicer-1, Dicer-2, drosha, and loquacious (Kao et al., 2016). However, no evidence was found for SID-1, a necessary component of systemic RNAi that is found in C. elegans

and many species of mammals, fish, and insects. These findings will allow for investigations of small RNA pathway evolution through studies in a crustacean.

The Parhyale genome assembly revealed several additional genomic features. Analysis of transcriptomes representing Parhyale cultures from different labs, although all derived from the original population extracted from the Shedd Aquarium, revealed high levels of heterozygosity in many genes (Kao et al., 2016), a finding consistent with polymorphisms observed in overlapping BAC sequences (a mean of 1.5-2.5% SNPs). Moreover, analysis of BAC sequences revealed many large (~100bp) indels between haplotypes. Consequently, greater care must be taken when using homology-dependent genetic manipulations such as CRISPR-Cas9 in this system.

Recently, the Parhyale community funded an updated genome sequencing effort, which included Dovetail Chicago and Hi-C methods for stitching together large contigs and bridging gaps over repetitive elements. This effort has resulted in a dramatically improved genome sequence, which will facilitate deeper investigations into the regulatory information contained in the Parhyale genome.

Taken together, there is now a wealth of genomic and bioinformatics data for Parhyale, including BACs, a high-quality genome sequence, assembled transcriptomes for various stages of development, and annotated small RNAs. Such resources will enable deeper genetic investigations in this emerging research system.

0.5 – Case Studies

With thoroughly characterized development, many methods for genetic manipulation, and ample genomic resources, Parhyale has become a platform for a diverse portfolio of investigations. Here, we examine two topics that highlight the advantages of using Parhyale as a research system.

Hox genes and body plan diversity

First discovered in Drosophila, the Homeotic or Hox family of genes has become appreciated as a conserved mechanism for A-P patterning of animal body plans. These transcription factors are usually found in a collinear genomic cluster that corresponds with their expression along the anterior-posterior (AP) axis (William and Krumlauf, 1992). Given their broad conservation and critical developmental function, Hox genes have been used to explore arthropod body plan diversity. Gene expression data for Hox genes is available for all of the major arthropod clades, including chelicerates, myriapods, crustaceans, and many diverse species of insects (Hughes and Kaufman, 2002)(partially summarized in Fig. 6b). These data suggest a correlation between Hox gene expression and body plan evolution. However, given the difficulties of homologizing body regions between different types of arthropods, it has been difficult to determine how exactly Hox expression and body plan evolution are related.

Within the arthropods, crustaceans offer an interesting opportunity for Hox gene analysis, as they exhibit immense diversity, particularly in regards to limb morphology, yet have body regions that can be readily homologized (VanHook and Patel, 2008), at least within some clades. Crustacean Hox gene expression domains correlate with the presence of particular limb types, suggesting that AP body regionalization is controlled by Hox expression. However, until recently, there has been no functional validation of this hypothesis. Furthermore, when looking across crustacean body plans, changes in Hox

gene expression also correlate directly with changes to body plan organization. This suggests that changes to Hox gene expression are responsible for body plan evolution in crustaceans (Abzhanov and Kaufman, 1999a, 1999b, 2000, 2000; Averof and Patel, 1997; Hughes and Kaufman, 2002).

With its unique status as a crustacean with advanced genetic tools and genomic resources, Parhyale has become an excellent system for studying crustacean Hox genes. Recent studies in Parhyale have provided additional crustacean Hox expression data, and the first forays into crustacean Hox gene function in Parhyale using transgenesis and CRISPR/Cas9 mutagenesis have offered strong support for previous hypotheses of Hox-mediated crustacean body plan evolution.

The first investigations into Parhyale Hox genes focused on Parhyale Ultrabithorax (Ph-Ubx). In malacostracan crustaceans, this Hox gene was previously correlated with the transition between feeding type appendages in the anterior thorax (maxillipeds) and the locomotory appendages found more posteriorly on the thorax. Previous work has demonstrated that a posterior shift in the expression of Ubx correlates with an increase in the number of maxillipeds across a variety of crustaceans, suggesting that the removal of Ubx from a given thoracic segment allows for the emergence of maxilliped identity (Averof and Patel, 1997).

In Parhyale, Ubx is expressed in segments T2-T8, which include the chelipeds, forward walking legs, and reverse jumping legs (Liubicich et al., 2009). The T1 segment in Parhyale has a reduced size and modified morphology, and is classified as a maxilliped. RNAi knockdown of Ubx induces a transformation of the T2 and T3 segments (normally chelipeds, or claws) towards a T1/Mxp-like identity, a result in agreement with previous evidence that suggested that loss or reduction of Ubx expression could generate maxilliped morphology. In a complementary experiment, (Pavlopoulos et al., 2009) described the effects of Ubx misexpression using a heat-shock inducible Ubx construct. Ectopically induced Ubx expression in the antenna, Mx2, Mxp, and T2-T3 (chelipeds) produced a homeotic transformation into T4-T5 (forward walking legs). Frequently, Mx2 could be transformed into Mxp, which the authors correlated with repression of the more anteriorly expressed Hox gene Ph-Sex combs reduced (Scr). Moreover, Ubx misexpression caused overall transformation of head segments into thoracic segments. These two studies were the first demonstration that Hox genes in Parhyale were necessary and sufficient to establish regional identity in the embryo. Moreover, they provided the first functional evidence for the hypothesis that changes in Hox expression could be responsible for changes in crustacean body plan organization.

A comprehensive analysis of the Parhyale Hox genes was first provided by (Serano et al., 2016)(summarized in Fig. 6b). The authors initially identified the Hox genes through degenerate PCR and RACE from cDNA. Using these sequences to probe the Parhyale BAC library allowed for the isolation of genomic regions containing labial, proboscopedia, Hox3, Deformed, Sex combs reduced, Antennapedia, abdominal-A, and Abdominal-B, in addition to the Ubx gene previously described. The only ancestral arthropod Hox gene that appeared to be missing from Parhyale was fushi tarazu. Overlapping BACs from this original data set determined that lab/pb and Dfd/Scr/Antp/Ubx were clustered together. Recent work using fluorescent in situ hybridization to detect nascent Hox transcripts has shown that the majority of the Hox genes are found in a single cluster, with the exception of Hox3 (Kao et al., 2016). The 2018 Parhyale genome assembly places all of the Hox genes in a single collinear cluster that spans approximately 3.5Mb.

Expression analysis of each of the Hox genes was performed at several stages of development using in situ hybridization and immunofluorescence using cross-reactive antibodies (Serano et al., 2016). In summary, the Hox genes are expressed with temporal and spatial collinearity: four Hox genes are expressed in the head: lab, pb, Hox3, and Dfd; two Hox genes are expressed in both the head and the thorax: Scr and Antp; Ubx is expressed in the thorax; abd-A is expressed in the thorax and abdomen, and Abd-B is expressed in the abdomen alone. Anterior genes are expressed earlier in development (e.g. lab and Dfd are first expressed at S8, ~16hpf), while more posterior genes are expressed later, depending on their position along the Hox cluster (e.g. Antp is first expressed at S12, ~56hpf, while Abd-B is first expressed at S17-S18, around ~86 hpf). The expression patterns identified for Parhyale add to the catalog of crustacean Hox gene expression data from crustaceans with diverse body plans, including from the branchiopods Artemia and Daphnia, the maxillopod Sacculina, and the malacostracans Porcellio, Procambarus and Asellus, thus enabling further comparisons of gene expression between species (Abzhanov and Kaufman, 1999a, 1999b, 2000a, 2000b; Averof and Akam, 1995; Blin et al., 2003; Brena et al., 2005; Copf et al., 2003; Mouchel-Vielh et al., 2002; Papillon and Telford, 2007; Shiga et al., 2002; Vick and Blum, 2010).

To interrogate how Hox genes specify regional identity in Parhyale, Martin et al. used both CRISPR-Cas9 mutagenesis and RNAi knockdown (Martin et al., 2016), examining the posterior six of the nine Parhyale Hox genes: Dfd, Scr, Antp, Ubx, abd-A, and Abd-B (summarized in Fig. 7). For each CRISPR-Cas9 mutagenesis experiment, the authors independently tested two sgRNAs to control for off-target effects. In summary, knockout of each of the Hox genes produced various homeotic transformations that revealed their function in specifying regional identity (Fig. 7). For example, knockout of abd-A caused a transformation of T6-T8 reverse jumping legs to T4-T5 forward walking leg identity, as well as a transformation of A1-A3 swimming legs to A4-A6 anchoring leg identity. This suggests that abd-A is necessary for the establishment of swimming leg and anchoring leg identity, and that these leg identities are created as a result of overlapping expression of abd-A with either Ubx or Abd-B. Together, the homeotic transformations identified in these experiments describe the "Hox code" in Parhyale that dictates how combinations of Hox gene expression generate different limb types.

By examining the expression patterns of Hox genes in other crustaceans, then, one can begin to make predictions about how the Hox codes in these crustaceans function to produce diverse body plans. For example, using a cross-reactive antibody (FP6.87) targeting Ubx/abd-A proteins, Martin et al. were able to discern that the abd-A expression boundary is 2 segments more posterior in the decapod crustacean Procambarus fallax and the mysid Mysidium columbiae (Martin et al., 2016). This shift in the posterior abd-A expression boundary correlates with an increased number of swimming legs. Based on the CRISPR-Cas9 phenotype of abd-A in Parhyale, one can infer that the change in pleopod number is directly induced by a change in the expression boundary of abd-A.

The present achievements in Hox gene analysis in Parhyale illustrate its particular advantages as a research system. In particular, the ability to perform molecular genetic manipulations (including RNAi and CRISPR-Cas9) allows for thorough analysis of gene function. As a crustacean, its complex body plan posits many avenues of future inquiry, including how Hox proteins act in combination to generate diverse limb types, and identification of genes downstream of the Hox genes that direct limb

morphogenesis. Understanding Hox gene regulation will also help us understand further details of how evolutionary changes in limb morphology have been achieved.

Regeneration

Regeneration is a widespread phenomenon found across animals. However, there does not appear to be a unifying theme to all regenerative processes, as different clades demonstrate different regenerative mechanisms and potentials. In vertebrates, such as the zebrafish, Danio rerio, and the axolotl, Ambystoma mexicanum, lineage-restricted progenitors contribute to regenerated organs, whereas in planarians, a population of pluripotent stem cells is responsible for regeneration (Tanaka and Reddien, 2011). Unlike each of these species, humans lack any extensive ability to regenerate. What, then, are the molecular and genetic factors that humans have lost, and regenerative animals share?

The current research organisms available for regenerative research offer particular challenges. For example, while axolotls exhibit immense regenerative potential in multiple tissues, their large genome size and slow development make genetic research a challenge. Moreover, the regenerative strategies of planarians appear to differ substantially from those of the vertebrates that have been studied. To develop a broader picture of the ancestral regenerative state, one might wish to examine more diverse species within the evolutionary tree.

Parhyale is an arthropod capable of adult limb regeneration, and thus presents an excellent opportunity for examining regenerative mechanisms in a previously unexplored phylogenetic position. Early research into the mechanisms of limb regeneration in Parhyale has offered new insights into the ancestral urbilaterian regenerative state. Recently, (Konstantinides and Averof, 2014) used Parhyale to describe a common cellular basis for muscle regeneration between arthropods and vertebrates. In axolotls, limb regeneration appears to be lineage-restricted: dermis can make skin or cartilage, but not muscle, while muscle cells are derived from a population of Pax3/7+ muscle stem cells called “satellite cells” (Kragl et al., 2009). Konstantinides and Averof capitalized on the early germ layer restriction of Parhyale embryos to label different germ layer-specific populations of cells using transposon-mediated lineage labeling. For example, by injecting a GFP-expressing Minos transposable element and transposase mRNA into El or Er blastomere cells, the authors labeled thoracic ectoderm cells. By amputating lineage-labeled limbs, they showed that ectodermal lineages only give rise to ectodermal cells. Meanwhile, injection into ml and mr cells labeled mesodermal lineages, including muscle. After injury, regenerated mesodermal tissue was derived from the mesodermal lineage. The authors did not identify any blastomeres that contributed to both the mesodermal and ectodermal lineages. This is also consistent with previous evidence for germ layer restriction following embryonic blastomere ablations (Price et al., 2010). Moreover, regeneration was also restricted locally to the portion of the body that was lineage-labeled; for example, ml cells could not contribute to mesodermal regeneration on the right side of the animal. These results suggest that regeneration in Parhyale is lineage-restricted, and that Parhyale does not have a general pool of lineage-restricted progenitors in the body that contribute to regeneration.

Upon deeper examination, the authors noticed a small population of DsRed-positive mesoderm-lineage-labeled cells associated with muscle fibers, which resembled satellite cells. Using a

Pax3/7 antibody known to cross-react across many species (Davis et al., 2005), they identified these muscle-associated cells as a Pax-3/7+ population of satellite-like cells (SLC). EdU labeling revealed that these cells also proliferate in response to limb amputation. Moreover, transplantation of fluorescently lineage-labeled muscle cells into unlabeled limbs produced fluorescently labeled muscle fibers in regenerated limbs. Together, these data suggest that Parhyale SLCs have similar function to satellite cells found in vertebrates. This result reveals that a homologous cell type between vertebrates and arthropods participates in limb regeneration, suggesting more broadly that the common ancestor of arthropods and vertebrates had a muscle-regeneration program that employed Pax3/7-expressing SLCs.

A recent paper by Alwes et al. has leveraged the unique optical properties of Parhyale to live-image limb regeneration (Alwes et al., 2016). As Parhyale limbs are small and highly transparent, it is possible to observe limb regeneration in real time at the level of single cells. This optical transparency allows for live lineage tracking that is not possible in larger vertebrate regenerative systems. To visualize the regenerative process, the authors glued transgenic Parhyale expressing either fluorescent histones or lineage-restricted markers within a live-imaging chamber, securing the amputated limb in place. These animals largely survived the procedure, and it was then possible to image the limb during the course of regeneration, for up to 4-5 days at a time. The authors used fluorescent microscopy to examine the migration and division of individual cells, as well as the broader morphogenesis of the regenerating limb. By tracking individual cells, the authors were able to determine that all epidermal cells were equally able to participate in the regenerative process; there is no specialized cell type responsible for regeneration. Moreover, cells largely retained their original proximodistal positions along the limb as they proliferated, but changed in their positional identities upon limb differentiation and morphogenesis; for example, cells originally found in the middle of the limb, but at the edge of the amputation site, would ultimately give rise to the distal tip of the regenerated limb.

By taking advantage of the unique phylogenetic position, optical qualities, and developmental strategies of Parhyale, researchers have been able to develop a deeper understanding of properties common across regenerative species. This unique research system is thus poised to become a platform for novel discoveries about animal regeneration for years to come.

Conclusion: Future Directions

Since its introduction in the late 1990s, the amphipod crustacean Parhyale hawaiensis has provided opportunity for a wealth of comparative developmental investigations. Parhyale has many qualities that make it an excellent research organism, including its ease of care, easily accessible embryos, and mosaic development. With an increasingly sophisticated molecular genetic toolkit and continued proliferation in genomic and bioinformatics resources, Parhyale is accessible to any researcher who wishes to capitalize on its unique qualities as a research system.

Parhyale has already been used to investigate the evolution of early development, gene function, body plans, and regeneration. What is in store for the future of Parhyale? During a 2016 conference at the Howard Hughes Medical Research Institute's Janelia Farm Campus, organized by Anastasios Pavlopoulos and Michalis Averof, researchers interested in topics from lignocellulose digestion to tissue forces to circadian rhythms presented early findings and ideas for new projects in

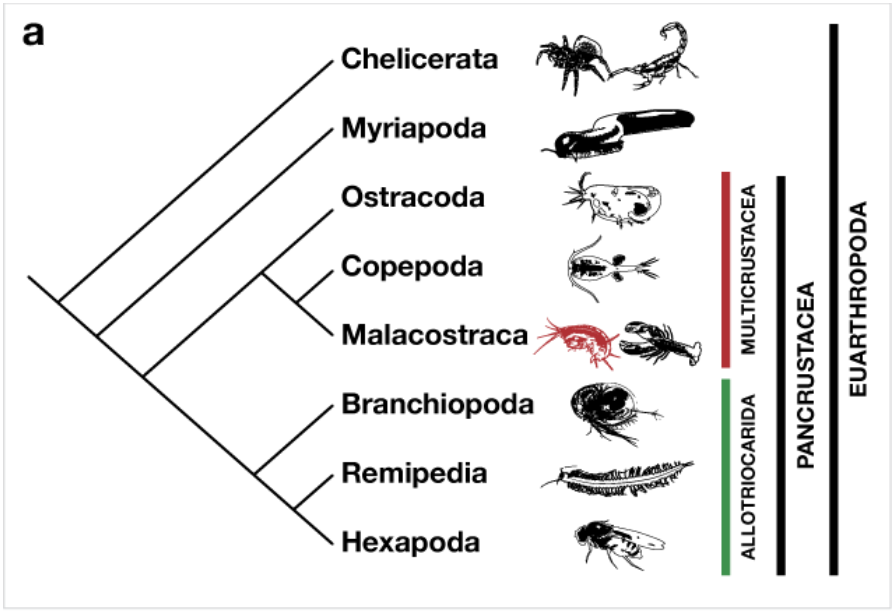
Parhyale. Others proposed novel transcriptomic datasets and functional genomic approaches to study Parhyale development. Moreover, the Parhyale community has begun to amalgamate protocols, bioinformatic tools, and resources to aid additional work in this system. In the last several years, a number of papers have been posted to bioRXiv and published in other journals that explore topics as diverse as the evolutionary origins of insect wings, developmental lineage tracing, and neuroanatomy. As more researchers continue to adopt Parhyale into their laboratories and examine unexplored aspects of its biology, it will be exciting to see what other research explorations will be enabled by this unique emerging research system.

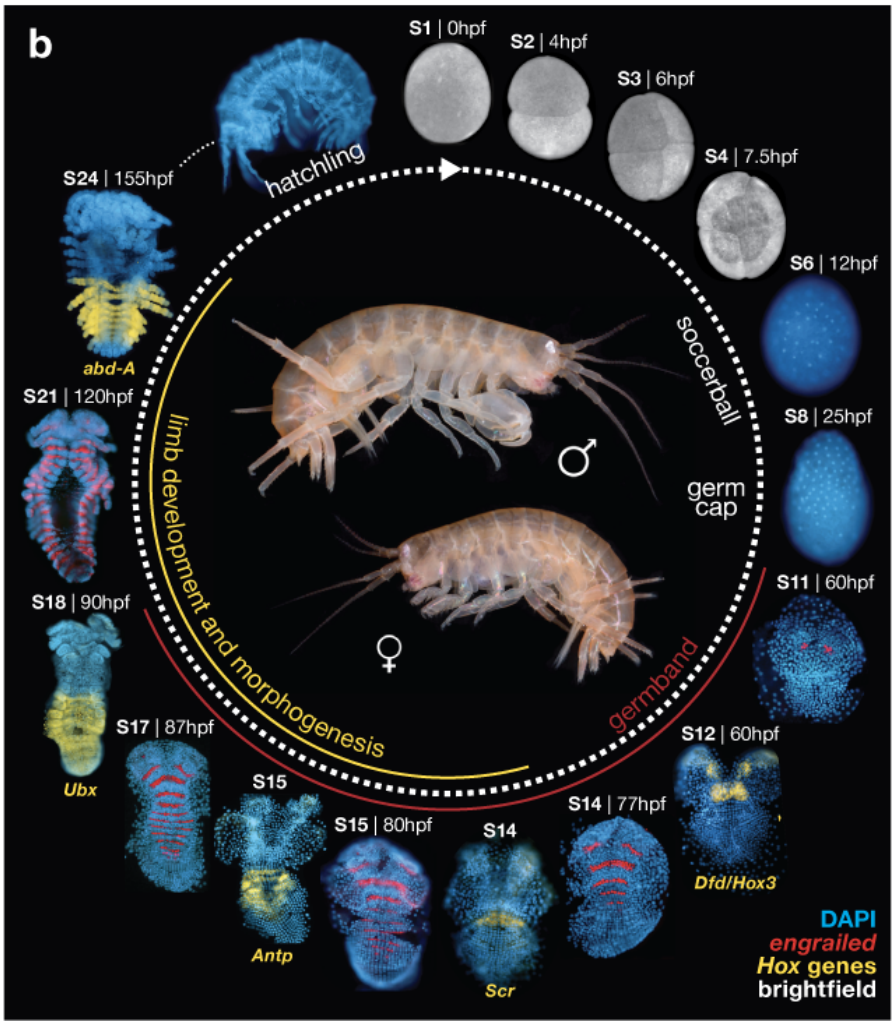
Figure 0.1: Phylogeny and development of *Parhyale hawaiensis*.

a) *Parhyale hawaiensis* is a malacostracan crustacean, and is within a clade with lobsters, crabs, and shrimp. Crustaceans are one of the four major groups of arthropods. The three remaining groups are chelicerates, including spiders and scorpions; myriapods, including millipedes and centipedes; and hexapods, including insects. (Schwentner et al., 2017).

b) Parhyale development takes place over ~10 days. The early embryo undergoes rounds of holoblastic cleavage (Stage 1-Stage 4) and cells eventually coalesce into a germ cap (Stage 8). Much of the embryo develops on the surface of the yolk during germband elongation. Parasegment formation can be visualized by the expression of engrailed (red). Over the course of embryogenesis, the Parhyale Hox genes (yellow) initiate expression over different regions of the body in a collinear anterior-posterior temporal sequence. The hatchling emerges as a miniature version of the adult. Sexually mature adults display sexual dimorphism: females are smaller and have visible ovaries, while males are larger and have larger chelipeds on T3.

hpf, hours post-fertilization; Dfd, Deformed; Scr, Sex combs reduced; Antp, Antennapedia; Ubx, Ultrabithorax; AbdA,





Abdominal-A.

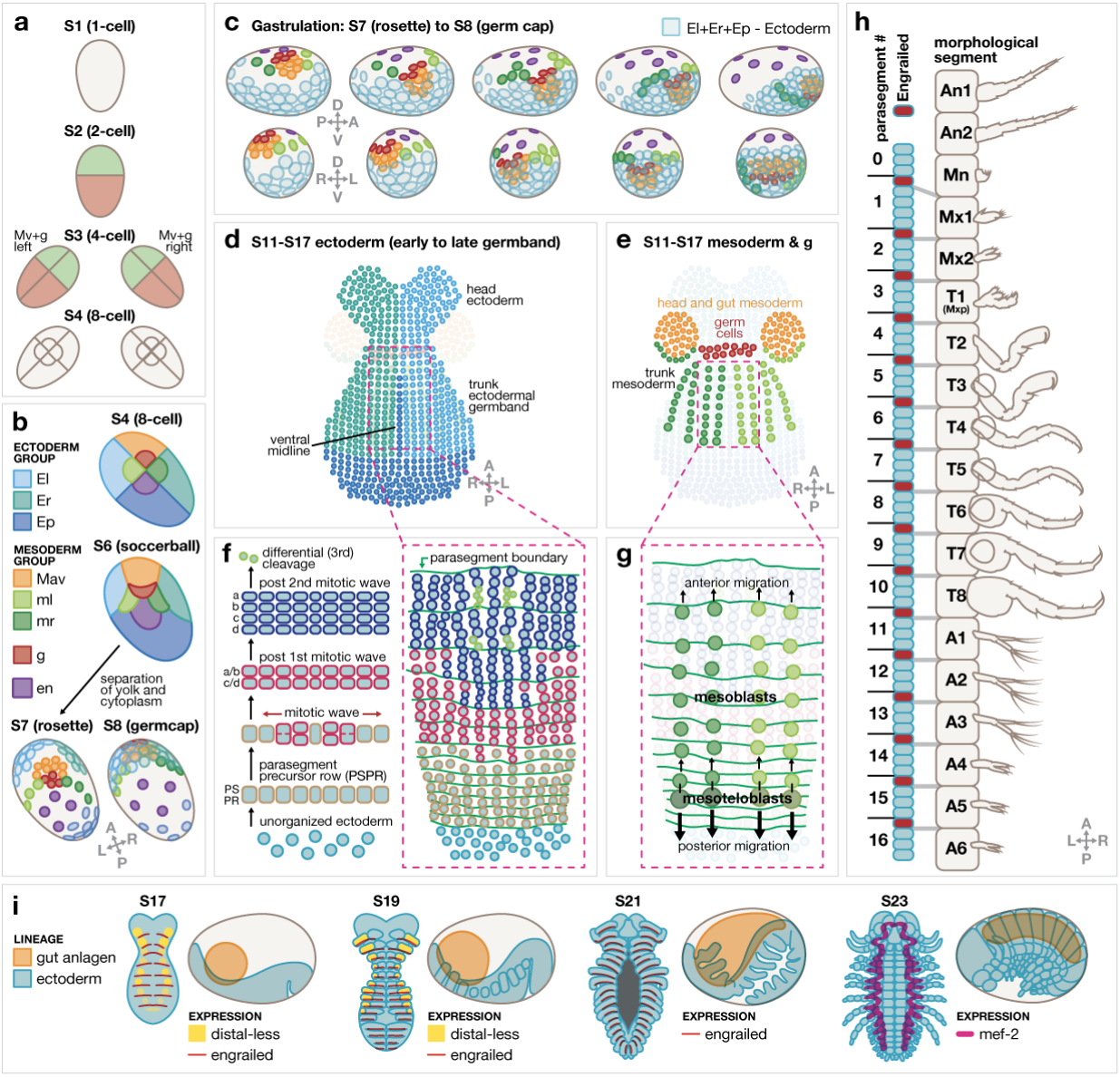


Figure 0.2: Parhyale cell lineage and development.

a) Following fertilization, the embryo undergoes a series of stereotyped holoblastic cleavages. At the 4-cell stage, the embryo can either undergo left- or right-cleavage, characterized by the position of the Mv+g cell.

b) Lineage tracing has revealed that descendants of cells of the eight-cell stage display germ layer restriction. The eight-cell embryo contains four macromeres (El, Er, Ep, Mav) and four micromeres (ml, mr, en, g). During the transition from S6 to S7, cells extrude their yolk to separate the cytoplasm from the yolk. The cells in the rosette of the S7 stage migrate beneath a layer of ectodermal cells in a process associated with gastrulation to generate the S8 germ cap stage.

c) Illustration of cell migrations during gastrulation.

d) Illustration of the contributions of El, Er, and Ep cells to the ectoderm at the germband stage. e) Illustration of the contributions of ml, mr, and Mav cells to the mesoderm and the position of the germ cells at the germband stage.

f) Left, schematic of the cell division cycles that give rise to the embryonic parasegments. Unorganized cells in the posterior of the embryo organize to form parasegment precursor rows (PSPRs), which then divide to once to generate a/b and c/d cells, then again to generate a, b, c, and d cells. These divisions occur as mediolateral waves of division, beginning

with the cells immediately next to the midline, and then spreading laterally. The midline cells divide shortly after the cells that are adjacent to them. Right, illustration of the organization of cells at the germband stage.

g) Migrations of the mesoteloblasts and mesoblasts during germband elongation. Illustration represents a subset of the mesoteloblasts. Mesoteloblasts migrate posteriorly and divide to deposit mesoblasts. Mesoblasts then migrate anteriorly to pair with their specified ectodermal parasegments.

h) Trunk ectodermal parasegments and their corresponding out-of-phase morphological segments. The “a” cell in each parasegment expresses engrailed, which marks the posterior of each morphological segment. i) Gene expression markers and morphological markers in Parhyale development. At S17-S19, engrailed marks the anterior of each parasegment, while distalless marks the developing limb buds. At S21, engrailed marks the parasegments, and at S23, mef-2 marks the developing mesoderm. The gut anlagen, which sequesters the yolk and develops into the gut, is a prominent feature of embryos, and is useful in classifying embryonic stages.

A, anterior; P, posterior; D, dorsal; V, ventral; L, left; R, right; An, antenna; Mn, mandible; Mx1, maxillule; Mx2, maxilla; Mxp, maxilliped.

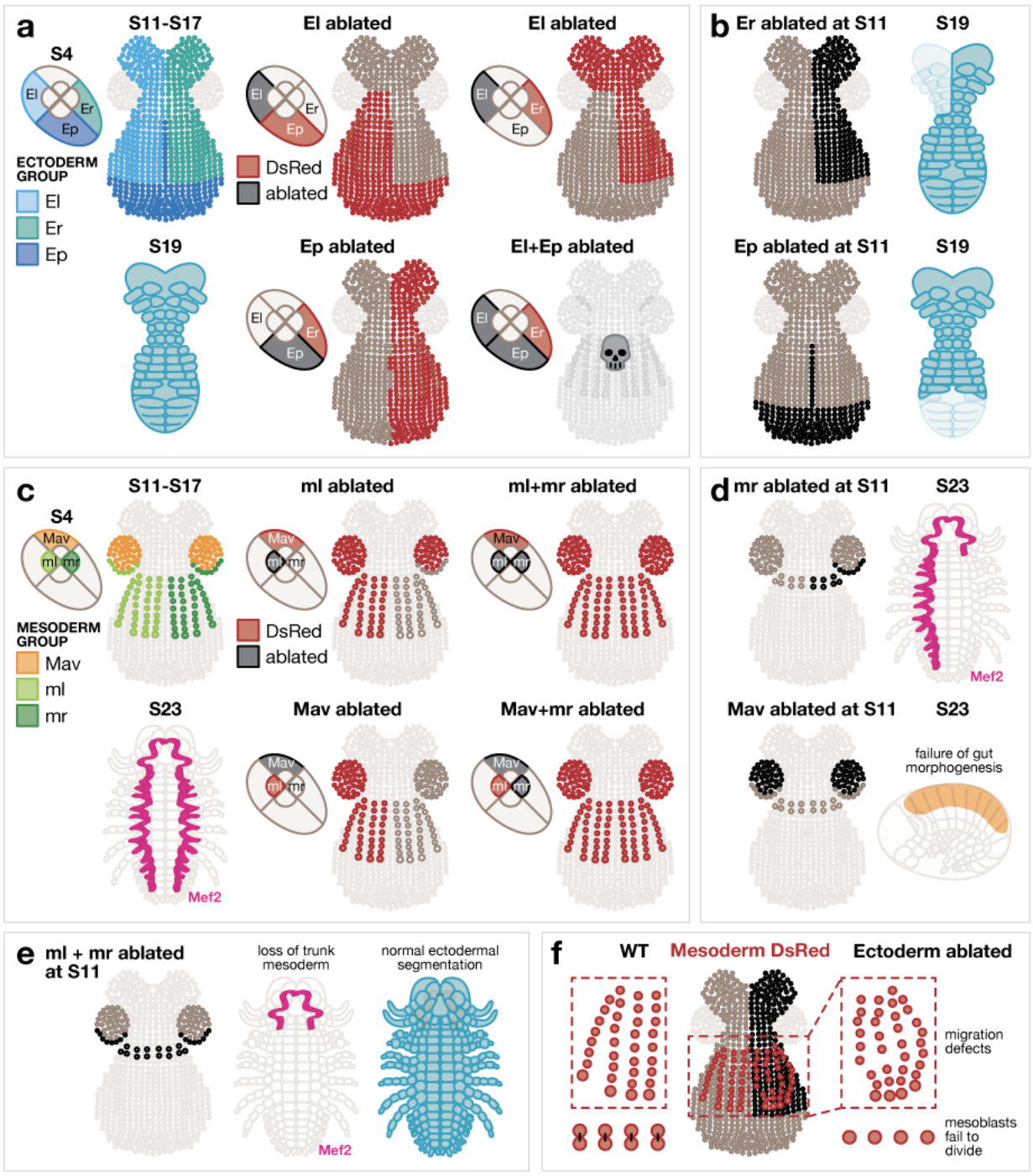


Figure 0.3: Germ layer compensation and germ layer interactions in Parhyale.

For ease of comprehension, in this figure, embryos are illustrated with the dorsal side facing towards the reader. During actual development, the dorsal side of the embryo faces the yolk, while the ventral side is superficial and clearly visible.

a) Left column: illustration of wildtype lineage patterns and morphology for ectoderm-lineage blastomeres. Middle and right columns: illustrations of ablation experiments, demonstrating intra-germ layer compensation in the ectoderm. Ablation of any two cells from the ectoderm lineage results in embryonic lethality.

b) Ablation of ectoderm lineage cells after the start of germband elongation. Germ layer compensation does not occur after the start of the germband stage.

c) Left column: illustration of wildtype lineage patterns and morphology for mesoderm-lineage blastomeres. Middle and right columns: illustrations of ablation experiments demonstrating intra-germ layer compensation in the mesoderm.
d) Ablation of mesoderm lineage cells after the start of germband elongation. Germ layer compensation does not occur after the start of the germband stage.
e) Ablation of the mesoderm after the start of germband elongation does not affect ectodermal segmentation.
f) Ablation of the ectoderm after the start of germband elongation leads to defects in mesoblast division and migration.
A, anterior; P, posterior; D, dorsal; V, ventral; L, left; R, right; An, antenna; Mn, mandible; Mx1, maxillule; Mx2, maxilla; Mxp, maxilliped.

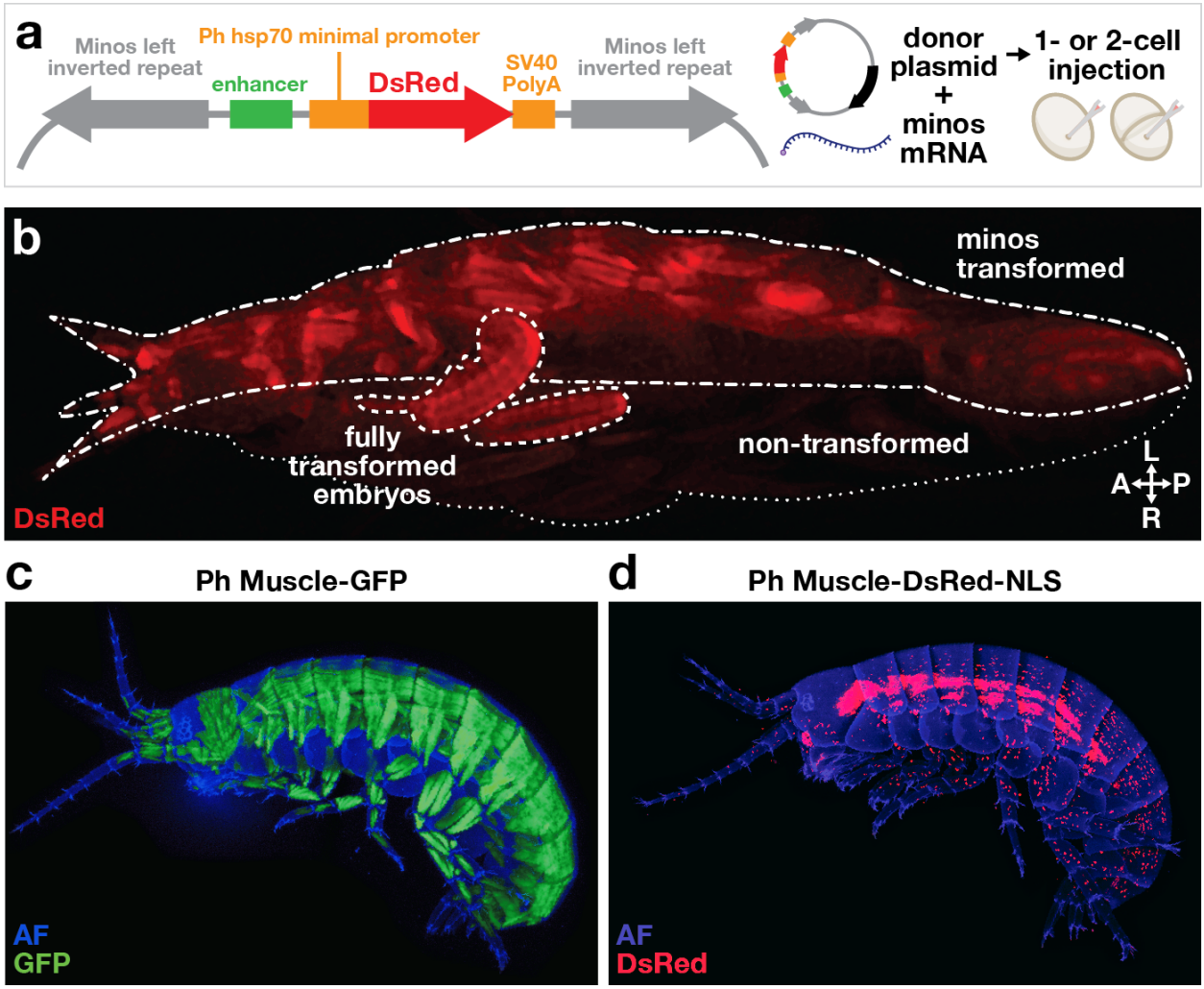


Figure 0.4: Transgenesis in Parhyale using the Minos transposase system.

a) A basic Minos transposase donor plasmid and a schematic of the transgenesis experiment.

b) Fluorescence image of a Minos-transformed female Parhyale that exhibits transformation in one-half of the body, with fully transgenic hatchlings in the brood pouch.

c) An example transgenic animal in which DsRed expression is specified by a Parhylale muscle enhancer.

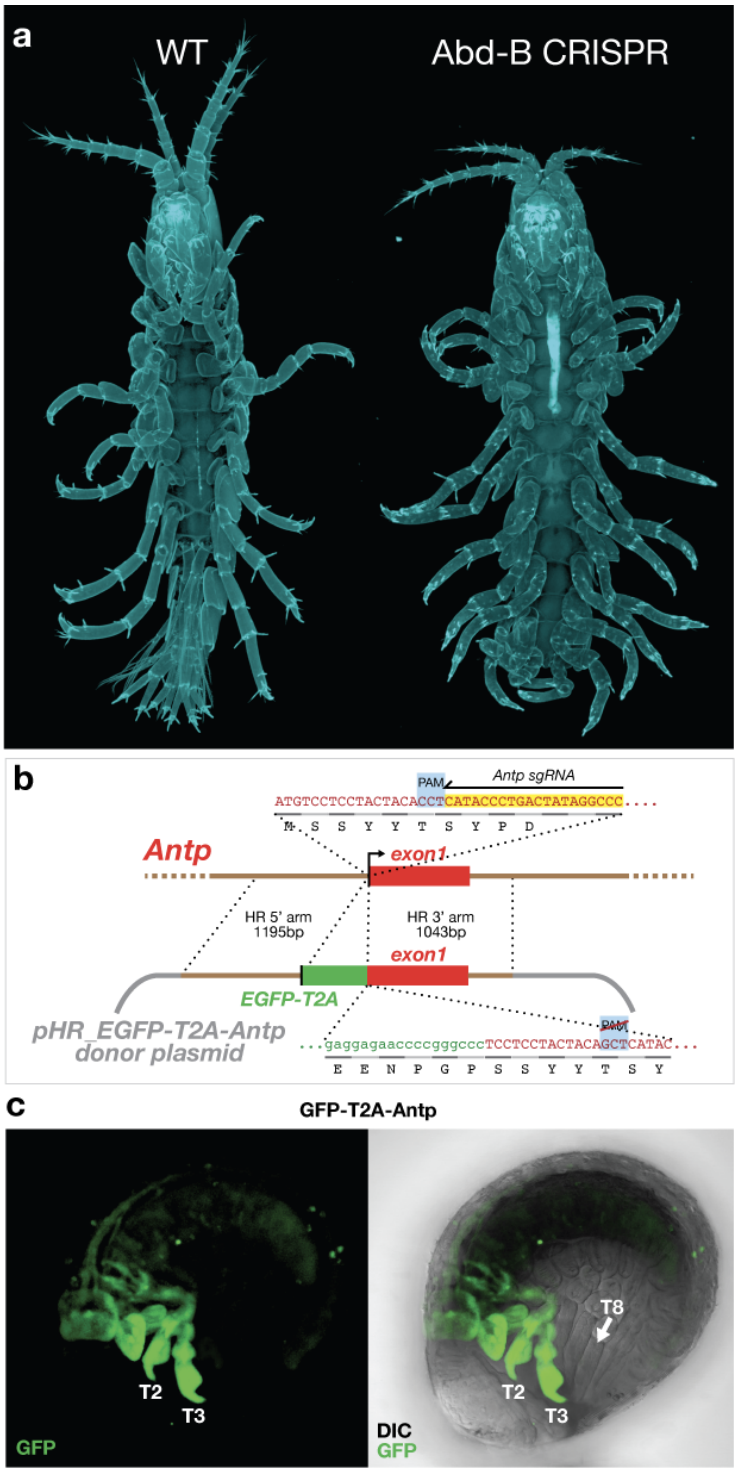
d) Same as in c, but using DsRed that contains a nuclear localization signal (NLS). The bright signal that runs the length of the body is autofluorescence.

Figure 0.5: CRISPR-Cas9 mutagenesis and transgenesis in *Parhyale*.

a) Bilateral CRISPR-Cas9 knockout of Abd-B. Left panel shows a wildtype Parhyale hatching. Right panel shows a CRISPR-Cas9 mutant for Abd-B. The abdominal appendages in the mutant are transformed to thoracic walking and jumping leg identities.

b) CRISPR-Cas9 homologous recombination plasmid and strategy to insert GFP into the Antp locus.

c) CRISPR-Cas9 homologous combination-mediated transgenesis. Transgenic embryo wherein GFP is expressed in segments that express Antp. Figure adapted from Martin et al 2016.



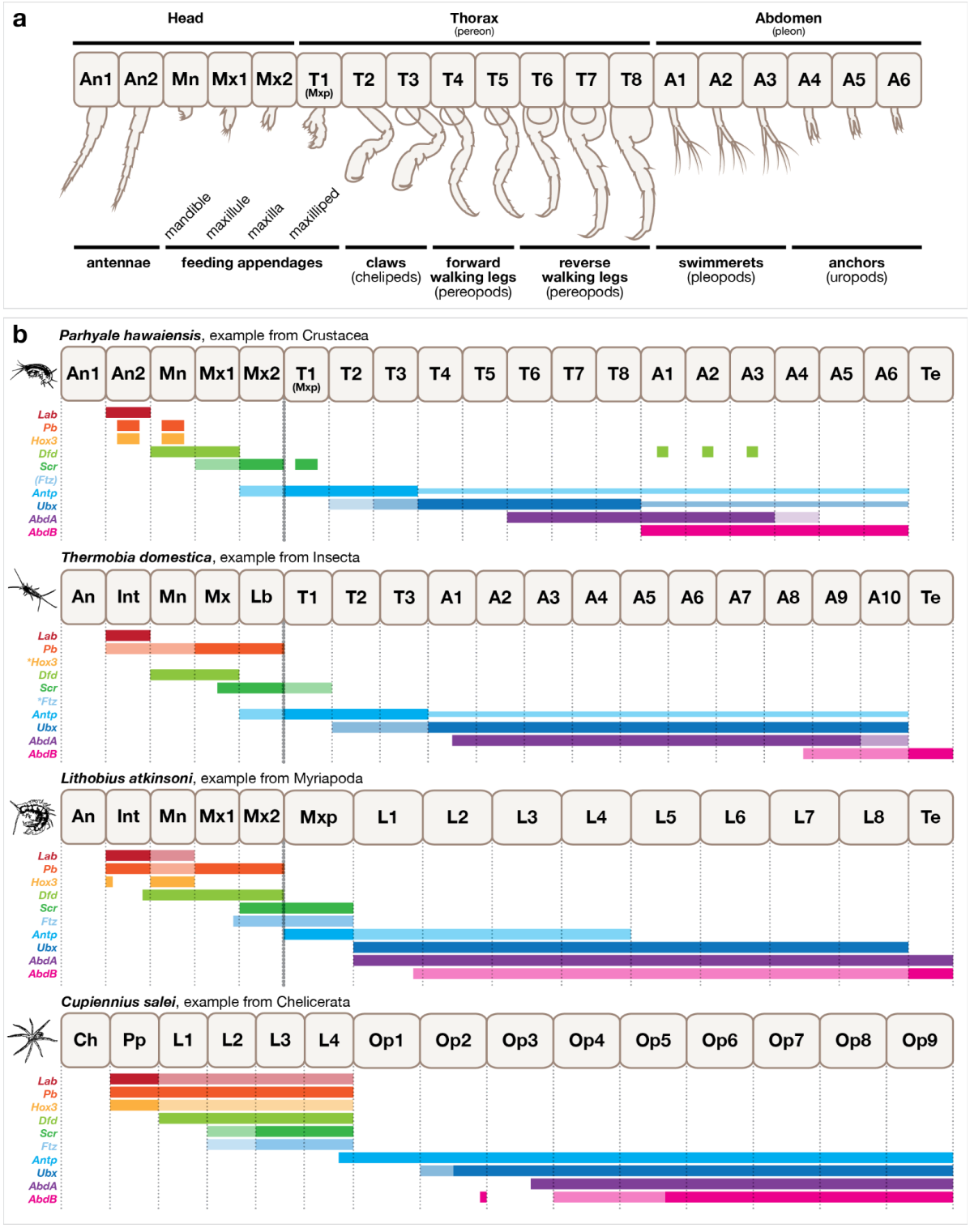


Figure 0.6: Body plan organization and Hox gene expression of Parhyale and other representative arthropods.

a) Body plan of Parhyale. b) Body plan and Hox gene expression of Parhyale and several representative arthropods. The expression of Hox genes correlates strongly with differing segment identities across the arthropods. Thermobia, Lithobius, and Cupiennius patterns adapted from Hughes and Kaufman, 2002.

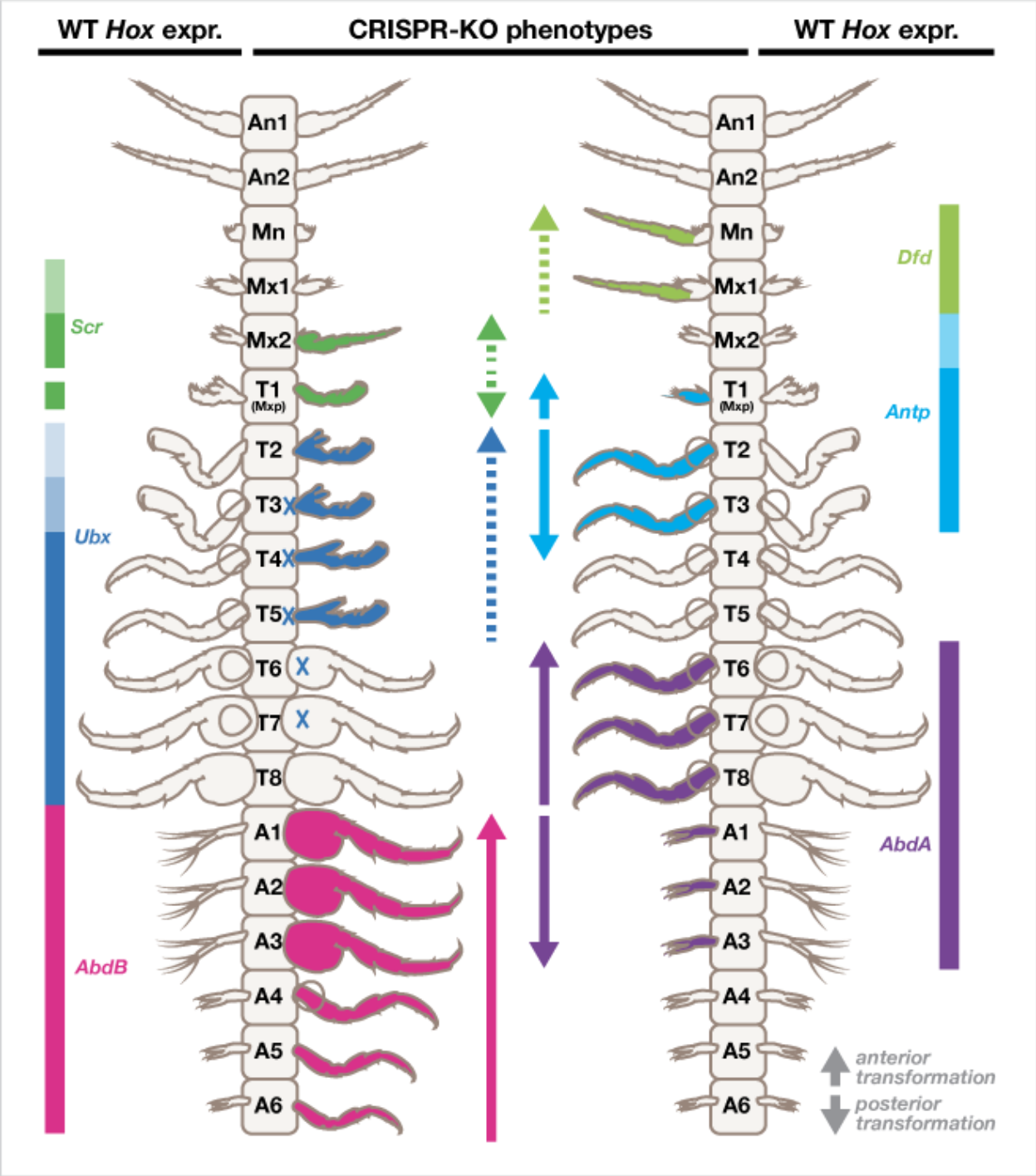


Figure 0.7: Homeotic transformations of segment identity in *Parhyale* through *Hox* gene knockout.

CRISPR-Cas9 knockout of Hox genes induces a range of different homeotic transformations, revealing the roles of these genes in specifying segment identity.

0.6 – References, Chapter 0

Abudayyeh, O.O., Gootenberg, J.S., Essletzbichler, P., Han, S., Joung, J., Belanto, J.J., Verdine, V., Cox, D.B., Kellner, M.J., Regev, A., et al. (2017). RNA targeting with CRISPR–Cas13. 280–284.

Abzhanov, A., and Kaufman, T.C. (1999a). Homeotic genes and the arthropod head: Expression patterns of the labial, proboscipedia, and Deformed genes in crustaceans and insects. Proc. Natl. Acad. Sci. 96, 10224.

Abzhanov, A., and Kaufman, T.C. (1999b). Novel regulation of the homeotic gene Scr associated with a crustacean leg-to-maxilliped appendage transformation. Development 126, 1121.

Abzhanov, A., and Kaufman, T.C. (2000). Crustacean (malacostracan) Hox genes and the evolution of the arthropod trunk. Development 127, 2239.

Alwes, F., Hinchen, B., and Extavour, C. (2011). Patterns of cell lineage, movement, and migration from germ layer specification to gastrulation in the amphipod crustacean Parhyale hawaiensis. Dev. Biol. 110–123.

Alwes, F., Enjolras, C., and Averof, M. (2016). Live imaging reveals the progenitors and cell dynamics of limb regeneration. eLife 5, e19766.

Ankeny, R.A. (2001). TIMELINE: The natural history of Caenorhabditis elegans research. Nat. Rev. Genet. 2, 474–479.

Averof, M., and Patel, N.H. (1997). Crustacean appendage evolution associated with changes in Hox gene expression. Nature 388, 682.

Bar-On, Y.M., Phillips, R., and Milo, R. (2018). The biomass distribution on Earth. Proc. Natl. Acad. Sci. 115, 6506.

Beeman, R.W. (1987). A homoeotic gene cluster in the red flour beetle. Nature 327, 247–249.

Blythe, M.J., Malla, S., Everall, R., Shih, Y.H., Lemay, V., Moreton, J., Wilson, R., and Aboobaker, A. (2012). High through-put sequencing of the Parhyale hawaiensis mRNAs and microRNAs to aid comparative developmental studies. PLoS ONE 7, e33784.

Brenner, S. (1974). The genetics of Caenorhabditis elegans. Genetics 77, 71.

Brown, S.J., Hilgenfeld, R.B., and Denell, R.E. (1994). The beetle Tribolium castaneum has a fushi tarazu homolog expressed in stripes during segmentation. Proc. Natl. Acad. Sci. 91, 12922.

Browne, W.E., Price, A.L., Gerberding, M., and Patel, N.H. (2005). Stages of embryonic development in the amphipod crustacean, Parhyale hawaiensis. Genesis 42, 124–149.

Browne, W.E., Schmid, B.G., Wimmer, E.A., and Martindale, M.Q. (2006). Expression of otd orthologs in the amphipod crustacean, Parhyale hawaiensis. Dev Genes Evol 216, 581–595.

Bucher, G., Scholten, J., and Klingler, M. (2002). Parental RNAi in Tribolium (Coleoptera). Curr. Biol. 12, R85–R86.

Chaw, R., and Patel, N.H. (2012). Independent migration of cell populations in the early gastrulation of the amphipod crustacean Parhyale hawaiensis. Dev Biol 371, 94–109.

Chen, F., M, P.-M., Shondra, and Davis, G.D. (2015). Gene editing using ssODNs with engineered endonucleases. Methods Mol Biol 1239, 251–265.

Choe, C.P., Miller, S.C., and Brown, S.J. (2006). A pair-rule gene circuit defines segments sequentially in the short-germ insect Tribolium castaneum. Proc. Natl. Acad. Sci. 103, 6560.

Christian, S.-E., Schultheis, D., Schwirz, J., Ströhlein, N., Troelenberg, N., Majumdar, U., Dao, V., Grossmann, D., Richter, T., Tech, M., et al. (2015). The iBeetle large-scale RNAi screen reveals gene functions for insect development and physiology. 7822.

Clarence R Shoemaker (1956). Observations on the Amphipod Genus Parhyale. Proc. U. S. Natl. Mus. 106, 345–358.

Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Joung, J., and Zhang, F. (2017). RNA editing with CRISPR-Cas13. Science 358, 1019.

Cuenot, L. (1905). Les races pures et leurs combinaisons chez les Souris (Notes et Revue). 3.

Dana, J. (1853). United States Exploring Expedition: 14: Crustacea.

Davis, G.K., and Patel, N.H. (2002). Short, long, and beyond: molecular and embryological approaches to insect segmentation. Annu Rev Entomol 47, 669–699.

Davis, G.K., D'Alessio, J.A., and Patel, N.H. (2005). Pax3/7 genes reveal conservation and divergence in the arthropod segmentation hierarchy. Dev. Biol. 285, 169–184.

Doudna, J.A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science 346, 1258096.

El-Sherif, E., Averof, M., and Brown, S.J. (2012). A segmentation clock operating in blastoderm and germband stages of Tribolium development. Development 139, 4341.

Extavour, C.G. (2005). The fate of isolated blastomeres with respect to germ cell formation in the amphipod crustacean Parhyale hawaiensis. Dev Biol 277, 387–402.

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature 551, 464.

Gerberding, M., Browne, W., and Patel, N. (2002). Cell lineage analysis of the amphipod crustacean Parhyale hawaiensis reveals an early restriction of cell fates. Development 129, 5789–5801.

Groth, A.C., Fish, M., Nusse, R., and Calos, M.P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics 166, 1775–1782.

Gupta, T., and Extavour, C.G. (2013). Identification of a putative germ plasm in the amphipod Parhyale hawaiensis. EvoDevo 4, 34.

Hannibal, R.L., Price, A.L., and Patel, N.H. (2012). The functional relationship between ectodermal and mesodermal segmentation in the crustacean, Parhyale hawaiensis. Dev Biol 361, 427–438.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell 157, 1262–1278.

Hughes, C.L., and Kaufman, T.C. (2002). Hox genes and the evolution of the arthropod body plan. Evol. Dev. 4, 459–499.

J Laurens Barnard (1965). Marine Amphipoda of atolls in Micronesia. Proc. U. S. Natl. Mus. 117, 459–551.

Janssen, R., Eriksson, B.J., Tait, N.N., and Budd, G.E. (2014). Onychophoran Hox genes and the evolution of arthropod Hox gene expression. Front. Zool. 11, 22.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821.

Kao, D., Lai, A.G., Stamataki, E., Rosic, S., Konstantinides, N., Jarvis, E., Donfrancesco, A., Natalia, P.-S., Semon, M., Grillo, M., et al. (2016). The genome of the crustacean Parhyale hawaiensis, a model for animal development, regeneration, immunity and lignocellulose digestion. Elife 5, 65789.

Knorr, E., Bingsohn, L., Kanost, M.R., and Vilcinskas, A. (2013). Tribolium castaneum as a Model for High-Throughput RNAi Screening. In Yellow Biotechnology II: Insect Biotechnology in Plant Protection and Industry, A. Vilcinskas, ed. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 163–178.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420.

Konstantinides, N., and Averof, M. (2014). A common cellular basis for muscle regeneration in arthropods and vertebrates. Science 343, 788–791.

Kontarakis, Z., Pavlopoulos, A., Kiupakis, A., Konstantinides, N., Douris, V., and Averof, M. (2011). A versatile strategy for gene trapping and trap conversion in emerging model organisms. Development 138, 2625–2630.

Kragl, M., Knapp, D., Nacu, E., Khattak, S., Maden, M., Epperlein, H.H., and Tanaka, E.M. (2009). Cells keep a memory of their tissue origin during axolotl limb regeneration. Nature 460, 60–65.

Lewis, E. (1978). A gene complex controlling segmentation in Drosophila. Nature 276, 565–570.

Liubicich, D.M., Serano, J.M., Pavlopoulos, A., Kontarakis, Z., Protas, M.E., Kwan, E., Chatterjee, S., Tran, K.D., Averof, M., and Patel, N.H. (2009). Knockdown of Parhyale Ultrabithorax recapitulates evolutionary changes in crustacean appendage morphology. Proc Natl Acad Sci USA 106, 13892–13896.

Lynch, J.A., El-Sherif, E., and Brown, S.J. (2012). Comparisons of the embryonic development of Drosophila, Nasonia, and Tribolium. Wiley Interdiscip. Rev. Dev. Biol. 1, 16–39.

Martin, A., Serano, J.M., Jarvis, E., Bruce, H.S., Wang, J., Ray, S., Barker, C.A., C, O., Liam, and Patel, N.H. (2015). CRISPR/Cas9 Mutagenesis Reveals Versatile Roles of Hox Genes in Crustacean Limb Specification and Evolution. Curr Biol 26, 14–26.

Nestorov, P., Battke, F., Levesque, M.P., and Gerberding, M. (2013). The Maternal Transcriptome of the Crustacean Parhyale hawaiensis Is Inherited Asymmetrically to Invariant Cell Lineages of the Ectoderm and Mesoderm. PLOS ONE 8, e56049.

Nüsslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. Nature 287, 795–801.

Oakley, T.H., Wolfe, J.M., Lindgren, A.R., and Zaharoff, A.K. (2013). Phylotranscriptomics to Bring the Understudied into the Fold: Monophyletic Ostracoda, Fossil Placement, and Pancrustacean Phylogeny. Mol Biol Evol 30, 215–233.

Özhan-Kizil, G., Havemann, J., and Gerberding, M. (2009). Germ cells in the crustacean Parhyale hawaiensis depend on Vasa protein for their maintenance but not for their formation. Dev Biol 327, 230–239.

Pace, R.M., Grbić, M., and Nagy, L.M. (2016). Composition and genomic organization of arthropod Hox clusters. EvoDevo 7, 11.

Paese, C.L.B., Schoenauer, A., Leite, D.J., Russell, S., and McGregor, A.P. (2018). A SoxB gene acts as an anterior gap gene and regulates posterior segment addition in a spider. eLife 7, e37567.

Parchem, R.J., Poulin, F., Stuart, A.B., Amemiya, C.T., and Patel, N.H. (2010). BAC library for the amphipod crustacean, Parhyale hawaiensis. Genomics 95, 261–267.

Parra, G., Korf, I., and Bradnam, K. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23, 1061–1067.

Pavlopoulos, A., and Averof, M. (2005). Establishing genetic transformation for comparative developmental studies in the crustacean Parhyale hawaiensis. Proc Natl Acad Sci USA 102, 7888–7893.

Pavlopoulos, A., Kontarakis, Z., Liubicich, D.M., Serano, J.M., Akam, M., Patel, N.H., and Averof, M. (2009). Probing the evolution of appendage specialization by Hox gene misexpression in an emerging model crustacean. Proc Natl Acad Sci USA 106, 13897–13902.

Poovachiranon, S., Boto, K., and Duke, N. (1986). Food preference studies and ingestion rate measurements of the mangrove amphipod Parhyale hawaiensis (Dana). 98, 129–140.

Price, A.L., and Patel, N.H. (2008). Investigating divergent mechanisms of mesoderm development in arthropods: the expression of Ph-twist and Ph-mef2 in Parhyale hawaiensis. J Exp Zool. Part B Mol Dev Evol 310B, 24–40.

Price, A.L., Modrell, M.S., Hannibal, R.L., and Patel, N.H. (2010). Mesoderm and ectoderm lineages in the crustacean Parhyale hawaiensis display intra-germ layer compensation. Dev Biol 341, 256–266.

Prpic, N.-M.M., and Telford, M.J. (2008). Expression of homothorax and extradenticle mRNA in the legs of the crustacean Parhyale hawaiensis: evidence for a reversal of gene expression regulation in the pancrustacean lineage. Dev Genes Evol 218, 333–339.

Quadros, R.M., Ohtsuka, M., Harms, D.W., Aida, T., Redder, R., Miura, H., Richardson, G.P., Behlke, M.A., Zeiner, S.A., Jacobi, A.M., et al. (2016). Easi-CRISPR: Efficient germline modification with long ssDNA donors. Biorxiv 69963.

Rehm, E.J., Hannibal, R.L., Chaw, R.C., Vargas-Vila, M.A., and Patel, N.H. (2009a). Injection of Parhyale hawaiensis Blastomeres with Fluorescently Labeled Tracers. Cold Spring Harb. Protoc. 2009, pdb.prot5128.

Rehm, E.J., Hannibal, R.L., Chaw, R.C., Vargas-Vila, M.A., and Patel, N.H. (2009b). Fixation and Dissection of Parhyale hawaiensis Embryos. Cold Spring Harb. Protoc. 2009, pdb.prot5127.

Rehm, E.J., Hannibal, R.L., Chaw, R.C., Vargas-Vila, M.A., and Patel, N.H. (2009c). Antibody Staining of Parhyale hawaiensis Embryos. Cold Spring Harb. Protoc. 2009, pdb.prot5129.

Rehm, E.J., Hannibal, R.L., Chaw, R.C., Vargas-Vila, M.A., and Patel, N.H. (2009d). In Situ Hybridization of Labeled RNA Probes to Fixed Parhyale hawaiensis Embryos. Cold Spring Harb. Protoc. 2009, pdb.prot5130.

Sarrazin, A.F., Peel, A.D., and Averof, M. (2012). A Segmentation Clock with Two-Segment Periodicity in Insects. Science 336, 338.

Schwager, E.E., Schoppmeier, M., Pechmann, M., and Damen, W.G. (2007). Duplicated Hox genes in the spider Cupiennius salei. Front. Zool. 4, 10.

Schwager, E.E., Pechmann, M., Feitosa, N.M., P, M., Alistair, and Damen, W.G. (2009). hunchback Functions as a Segmentation Gene in the Spider Achaearanea tepidariorum. 1333–1340.

Schwentner, M., Combosch, D.J., Pakes Nelson, J., and Giribet, G. (2017). A Phylogenomic Solution to the Origin of Insects by Resolving Crustacean-Hexapod Relationships. Curr. Biol. 27, 1818–1824.e5.

Serano, J.M., Martin, A., Liubicich, D.M., Jarvis, E., Bruce, H.S., La, K., Browne, W.E., Grimwood, J., and Patel, N.H. (2016). Comprehensive analysis of Hox gene expression in the amphipod crustacean Parhyale hawaiensis. Dev Biol 409, 297–309.

Streisinger, G., Walker, C., Dower, N., Knauber, D., and Singer, F. (1981). Production of clones of homozygous diploid zebra fish (Brachydanio rerio). Nature 291, 293–296.

Sturtevant, A. (1965). The "Fly Room." Am. Sci.

Tanaka, E.M., and Reddien, P.W. (2011). The cellular basis for animal regeneration. Dev Cell 21, 172–185.

Tomoyasu, Y., Wheeler, S.R., and Denell, R.E. (2005). Ultrabithorax is required for membranous wing identity in the beetle Tribolium castaneum. Nature 433, 643–647.

VanHook, A.M., and Patel, N.H. (2008). Crustaceans. Curr Biol 18, R547–R550.

Vargas-Vila, M.A., Hannibal, R.L., Parchem, R.J., Liu, P.Z., and Patel, N.H. (2010). A prominent requirement for single-minded and the ventral midline in patterning the dorsoventral axis of the crustacean Parhyale hawaiensis. Development 137, 3469–3476.

William, M., and Krumlauf, R. (1992). Homeobox genes and axial patterning. Cell 68, 283–302.

Wolff, C., and Scholtz, G. (2002). Cell lineage, axis formation, and the origin of germ layers in the amphipod crustacean Orchestia cavimana. Dev Biol 250, 44–58.

Zeng, V., and Extavour, C.G. (2012). ASGARD: an open-access database of annotated transcriptomes for emerging model arthropod species. Database Oxf. 2012, bas048.

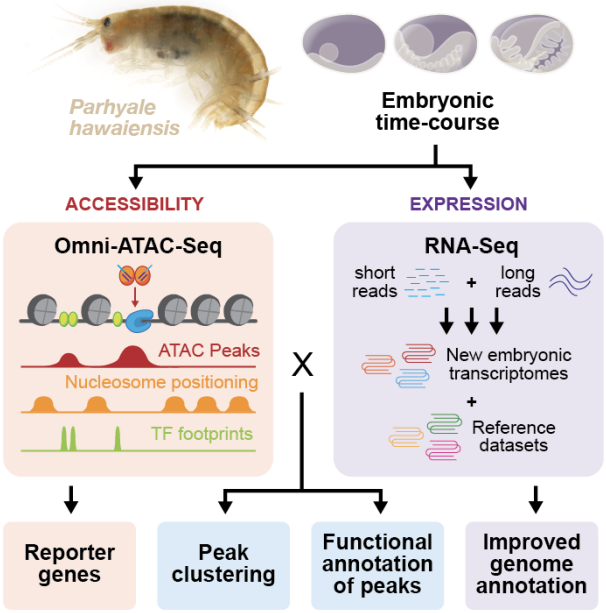
Zeng, V., Villanueva, K.E., S, E.-C., Ben, Alwes, F., Browne, W.E., and Extavour, C.G. (2011). De novo assembly and characterization of a maternal and developmental transcriptome for the emerging model crustacean Parhyale hawaiensis. BMC Genomics 12, 581.

1 – Identification and classification of cis-regulatory elements in the amphipod crustacean *Parhyale hawaiensis*

Dennis A Sun[1*], Jessen V Bredeson[1], Heather S Bruce[2], and Nipam H Patel[2,3*]
1. Department of Molecular and Cell Biology, University of California, Berkeley
2. Marine Biological Laboratory, Woods Hole, MA
3. Department of Organismal Biology and Anatomy, University of Chicago, Chicago, USA
*corresponding authors

Abstract

Emerging research organisms enable the study of biology that cannot be addressed using classical "model" organisms. The development of new data resources can accelerate research in such animals. Here, we present new functional genomic resources for the amphipod crustacean Parhyale hawaiensis, facilitating the exploration of gene regulatory evolution using this emerging research organism. We use Omni-ATAC-Seq, an improved form of the Assay for Transposase-Accessible Chromatin coupled with next-generation sequencing (ATAC-Seq), to identify accessible chromatin genome-wide across a broad time course of Parhyale embryonic development. This time course encompasses many major morphological events, including segmentation, body regionalization, gut morphogenesis, and limb development. In addition, we use short- and long-read RNA-Seq to generate an improved Parhyale genome annotation, enabling deeper classification of identified regulatory elements. We discover differential accessibility, predict nucleosome positioning, infer transcription factor binding, cluster peaks based on accessibility dynamics, classify biological functions, and correlate gene expression with accessibility. Using a Minos transposase reporter system, we demonstrate the potential to identify novel regulatory elements using this approach, including distal regulatory elements. This work provides a platform for the identification of novel developmental regulatory elements in Parhyale, and offers a framework for performing such experiments in other emerging research organisms.



Primary Findings

- Omni-ATAC-Seq identifies cis-regulatory elements genome-wide during crustacean embryogenesis
- Combined short- and long-read RNA-Seq improves the *Parhyale* genome annotation
- ImpulseDE2 analysis identifies dynamically regulated candidate regulatory elements
- NucleoATAC and HINT-ATAC enable inference of nucleosome occupancy and transcription factor binding
- Fuzzy clustering reveals peaks with distinct accessibility and chromatin dynamics
- Integration of accessibility and gene expression reveals possible enhancers and repressors
- Omni-ATAC can identify known and novel regulatory elements

1.0 – Introduction

Advances in genomic techniques have facilitated genetic research in emerging research organisms. As sequencing tools become more sophisticated and less expensive over time, understanding what genes an organism expresses at any time in development becomes more straightforward. The expansion of genome sequencing has accelerated to such a degree that cost per gigabase of genome sequence has decreased at a rate faster than predicted by Moore's Law (Wetterstramd), enabling the inexpensive generation of gene sequence. Moreover, the establishment of techniques, including RNAi and CRISPR-Cas9 mutagenesis, has made genetic ablation available to researchers working in a range of different research organisms, enabling the rapid characterization of gene function. RNAi knockdown has been applied to various insects (Christian et al., 2015; Mito et al., 2011), chelicerates (Sharma et al., 2013), flatworms (Rouhana et al., 2013; Srivastava et al., 2014), and numerous other organisms to study gene function. More recently, CRISPR-Cas9 mutagenesis has enabled both targeted ablation and targeted transgenesis in organisms as diverse as cephalopods (Crawford et al., 2020) and reptiles (Rasys et al., 2019). Using these tools, it is now possible to identify and perturb the function of genes across diverse organisms to understand how genes and genomes evolve across the tree of life.

While changes to gene function have been shown to play a significant role in evolutionary processes, recent studies have revealed how changes to gene regulation can also play a meaningful role in shaping the diversity of cell types, morphologies, and populations across clades (Mattioli et al., 2020; Signor and Nuzhdin, 2018; Wittkopp et al., 2004). The evolution of gene regulation is governed by changes in the function of two gene-regulatory modes: changes to trans-regulatory factors and changes to cis-regulatory elements. Trans-regulatory factors may include transcription factors, noncoding RNAs, and other diffusible elements that modify the expression of distant genes, while cis-regulatory elements are considered regions of the genome that locally mediate gene regulation.

The identification and characterization of possible trans-regulatory factors in a genome can be accomplished with relative ease using genome sequencing. For example, functional annotation of transcriptomes is able to easily identify transcription factors based on highly conserved protein domains (Ofran et al., 2007; Wang and Brown, 2006). Many bioinformatic tools also exist to identify microRNAs and long noncoding RNAs through sequence-based approaches (Duan et al., 2020; Liu et al., 2014). RNA-Seq approaches can also enable the identification of differentially expressed transcripts

which might be important for biological function (Costa-Silva et al., 2017). Thus, it is possible to identify many of the trans-regulatory factors within an organism's genome with straightforward RNA sequencing approaches.

Cis-regulatory elements (CREs), on the other hand, have historically proven difficult to identify even in model organisms. Unlike protein-coding sequences, cis-regulatory elements are composed of complex assemblies of transcription factor binding sites, which can be computationally challenging to identify (Li et al., 2015; Wasserman and Sandelin, 2004). These sites can change dramatically over short periods of time; for example, in the context of the *even-skipped* stripe 2 enhancer across several *Drosophila* species, the position of critical transcription factor binding sites shifted in as short a span as 1–2 million years divergence (Ludwig et al., 1998). Such rapid evolutionary flexibility, while important for driving short- and long-term evolutionary processes, makes the identification of such elements in novel research organisms a considerable challenge.

In recent years, novel techniques for identifying CREs have been developed, which rely on the fact that CREs often occur in regions of accessible chromatin. The earliest, including DNAse I hypersensitivity and Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE-Seq) enabled researchers to locate accessible chromatin regions and discover novel regulatory elements in a range of research organisms (Giresi et al., 2007; Lai et al., 2018; Pérez-Zamorano et al., 2017). However, such techniques were hampered by low signal-to-noise ratios, the requirement for very large amounts of tissue, and technically challenging protocols. In 2013, Buenrostro et al. published papers describing the Assay for Transposase-Accessible Chromatin with next-generation sequencing (ATAC-Seq), a novel technique using a hyperactive version of the Tn5 transposase, as a way of identifying accessible chromatin regions genome-wide (Buenrostro et al., 2015). In this paper, they demonstrated that ATAC-Seq performed comparably to previous accessibility techniques with reduced tissue requirements and increased convenience and speed.

ATAC-Seq has proved to be an effective technique for traditional model systems, and has offered a true revolution for emerging model systems. The low material input requirements of ATAC-Seq are of particular use in emerging research organisms, for which generating the millions of cells required for previous techniques would prove challenging. Researchers have applied ATAC-Seq to gar embryos in order to study the evolution of limb development (JL and Shubin, 2015) and regenerating acoel flatworms and hydra to study regeneration (Cazet et al., 2021; Gehrke et al., 2019), among numerous other examples. Thus, ATAC-Seq is now established as a generalizable tool that is particularly well-suited to accelerating work in emerging research organisms.

In 2017, Corces et al. reported a further-improved version of ATAC-Seq, Omni-ATAC-Seq, which provides several improvements to the standard ATAC-Seq protocol (Corces et al., 2017). In particular, it has been reported to have a greater signal-to-noise ratio, decreased mitochondrial read contamination, and applicability to fixed and frozen tissues, with only minor modifications to the buffers and detergents used in the standard ATAC-Seq protocol.

In this paper, we apply the Omni-ATAC-Seq protocol to embryos of the amphipod crustacean *Parhyale hawaiensis*, an emerging research organism for the study of arthropod development, evolution, and regeneration (Paris et al., 2021; Stamataki and Pavlopoulos, 2016; Sun and Patel, 2019). We identify dynamic regions of open chromatin ("peaks") across a broad swath of *Parhyale*

developmental time. We comprehensively analyze our Omni-ATAC-Seq data to predict the position of nucleosomes along the genome and infer the footprints of transcription factors bound to peaks. Using fuzzy clustering (Kumar and E Futschik, 2007), we partition our peaks into groups based on similar accessibility trajectories, revealing groups of peaks with different transcription factor footprint enrichment and nucleosome occupancy that are active at different points in development. In addition, we use short- and long-read RNA-Seq to improve the *Parhyale* genome annotation and investigate the relationship between accessibility and gene expression over time during development.

Parhyale has served as a platform for foundational discoveries about such processes as body plan evolution (Martin et al., 2015), the evolution of arthropod limbs (Bruce and Patel, 2020), and the evolution of regeneration (Konstantinides and Averof, 2014). By facilitating the identification of regulatory elements in this research organism, our work will enable other researchers to investigate the complexities of gene regulatory evolution in these processes and others. Furthermore, by enabling the assessment of cis-regulatory elements in *Parhyale*, we open avenues to investigate fundamental mechanisms of gene regulation in the understudied non-insect crustacean clade.

While this work primarily focuses on *Parhyale*, our methods can provide examples of how one can perform thorough analyses of ATAC-Seq and RNA-Seq data using existing tools to generate hypotheses about gene expression dynamics and regulatory element function. Such approaches can be applied to a diverse range of organisms, and will facilitate deeper understanding of gene regulation across the tree of life.

1.1 – Results

1.1.1 – Omni-ATAC-Seq identifies open chromatin across *Parhyale* developmental stages

To identify developmental regulatory elements, we performed Omni-ATAC-Seq on 15 stages of *Parhyale* embryonic development (Fig. 1). For each developmental stage, we generated two Omni-ATAC-Seq libraries using five embryos each, and fixed and DAPI stained sibling embryos to confirm staging (Fig. 1A). The developmental stages catalogued in our dataset are illustrated on a timeline in Fig. 1B, and representative fixed, DAPI-stained sibling embryos are illustrated in Fig. 1C. Additional brightfield images of embryos prior to tagmentation can be found in Supp. Fig. 1A.

To assess the quality of our libraries, we employed a battery of standard tests used to assess the quality of ATAC-Seq data (Supp Fig. 2A). Omni-ATAC-Seq libraries were also compared to previous ATAC-Seq libraries (see Methods), and had greater library size, higher numbers of non-duplicated reads, and lower amounts of mitochondrial DNA contamination (Supp. Fig. 2B, C). Fragment size analysis of Omni-ATAC-Seq reads after mapping revealed distinct 1-nucleosome and 2-nucleosome peaks, indicating that tagmentation was efficient (Supp Fig. 2E, F).

To determine whether our Omni-ATAC-Seq data represented genuine enrichment in accessible chromatin regions, we examined promoter accessibility genome-wide. A hallmark of successful ATAC-Seq experiments is strong enrichment of reads in gene promoters (Yan et al., 2020). We evaluated our ATAC-Seq using the most recent phaw_5.0 gene annotations (obtained from Leo Blondel, hereafter referred to as the "MAKER annotation") at the start of annotated mRNAs and across mRNA lengths, and compared our results to RNA-Seq read pileups (Fig. 2A–B). As expected, Omni-ATAC-Seq signal is enriched symmetrically at mRNA starts, whereas RNA-Seq reads are

enriched 3' of mRNA starts. In addition, Omni-ATAC-Seq data shows greater enrichment at promoters than over annotated mRNA regions, and decreased enrichment outside of annotated mRNAs. These data suggest that our Omni-ATAC-Seq performed as expected in identifying promoter regions genome-wide.

To assess whether our libraries were capable of capturing significant variation over time, we performed Principal Component Analysis (PCA) on our libraries (Fig. 2C). PCA revealed that our libraries were primarily separated along two principal components (PCs), PC1 (60% of variation) and PC2 (18% of variation), with a considerable drop in variation explained in other PCs (Supp. Fig. 2D). PC1 appeared to be associated with developmental time, with earlier developmental stage libraries showing a negative loading, and later developmental stages showing a positive loading. Within the PCA plots, samples from the S21 and S22 time points appeared to be more separated in PC2 relative to the other timepoints. A discussion of this observation and its implications can be found in the Supplementary Methods and Supplementary Figures 2.2 and 2.3C.

To identify regions of dynamically accessible chromatin, we used the ImpulseDE2 (IDE2) pipeline (Fischer et al., 2018). IDE2 differs from other software for differential expression analysis in that it allows the investigation of trajectories of dynamic expression over large numbers of timepoints. It does so by modeling a gene expression trajectory as an "impulse" function that is the product of two sigmoid functions (Chechik and Koller, 2009; Yosef and Regev, 2011). This approach enables the modeling of a trajectory of gene expression in three parts: an initial value, a peak value, and a steady state value, thus summarizing an expression trajectory using a fixed number of parameters. With the ability to capture the differences between early, middle, and late expression values for each gene in a dataset, IDE2 also enables the detection of transient changes in gene expression or accessibility during a time course. Identifying differential expression over large numbers of timepoints is difficult for more categorical differential expression software such as edgeR and DESeq2, which generally use pairwise comparisons between timepoints to assess change over time (Love et al., 2014; Robinson et al., 2010).

We used Genrich (Gaspar) to call significant peaks (merged q-value between replicates < 0.05) in each of our 15 timepoints independently, and merged overlapping peaks across timepoints using bedtools (Quinlan and Hall, 2010), yielding 190,078 genomic regions which we used as our "peaks" in downstream analyses, including IDE2. The dynamic accessibility of these peaks is illustrated in a heatmap in Fig. 2D. Of these peaks, 163,227 (85.87%) were classified as having statistically significant variation (padj < 0.01) over our time-course; 60,909 (37.32%) were classified as having transient expression dynamics; and 88,231 (54.05%) were classified as showing positively or negatively monotonic expression dynamics. These results indicate that we were able to identify many dynamically accessible regions across the *Parhyale* genome.

IDE2 produces a fitted model of accessibility for each of the peaks used in the analysis (Fischer et al., 2018)(see Supp. Fig. 2.3C, D for examples). We used these model fits to summarize the global properties of the dynamic peaks we identified. First, we estimated the number of models that achieved their maximum accessibility during each of the 15 timepoints sampled (Fig. 2F). The timepoints during which models most frequently showed their highest accessibility were S13, S21, and S27, representing approximately the earliest, middle, and latest timepoints.

We then used these IDE2 model fits to understand how the variance of our libraries captured by the PCs was related to accessibility dynamics (Supp. Fig. 2G). We extracted the 100 peaks with the most positive and negative loadings in PC1 and PC2 and plotted their IDE2 model fits. Based on these plots, we inferred that PC1 captured variation along developmental time, with positively-loaded peaks showing a "low early, high late" accessibility trajectory, and negatively-loaded peaks showing a "high early, low late" accessibility trajectory. Peaks with high positive loading in PC2 appeared to be those with highest accessibility in the middle developmental stages.

These global analysis results indicate that our Omni-ATAC-Seq experiments captured information normally found in ATAC-Seq data, including strong enrichment at promoters. Overall, the low mitochondrial read contamination, large library size, and low fraction of duplicated reads suggest that our Omni-ATAC-Seq data are of high quality. In addition, the results of the IDE2 differential accessibility analyses indicate that the vast majority of accessible regions in the *Parhyale* genome show dynamic accessibility over developmental time.

1.1.2 – Improving the *Parhyale* genome annotation using short- and long-read RNA-Sequencing

While Omni-ATAC-Seq signal showed enrichment across annotated mRNA starts in a genome-wide analysis, careful examination of individual genes and gene models indicates that many of the MAKER gene annotations are fragmented. For example, among the *Parhyale Hox* genes, 7 of 9 the genes showed discrepancies in annotation compared to previously-published Rapid Amplification of cDNA Ends (RACE) data (Serano et al., 2016), with *lab, pb*, *Ubx*, and *abd-A* showing fragmented gene models (Supp. Fig. 3.1A, B).

A more accurate genome annotation would improve both genome-wide analyses and enable the precise classification of candidate regulatory elements as promoters, exonic and intronic regulatory elements, and intergenic regulatory elements. To generate a more complete genome annotation, we performed RNA-Seq using two approaches: short-read sequencing using the Illumina NovaSeq platform and long-read sequencing using Oxford Nanopore technology. For four developmental stages (S13, S19, S21, and S23), we generated triplicate RNA-Seq libraries, which were sequenced using a short-read approach (Illumina NovaSeq) and a long-read approach (Oxford Nanopore) (Fig. 3A, B; representative embryo images in Supp. Fig. 1B).

We assembled multiple transcriptomes from each of the sequencing approaches (Fig. 3B). Using the Trinity (Haas et al., 2013) pipeline to assemble the short-read sequences, we generated four developmental stage-specific *de novo* transcriptomes (Trinity-S13 to Trinity-S23) and two merged transcriptomes, one *de novo* (Trinity-all) and one genome-guided (Trinity-GG). Additionally, we generated two transcriptomes with reads from several developmental stages covering the time course of limb development (Trinity-limb (new) and Trinity-limb (old), see Methods). Although the merged transcriptomes were generated from all four developmental stages (Trinity-all, Trinity-GG), they scored lower compared to stage-specific transcriptomes when CDS completeness was assessed using BUSCO (Fig 3D). One possible explanation for this could be an increase in transcript assembly fragmentation from de Bruijn graph assembly caused by the high heterozygosity previously described of the *Parhyale* genome (Kao et al., 2016). For long-read sequencing, we mapped the reads to the phaw_5.0 genome

using minimap2 and assembled a transcriptome using StringTie2 (Kovaka et al., 2019; Li, 2018). We also used StringTie2 to generate a combined transcriptome containing both short and long reads. While the long-read transcriptome (StringTie2 L) yielded a low BUSCO score, the short- + long-read transcriptome (StringTie2 SL) scored comparably to our other transcriptomes (Fig. 3D).

To generate a more complete genome annotation, we used the Mikado pipeline to merge our assembled transcriptomes with the previously-generated Kao et al. 2016 transcriptome and the MAKER genome annotation. The Mikado pipeline and others (e.g. EvidentialGene) take advantage of the previously observed variable ability of transcriptome assemblers to produce complete gene models for different types of genes (Gilbert, 2019; Venturini et al., 2018). Such software uses a variety of metrics to compare transcripts from different sources to determine the "best" gene models for each gene region in the genome. We evaluated the Mikado-merged transcriptome (hereafter "Mikado transcriptome") using BUSCO, and observed a comparable score to the best of the individual transcriptomes (90.9% complete), and a marked improvement compared to the MAKER annotation (80.9% complete).

We first examined the Mikado transcriptome models representing the *Parhyale Hox* cluster genes and compared those to the RACE data and the MAKER annotation (Supp. Fig. 3 C–E). The gene models from the Mikado transcriptome showed an improvement relative to the MAKER models overall, with complete models for *pb, Hox3, Dfd* and *Scr*, which were fragmented in the MAKER annotation. While the Mikado pipeline was only able to recover one transcript isoform each for the three posterior *Hox* genes, for *abd-A* and *Ubx*, the isoforms identified were more complete than those the MAKER annotation. Moreover, the Mikado-merged transcriptome was able to produce a more complete gene model for *Hox3* than the previous RACE data or MAKER annotation. Combining the RACE data with our Mikado transcriptome, we have assembled the most complete annotation of the *Parhyale Hox* cluster to date.

To quantify the level of gene model fragmentation in our dataset, we generated a series of manual gene annotations based on RACE sequences. Among 143 previously-generated RACE transcripts, we selected 49 multi-exonic transcripts that appeared to have a single promoter based on the Omni-ATAC-Seq data (Supp. Fig. 3D). For each of these RACE transcripts, we manually annotated the extent of the first and last exon by comparing RNA-Seq read pileups to RACE data and the current genome annotation. We also identified the Omni-ATAC-Seq peak most likely to capture the gene promoter, based on the strength of the peak (as evaluated by the number of timepoints over which we observed a statistically significant peak), as well as overlap to RNA-Seq read data. We used these manual annotations (see Supp. Table 2) to evaluate the gene models from each of the different transcript sources (transcriptomes or gene annotations).

For each of the 49 genes in our dataset, we evaluated whether any models in each of the transcript sources in our dataset overlapped the promoter peak (Supp. Fig. 3E). The ability to unambiguously identify promoters is essential for downstream analyses, including building reporter constructs or targeting CRISPR guides to the 5'-most end of genes. Among the transcript sources, the Trinity *de novo* transcriptomes had the highest proportion (0.96) of genes where at least one model overlapped with the promoter peak. However, the Trinity models we observed often (0.98, 48/49 RACE genes examined for the Trinity-limb transcriptome) contained numerous spurious transcript

fragments in introns, exons, and 3'UTR regions (see Supp. Fig. 3.2 for summary statistics and examples). Among the remaining transcript sources, the Mikado transcriptome had the highest fraction of genes where at least one model overlapped with the promoter peak (0.84).

In addition, we assessed the degree of fragmentation of gene models across the 49 RACE transcripts (Supp. Fig. 3.1D, Methods). For each gene, we determined first whether any single gene model spanned the first and last exon, representing a complete "single" transcript. If there was not a single model, we next assessed whether two separate models overlapped with the first and last exon, or if all models only overlapped either the first or last exon. These results together formed a measure of transcript completeness. Comparing the different transcript sources, the StringTie2 SL transcriptome had the highest fraction (0.76) of complete "single" transcripts, with the Mikado transcriptome having a slightly lower fraction of "single" transcripts (0.73)(Supp. Fig. 3E).

When we directly compared the completeness measures between the Mikado transcriptome and the MAKER genome annotation, we observed a marked improvement. Notably, the Mikado transcriptome had a greater proportion of RACE genes with promoter-peak overlap (Mikado 0.84 > MAKER 0.71), as well as a greater fraction of "single" transcripts (Mikado 0.73 > MAKER 0.55)(Fig. 3D).

To further evaluate the quality of the Mikado transcriptome as a reference, we compared its performance to the MAKER annotation using a variety of metrics (Fig. 3E). The Mikado transcriptome produced a larger number of total gene models (n = 21,218) than are found in the MAKER annotation (n = 15,105). Using bedtools, we attempted to assign a candidate promoter peak to each transcript in each genome annotation (see Methods), and observed that the Mikado transcriptome had a larger number of transcripts with an assigned promoter (Mikado n = 18,946; MAKER n = 11,909), further suggesting improved genome-wide gene completeness.

We further examined the quality of the Mikado and MAKER transcriptomes by evaluating Omni-ATAC-Seq signal and RNA-Seq signal genome-wide at mRNA starts (Fig. 3F, G). We observed a comparable RNA-Seq signal, but a higher Omni-ATAC-Seq signal for the Mikado transcriptome as compared to the MAKER annotation, suggesting a stronger enrichment of promoters in our dataset. These results together suggest that the Mikado transcriptome produces a more complete genome annotation, capturing more of the *bona fide* genes in the *Parhyale* genome with greater accuracy than previously available.

To further assess the quality of the two genome annotations, we performed automated functional annotations to assign gene names and functions. We used two approaches: eggNOG and OrthoFinder (Emms and Kelly, 2019; Huerta-Cepas et al., 2019). eggNOG is a rapid and lightweight genome annotation software, and assigns gene names, KEGG pathway information, and GO terms, among numerous other metrics, to gene models. OrthoFinder facilitates the identification of orthogroups between provided peptide libraries, enabling automated comparisons between gene lists in species of interest.

We observed that the Mikado transcriptome included a greater number of GO-term assigned genes, eggNOG named genes, and eggNOG uniquely-named genes than found in the MAKER annotation (Fig. 3E). We used OrthoFinder to assign orthogroups between the Mikado or MAKER transcriptomes and the list of all *Drosophila melanogaster* peptides from the UNIPROT-SWISSPROT

database. The Mikado and MAKER transcriptomes produced similar numbers of orthogroups; however, the Mikado transcriptome produced a greater number of 1:1 orthogroups, which proved useful for downstream GO term enrichment analysis (summarized in Fig. 3E, see Supp. Fig. 3.3 for further explanation of orthogroup size comparisons).

Overall, we observed an improvement across nearly all metrics we used to compare the Mikado transcriptome to the MAKER annotation. The Mikado transcriptome appeared to produce more complete gene models than in the MAKER annotation, and more of these models could be annotated using automated methods such as eggNOG and OrthoFinder. Moreover, the Mikado transcriptome draws from a diverse range of tissue sources, as the previous Kao transcriptome and MAKER annotations that contribute to the Mikado transcriptome also incorporated data from adult tissues and other embryonic stages (See Supp. Fig. 3.1G for contributions to Mikado from each transcript source). Given these results, we used the Mikado transcriptome as our reference genome annotation for downstream analyses.

Using the Mikado transcriptome as a reference, we assigned Omni-ATAC-Seq peaks to a number of different spatial categories ("peak types") along the genome (Fig. 3H, see Methods). We observed approximately equivalent numbers of peaks over gene bodies (34.4%: 6.2% exonic, 28.2% intronic) and in intergenic regions (37.2%: 7% proximal intergenic, 30.2% distal intergenic). We further partitioned the intergenic peaks into proximal and distal segments, with distal intergenic peaks representing those peaks >10kb away from the nearest gene. We established this cutoff with the rationale that peaks beyond this distance would be considerably more difficult to isolate as single fragments combined with a promoter peak using PCR, agnostic of their orientation with respect to the promoter element. Such peaks could not have been easily identified using previous transgenic approaches, and are thus distinguished from "proximal" peaks by their necessary identification using techniques such as ATAC-Seq. The majority of intergenic peaks in our dataset were distal intergenic peaks (82%; 30.2%/ 37.2%), indicating that many intergenic regulatory elements could not have been identified using previous approaches. The average distance of intergenic peaks was 73,351bp away from the nearest gene (Supp. Fig. 2.3B). Nearly a fifth of regulatory elements (18.2%) were also located on contigs that did not contain genes, and were classified as "unknown".

As our peaks represent the merged combination of peaks across individual timepoints, we also evaluated the proportion of peak types from each time-point (Fig. 3H). Peaks analyzed in this way could reveal differences in the spatial distribution of significant peaks across the genome over developmental time. After normalizing for the number of "unknown" peaks, the proportion of peaks belonging to intergenic and promoter regions declined slightly as development progressed, while the proportion of gene body peaks increased (Supp. Fig. 2.3E). There appeared to be slightly more intergenic peaks than gene body peaks at all developmental time points, and the ratio of intergenic peaks to gene body peaks declined very slightly over time (Supp. Fig. 2.3F). These data indicate that intergenic and gene body peaks have different enrichment trajectories over time. As development progresses, gene body peaks are increasingly enriched, and may have a greater impact on gene regulation.

1.1.3 – Inferring nucleosome positioning and transcription factor footprints from Omni-ATAC-Seq

ATAC-Seq and other accessibility approaches primarily identify regions of open chromatin in the genome. However, with sequencing performed at sufficient depth, one can infer additional information from the pattern of insertions of ATAC-Seq adapters into the genome. Using our Omni-ATAC data, we predicted the temporal positioning of nucleosomes near Omni-ATACpeaks using NucleoATAC and inferred the footprints of transcription factors bound to accessible peaks using HINT-ATAC.

NucleoATAC enables the prediction of nucleosome positioning from ATAC-Seq data (Schep et al., 2015). We applied NucleoATAC to a window of +/- 500bp around our Omni-ATAC-Seq peaks at each developmental stage. NucleoATAC depends on a stereotypical signal of Tn5 insertion around nucleosomes referred to as a "V-plot", which resembles the fragmentation pattern around chromatin subjected to chemical fragmentation. We observed across developmental time-points a clear V-plot signal generated by NucleoATAC (example of V-plot from S21 libraries shown in Supp. Fig. 4A).

A well-established signal of nucleosome positioning found across animals is the presence of strong +1 and -1 nucleosome positioning around gene promoters, and a depletion of nucleosomes at RNA-Pol II binding sites (referred to as nucleosome-free region, or NFR) (Brogaard et al., 2012; Radman-Livaja and Rando, 2010). We visualized NucleoATAC signal and nucleosome occupancy around mRNA starts genome-wide, and observed strong signals associated with +1 and -1 nucleosomes (Fig. 4 A, B; Supp. Fig. 4 B–D). We also observed a depletion of nucleosomes and a decrease in NucleoATAC signal between the +1 and -1 nucleosomes, which may represent nucleosome-free regions. This signal was observed consistently in each of our stage-specific libraries (Supp. Fig. 4 E, F). These data indicate that we were able to identify the stereotypical signal of nucleosomes at promoters from our Omni-ATAC data. Moreover, these results suggest that NucleoATAC is more generally able to infer nucleosome positions and nucleosome-free regions within our dataset genome-wide.

In addition to inferring nucleosome positions, deeply sequenced ATAC-Seq data also enables the inference of transcription factor footprints. We used HINT-ATAC to identify transcription factor footprints in Omni-ATAC-Seq peaks from each developmental stage (Li et al., 2019) and evaluated the enrichment of these TF footprints across developmental time. We observed that HINT-ATAC enriched TF footprints appeared to group into three broad clusters based on the timing of highest enrichment: early, S21/S22, and late (Fig. 4D, E). These trends were true both when examining all JASPAR CORE position frequency matrices (PFMs) (Fornes et al., 2020), and when examining only those JASPAR CORE PFMs from *Drosophila*. These results are consistent with the general trends observed in PCA, which suggest that the early, middle, and late timepoints in development have the greatest variation.

The data generated from these analyses enable the visualization of chromatin accessibility, predicted nucleosome position, inferred transcription factor binding, and RNA-Seq expression at individual loci across around half of *Parhyale* development. The strengths of these data are illustrated in Fig 4C at the predicted promoter of the *Parhyale Engrailed-1* locus for stage S21. Thus, for any

genomic region of interest, one can develop a comprehensive prediction of the local chromatin environment at each of the 15 timepoints in our dataset.

1.1.4 – Identifying gene regulatory programs from Omni-ATAC-Seq using fuzzy clustering

To assess the potential for our dataset to provide new biological insights, we attempted to identify distinct gene regulatory programs based on differential accessibility. We used the Mfuzz package to perform fuzzy c-means clustering on the matrix of read counts generated from our Omni-ATAC-Seq peaks (Kumar and E Futschik, 2007). Fuzzy clustering differs from categorical clustering approaches, such as hierarchical clustering and k-means clustering, in that it assigns a probability for each element in the dataset to fall into each of the identified clusters. For example, rather than a given peak falling into either cluster A or B exclusively, a given peak might have a 95% chance of falling into cluster A and a 5% chance of calling into cluster B.

The Mfuzz package allows the specification of two parameters: c, or the number of clusters to create from the data, and m, the stringency of the clustering. To determine the optimal number of clusters in our fuzzy c-means clustering, we varied the c-value from 3 to 13, and evaluated the quality of clustering using the cluster overlap metric provided by Mfuzz, as well as the silhouette score metric (Supp. Fig. 5.2)(Kumar and E Futschik, 2007). We determined that 9 clusters yielded the highest mean silhouette score, while avoiding clusters below the silhouette score average and minimizing size variability between clusters. Figure 5A plots with t-stochastic neighbor embedding (t-SNE, implemented in scikit-learn) (Hinton and Roweis, 2003; Maaten and Hinton, 2008; Pedregosa et al., 2011) the top 10,000 most significant peaks called by IDE2, where each point is colored by its IDE2 model max fit, while Figure 5B illustrates the same clusters, colored by their Mfuzz cluster.

We evaluated the differences between clusters through a variety of approaches. Figure 5D shows the standardized raw accessibility scores of peaks in each of the 9 clusters over time. We first assessed the spatial category distribution of each cluster compared to the global average (Figure 5C). Clusters 1, 2, and 8 appeared to be enriched for “unknown” regions, while clusters 3-7 appeared to show slightly greater proportions of intronic elements than distal intergenic elements, counter to the global average. Cluster 8 is notable in that it is both the smallest cluster, and the cluster where a majority of peaks are found in “unknown” regions. In addition, we used the IDE2 model fits for each cluster to evaluate the general accessibility dynamics of each cluster (Supp. Fig. 5.1B, C). Clusters 1 and 2 appear to be enriched for peaks that achieve high accessibility early in development and decrease in accessibility over time. Clusters 3–6 appear to be enriched for peaks that achieve maximum accessibility at different timepoints along developmental time. Cluster 7 appears to be enriched for peaks that show low accessibility early in development, and increased accessibility late in development. Finally, clusters 8 and 9 showed a more striking pattern, with Cluster 8 composed of peaks that appear to have a pulsatile increase in accessibility around stages S21–S22, and Cluster 9 composed of peaks with an opposite pattern: those with a sudden pulsatile decrease around S21–S22. Thus, the 9 clusters identified in our analysis appeared to capture distinct patterns of accessibility throughout time.

We further investigated the clusters identified by Mfuzz using additional metrics in order to understand the biological significance of these clusters. First, we investigated the relationship between accessibility dynamics and nucleosome positioning between clusters (Fig. 5E, Supp. Fig. 5.1D). For

each developmental time point, we visualized Omni-ATAC accessibility (solid line, black axis) compared to NucleoATAC signal (dashed line, blue axis) for peaks belonging to each cluster. We observed that peaks belonging to Clusters 3–7 and 9 all appeared to have an inverse relationship between accessibility and NucleoATAC signal; this inverse relationship was most obvious at the time points during which these peaks achieved their maximum accessibility (Fig. 5E, blue dashed boxes). Moreover, these peaks all appeared to have a depletion of nucleosome occupancy at peak centers relative to the region around peaks (Supp. Fig. 5.1D, solid line, pink axis). By contrast, peaks in clusters 1, 2, and 8 appeared to show a counterintuitive and opposite result: both high NucleoATAC signal and high nucleosome occupancy at timepoints during which those peaks achieved maximum accessibility (Fig. 5E, Supp. Fig. 5.1D, solid line, pink axis). As with the clusters that showed an inverse relationship between accessibility and occupancy, the clusters that showed a parallel relationship between accessibility and occupancy showed the strongest relationship at those timepoints during which they achieved maximum accessibility (Fig. 5E, red dashed boxes).

To assess the biological function of each of the 9 Mfuzz clusters, we performed GO term enrichment analysis. To assign functions to each peak, we extracted the *Drosophila melanogaster* gene assigned to the nearest gene for each peak, and performed GO enrichment analysis on the list of unique gene names associated with peaks from each cluster relative to the list of all OrthoFinder-assigned gene names in the genome. Fig. 6A illustrates the top 10 most-enriched GO terms for each of the clusters. While we were not able to identify significantly enriched GO-terms (FDR < 0.05) for Clusters 1, 2, and 8, Clusters 3–7 and 9 all showed enrichment of developmental GO-terms. These enriched GO-terms generally appeared to match the expected biological functions of peaks based on their accessibility dynamics. For example, Cluster 3 peaks, which generally achieve maximum accessibility during germ band elongation stages in development, showed enrichment for GO terms including head segmentation and neuroblast fate, two processes which take place during this developmental period. Meanwhile, Cluster 7 peaks appeared to be enriched for neuronal terms (neuropeptide signaling pathway, synaptic target recognition, neuron recognition), potentially reflecting the function of genes near these peaks which achieve their maximum accessibility towards the very end of the developmental time-course, during which much of the morphology of the organism has been specified, and the embryo begins to twitch inside the egg. Notably, Clusters 3–7 showed slightly greater enrichment of intronic peaks than distal intergenic peaks, suggesting that intronic peaks may be more likely to regulate developmental processes. These results resemble the observation from recent work showing that tissue-specific enhancers are enriched in intronic regions in various human cell types (Borsari et al., 2021).

To further investigate the clusters in our dataset, we examined the enrichment of TF footprints generated from HINT-ATAC in each of our clusters. For each cluster, we compiled all of the TF footprints predicted across developmental stages for all peaks within that cluster. We compared the enrichment of those TF footprints to our randomly generated background data, and examined all JASPAR CORE transcription factors, as well as the subset of JASPAR CORE transcription factors found in *Drosophila*.

We observed strong differential enrichment of different groups of transcription factors in each of our 9 clusters. Notably, particular families of transcription factors appeared enriched in several of

the clusters. For example, Cluster 1 and Cluster 2 appeared to have enrichment for footprints matching C2H2 zinc-finger transcription factors and homeodomain transcription factors. Clusters 4 and 7 showed enrichment for footprints matching beta helix-loop-helix (b-HLH) transcription factors. Meanwhile, clusters 5, 8, and 9 showed strong enrichment for footprints of ethylene response factor (ERF), DM-type intertwined zinc-finger factor (DM IZF), and forkhead domain (FKH) transcription factor families, respectively.

Given that many JASPAR position frequency matrices (PFMs) come from non-arthropod sources, we attempted to identify possible *Drosophila* motif matches for the enriched transcription factor footprints found in the JASPAR CORE dataset by comparing the most-enriched JASPAR CORE and JASPAR *Drosophila* PFMs. We used the STAMP tool through the JASPAR webpage to align motif sequences based on similarity (Mahony and Benos, 2007). We identified groups of aligned PFMs between non-*Drosophila* and *Drosophila* PFMs (see Supp. Fig. 6). Selected pairs of motifs are shown in Fig. 6C for each cluster (non-*Drosophila* labeled in black, *Drosophila* labeled in red). For some pairs, we were able to identify candidate *Drosophila* transcription factors that appeared highly similar to non-*Drosophila* sequences. For example, in Cluster 9, we observed strong enrichment for FKH family transcription factors, whose PFMs matched the br(var.4) and slp1 PFMs. This suggests that peaks found in Cluster 9 may be regulated by br(var.4) and slp1 binding in *Parhyale*.

For other clusters, we had variable success. For example, we were unable to identify a strong *Drosophila* match to explain the strong enrichment of ERF transcription factors (found in *Arabidopsis thaliana*) in Cluster 5. Nonetheless, footprints matching ERF transcription factors appeared to be highly enriched in Cluster 5, suggesting that there may be a transcription factor within the *Parhyale* genome whose PFM matches that of ERF transcription factors, and which may be important for the regulation of Cluster 5 peaks.

Altogether, our results indicate that clustering Omni-ATAC data using accessibility can identify groups of transcription factors with similar accessibility trajectories. In addition, these clusters can be further analyzed using nucleosome occupancy, GO enrichment, and transcription factor footprint enrichment to understand possible biological functions and genetic mechanisms behind differential accessibility.

1.1.5 – Concordant and discordant expression and accessibility dynamics appear across development

To understand the relationship between peak accessibility and gene expression, we examined the correlation between Omni-ATAC accessibility in individual peaks with respect to RNA-Seq expression for the nearest gene at each of the developmental stages for which we generated RNA-Seq data. We first examined the relationship between accessibility at all peaks compared to their nearest genes (Supp. Fig. 7A) and observed a very weak (Pearson $R^2 = 0.11–0.15$, Spearman $\rho = 0.1–0.14$) but statistically significant ($p < 0.001$) positive correlation between peak accessibility and the expression of nearby genes. When we examined only promoter peaks (Fig. 7B), we observed a slightly greater correlation between accessibility and gene expression (Pearson $R^2 = 0.19–0.22$, Spearman $\rho = 0.17–0.25$) correlation that was also significant ($p < 0.001$). Notably, at each developmental stage, we observed a higher correlation between accessibility and expression at promoter peaks than at all peaks.

We expect that regulatory elements with a variety of impacts on gene expression might be found among our ATAC-Seq peaks. To further classify peaks, we examined the relationship between the change in accessibility of individual peaks and the change in expression of nearby genes. We used DESeq2 to evaluate the $\log_2$-fold change of both expression (RNA-Seq) and accessibility (ATAC-Seq) for all genes and peaks. We then correlated the differential accessibility and gene expression for all peak-gene pairs where both the differential accessibility of a peak and expression of the peak's nearest gene were statistically significant (FDR/ padj < 0.05). When examining all peaks, we observed a weak but significant correlation between $\log_2$-fold change in accessibility and gene expression (Pearson R^2 = 0.25–0.27, Spearman ρ = 0.27–0.31)(Fig. 7A). This correlation was slightly stronger at promoter peaks (Pearson R^2 = 0.28–0.40, Spearman ρ = 0.35–0.41)(Fig. 7C).

Among our differential accessibility and expression analyses, we classified two broad classes of peak-gene pairs representing concordant and discordant accessibility and expression dynamics. Peaks with concordant accessibility and gene expression were those where the sign of the $\log_2$-fold change in expression and accessibility were in agreement—for example, peaks where an increase in accessibility was observed concurrently with an increase in expression. Meanwhile, peaks with discordant accessibility and gene expression were those where an increase in accessibility was observed concurrently with a decrease in expression, or vice versa.

We observed both classes of peak-gene pairs among all peaks, but also among promoter peaks. Over time, the number of concordant peak-gene pairs increased from 1,888 at the S19-vs-S13 comparison, to 5,738 at the S21-vs-S19 comparison, and then decreased to 5,240 at the S23-vs-S21 comparison (Supp Fig. 7D). Meanwhile, the number of discordant peaks gradually increased over time (761 to 2,231 to 3,115 peak-gene pairs)(Supp Fig. 7D). These trends were also observed for the promoter-only peaks. The gradual increase in discordant peaks may indicate an increase in repressive gene regulatory mechanisms over time, such as chromatin compaction, as developmental gene networks become refined over the course of differentiation.

Overall, the classification of peaks as concordant or discordant in accessibility and gene expression may provide downstream users of this data with hypotheses about cis-regulatory element function. Given that cis-regulatory elements have recently been shown to function as both activating elements and silencers, depending on tissue context (Gisselbrecht et al., 2020; Halfon, 2020), we cannot directly map concordant peaks to enhancers or discordant peaks to silencers. Without information about higher-order chromosome contacts, it is also difficult to precisely assign a given regulatory element's function to a particular gene of interest.

However, information about concordance and discordance of peaks could be useful for researchers deciding which among many peaks surrounding a gene of interest could be most fruitful for reporter construction. To facilitate analysis of accessibility and expression for peak-gene pairs of interest, we have included functions enabling visualization of these two factors for arbitrary peak-gene pairs, illustrated in Supp. Fig. 7.2. The visualization tool displays accessibility and gene expression for user-selected Omni-ATAC peaks and Mikado genes over RNA-Seq stages (S13, S19, S21, S23), fold change in accessibility and expression between adjacent timepoints, significance of fold change as evaluated by DESeq2, and concordance/discordance assignment at each time point.

1.1.6 – Omni-ATAC-Seq re-identifies known *Parhyale* regulatory elements

To assess the practical usefulness of our Omni-ATAC-Seq dataset, we compared our Omni-ATAC-Seq peaks to known regulatory elements in *Parhyale*. A very limited set of regulatory elements have been described in *Parhyale*: a muscle reporter, PhMS (Pavlopoulos and Averof, 2005); a heat shock element, HS2a (Pavlopoulos et al., 2009); an embryonic ubiquitous reporter, PEB (Danielle Liubicich, unpublished); and two reporter constructs for *Parhyale Opsin-1* and *Opsin-2* (Ramos et al., 2019). We were unable to locate any Omni-ATAC-Seq peaks for the *Parhyale Opsin* genes, likely owing to their late expression (outside of our developmental time course) and very low cell number (a handful of cells per eye). However, strong Omni-ATAC-Seq enrichment was observed at the PhMS, HS2a, and PEB elements, as illustrated in Supp. Fig. 8.1.

Each of these reporter constructs consists of the first exon and intron of a gene. In each case, we observed an Omni-ATAC peak overlapping the 5' end of the gene. For the PhMS reporter (Supp. Fig. 8.1B), previous work had identified a cluster of putative b-HLH transcription factor binding sites. These sites were captured within our Omni-ATAC peak. The HS2a reporter has been reported to contain two binding sites for heat shock factor (HSF) proteins upstream of a minimal promoter (PhHsp70), both of which are captured within our Omni-ATAC peak (Supp. Fig. 8.1D). Finally, the PEB element appears to contain two Omni-ATAC peaks. These results suggest Omni-ATAC is able to identify the position of functional cis-regulatory elements, and has been able to capture those regions important for cis-regulatory element function.

1.1.7 – *Minos* transposase reporter assays can reveal novel promoters and distal enhancers

Given that our Omni-ATAC peaks were able to capture known regulatory elements, we attempted to use these peaks to identify novel regulatory elements. Careful examination of developmentally important genes revealed that many are surrounded by large numbers of peaks (>10) spread over large genomic distances. To further filter our Omni-ATAC peaks, we identified regions of sequence conservation to another amphipod crustacean, *Hyalella azteca* (Poynton et al., 2018), using the VISTA sequence alignment software (Ratnere and Dubchak, 2009)(Supp. Fig. 8.2A). *Hyalella* serves as a useful comparison to *Parhyale*, as its genome size is smaller (1.05 Gb), but as it is also an amphipod crustacean, we expect that key developmental regulatory elements might have some level of sequence conservation.

Among our Omni-ATAC peaks, we took two strategies to identify candidate reporter elements. In the first approach, we identified all peaks within 5kb of mRNA starts. In doing so, we were able to identify a handful of genomic regions in which we were able to locate both a putative promoter and a candidate proximal enhancer. We also identified several putative promoters, which we tested in isolation. In the second approach, we examined the genomic regions around important developmental genes of interest. Many of the genes we examined showed very large numbers of strong peaks, and were therefore intractable to thorough analysis. For the purposes of this study, we focused on three regions: the region around the *Engrailed-1* gene, the region around the *Sp-69* gene, and the region around the *Heat shock protein 70* complex, where two previous cis-regulatory elements (PhMS and HS2a) had been identified.

To assess the function of candidate regulatory elements identified by Omni-ATAC-Seq, we employed a *Minos* transposase reporter assay (Fig. 8A). In this assay, we injected *Minos* transposase mRNA along with a transposon donor plasmid containing a reporter gene construct that expressed DsRed to 1- and 2-cell embryos. Once per day, from 3–10dpf, we screened for DsRed expression on a Zeiss LSM780 confocal microscope.

The reporters we examined are summarized in Table 1. In the course of our experiments, we observed numerous spontaneous expression patterns (Supp. Fig. 9). These expression patterns may represent rare integration events, or may reflect fortuitous insertions into regions of the genome with promiscuous regulatory sequences that facilitate expression.

One expression pattern observed from the Sp69-p3 enhancer appeared to be repeated in 5 embryos along the ~300 we injected. This expression pattern was characterized by strong expression in the head, antenna 1, and antenna 2 (Supp. Fig. 10). In the same set of experiments, we also observed three spontaneous expression patterns that appeared to differ amongst each other, labeling different regions of the limbs and gills. It is possible that the repeated expression observed from the Sp69-p3 enhancer represents a genuine expression pattern. We tested this hypothesis by raising three Sp69-p3 head-expressing animals to adulthood and mating the two adults that showed the strongest expression during development. However, none of the offspring of this cross showed obvious head expression, or any obvious DsRed expression elsewhere in the embryo.

Among the novel reporter constructs we tested, two showed robust expression: pMi(Hsc70-4) and pMi(ne1)-Hsp70-p2. The pMi(Hsc70-4) construct was constructed using the putative promoter peak of the Hsc70-4 gene (Fig. 8B, D). This reporter showed strong expression in a cluster of neurons associated with the eye beginning at around 9dpf (Fig. 8C). The expression pattern of these neurons appears to be distinct from those observed using a synthetic 3XP3 enhancer, which has previously been used as a positive marker of transformation *Parhyale* (Pavlopoulos and Averof, 2005). Thus, it is possible to build new reporters from our Omni-ATAC peaks using putative promoter peaks.

The second reporter that showed robust expression was the pMi(ne1)-Hsp70-p2 construct, which contains the PhHsp70 minimal promoter along with a strong peak extracted from the Hsp70 cluster, about 30kb away from the location of the minimal promoter (Fig. 8E, G). Hsp70-p2 drives strong expression in cells of the embryonic yolk (Fig. 8F). Some yolk cells labeled by this reporter moved dynamically around the yolk, and while others appeared static and showed spongiform morphology (Supp. Video 1). We observed defects in the development of the dorsal portion of the embryo in mid-stage (~6–7dpf) embryos injected with this construct, which stereotypically displayed a separation between the embryo proper and the eggshell (Fig. 8F, bottom panel, marked with arrow). While we were unable to establish a genetic line using this construct, we observed strong and reproducible expression in many embryos. This reporter is the first indication that distal enhancers exist and can be identified in *Parhyale*.

While it remains challenging to identify novel cis-regulatory elements in *Parhyale*, our results suggest that functional regulatory elements can be identified from our Omni-ATAC data. Among the numerous peaks identified in our dataset, we expect that many will reveal novel gene regulatory dynamics. Future work to identify such elements will require the optimization of reporter expression and transgenic strategies to increase the throughput and sensitivity of reporter gene assays.

1.2 – Discussion

New genomic data enable deeper understanding of underlying biology. This is especially true in emerging research organisms, for which data and resources are limited. This work provides a wealth of genomic resources for the amphipod crustacean *Parhyale hawaiensis*. In addition to providing Omni-ATAC-Seq and RNA-Seq data for a broad developmental time course, our work generates multiple new transcriptomes, an updated genome annotation, a catalog of dynamically accessible chromatin regions, and predictions of nucleosome occupancy and transcription factor binding. Moreover, our thorough analysis of regulatory element dynamics through clustering, GO enrichment, predicted TF binding, and correlations with RNA-Seq present numerous hypotheses for cis-regulatory element function. Such data will support researchers in the growing *Parhyale* community in efforts to identify and characterize developmental regulatory elements, and provide the foundations for more advanced approaches, including single-cell sequencing techniques, which rely on high-quality reference datasets.

1.2.1 – The *Parhyale* genome contains many distant and dynamic regulatory elements

From our data, we were able to glean new information about the global dynamics of gene regulation during *Parhyale* development, as well as the composition of the *Parhyale* genome. We observed that the vast majority of our peaks showed dynamic accessibility over developmental time, and that peaks could have a variety of different temporal dynamics, including transient increases and decreases in accessibility alongside more absolute increases and decreases. These results indicate that much of the *Parhyale* genome undergoes dynamic changes in accessibility over developmental time.

With an improved genome annotation, we were also able to determine that many (~37.2%) regulatory elements in *Parhyale* are located in intergenic regions, and that among all elements at all timepoints, elements located >10kb away from the nearest gene (distal intergenic peaks) were the largest group, followed by intronic peaks. These results are notable, given that many of the currently studied protostome genomes are small (for example, *Drosophila melanogaster* is 180Mb, *Tribolium castaneum* is 200Mb, and *Bombyx mori* is 530Mb). In such organisms, regulatory elements tend to occur within short distances from promoters, and it is common in such systems to attempt to build reporter genes using a short window upstream of the promoter of a gene of interest.

Parhyale is an example of an arthropod with a large genome (3.6 Gb, or ~10% larger than *Homo sapiens* at 3.2 Gb). Given the large genome size, and the large proportion of elements located distant to genes, we expect that regulatory elements within *Parhyale* may be located more distantly from gene promoters than in other arthropods. Future attempts at building reporter genes in this organism will need to account for the large number of distant regulatory elements, which may be critical to proper reporter expression.

Our data are among the first to identify regulatory elements genome-wide in a non-insect arthropod (Gatzmann et al., 2018; Kissane et al., 2021), and also examine one of the largest sequenced arthropod genomes available to date (Kao et al., 2016). These data will enable other researchers to examine the relationship between genome size and regulatory element composition across more diverse taxa, and will be useful in developing a deeper understanding of how genome organization affects gene expression across the metazoans.

1.2.2 – Combined short- and long-read transcriptomes can improve genome annotations

In addition to providing a resource for the *Parhyale* community, our work can also serve as a guide for other researchers working on developing such datasets in emerging research organisms. First, our work demonstrates the utility of combining short- and long-read sequencing in generating more accurate genome annotations for emerging research organisms. Even with a small number of sequences (~1.2 million reads) generated using inexpensive Nanopore sequencing, we were able to assemble a transcriptome with a moderate BUSCO score (71%). Combining short reads and long reads (StringTie2 SL), we were able to construct a transcriptome with both a high BUSCO score (91.8%) and low rate of gene fragmentation.

Moreover, the completeness of the StringTie2 SL transcriptome met or exceeded that of several of our Trinity-derived transcriptomes. Examination of individual genomic regions in the *Parhyale* genome indicates that Trinity generates numerous spurious transcripts, which can be confounding for whole-genome analysis (see Supp. Fig. 3.2 for examples). Thus, for other researchers with access to a genome assembly, we would strongly recommend performing both short- and long-read sequencing, and assembling using multiple different assembly strategies to generate higher quality genome annotations. For those without access to a high-quality genome, we would caution against relying strictly on transcripts generated from software such as Trinity, which in our system generated many transcripts that did not appear to match with previous RACE data. Filtering transcripts by expression values, as well as employing transcriptome-merging pipelines such as Mikado (in the case of a high-quality genome) and EvidentialGene (in the absence of a genome), would likely help remove such spurious transcripts.

1.2.3 – Deep analysis of ATAC-Seq data can reveal more than just dynamic accessibility

While the most direct analysis of ATAC-Seq data can enable the identification of dynamic regulatory elements, our work uses additional tools to infer nucleosome positioning and transcription factor binding from our ATAC-Seq data. Using NucleoATAC, we were able to recover clear signals of nucleosome positioning at promoters, a quality observed in numerous organisms. Using HINT-ATAC, we were able to predict transcription factor binding across developmental stages. Such analyses enable a glimpse of the possible chromatin landscape of the genome, particularly in the absence of more direct and specific, but also more time-consuming and expensive approaches such as ChIP-Seq or CUT&RUN.

In addition to performing inference, we were able to identify clusters of peaks with potentially distinct biological functions using fuzzy clustering. These clusters of peaks appeared to differ in their accessibility dynamics, nucleosome positioning, and enrichment of transcription factor footprints, indicating that meaningful biological differences can be gleaned from deeper analysis of ATAC-Seq data. For example, we observed strong enrichment in Cluster 9 for FKH transcription factors footprints matching the PFMs for *Drosophila* br(var.4) and slp1. The strongest peaks in this cluster appeared to share a common decrease in accessibility during stages S21–S22, and an enrichment for genes with GO terms related to cytoskeletal function. This group of peaks appeared to show a decrease in accessibility at a time point during which the embryo undergoes a dramatic morphogenetic event in

which it splits along the midline. Based on our data, it is possible that FKH domain-containing transcription factors may play a role in that important morphogenetic event. Using our analyses as a starting point, other *Parhyale* researchers may be able to make hypotheses about their own developmental processes of interest.

1.2.4 – Omni-ATAC-Seq identifies old and new regulatory elements, including enhancers

Our data were able to capture most of the previously-identified regulatory elements that had been described for *Parhyale*. In addition, we were able to demonstrate that our data contain new regulatory elements. Notably, we demonstrate the first identification of a distal regulatory element, Hsp70-p2, located about 30kb from the PhHsp70 minimal promoter. These results are the first identification of an enhancer separated by a large distance from a minimal promoter element in this organism.

Our results demonstrate that Omni-ATAC-Seq is able to identify novel regulatory elements, and we expect that numerous new reporter genes will be built from these data. Previous work relied on examining candidate regulatory elements within individual genomic regions, or attempting to build new reporters through random integration events. With this dataset, it is now possible to identify candidate regulatory elements for any gene of interest and to screen for expression patterns using the Minos reporter system. Moreover, by evaluating concordance and discordance of gene expression and accessibility using this dataset, it is possible to further refine the list of candidates for reporter construction. As this approach enables the direct investigation of candidate regulatory regions with known sequence, this will also enable the generation of novel synthetic reporters generated by modifying endogenous sequences. Such future directions will enable thorough dissection of regulatory elements to determine how transcription factor binding influences gene expression during *Parhyale* development.

While we were unable to identify regulatory elements that drove clear expression patterns for the handful of genes we tested, numerous additional peaks near these genes remain to be investigated. Future work to build new reporters in *Parhyale* may require the development of more efficient transgenesis and screening strategies. For example, improving the efficiency of CRISPR-mediated homologous recombination or CRISPR-mediated NHEJ transgenesis could enable researchers to insert a reporter near a candidate regulatory region of interest. Moreover, future approaches will also need to account for the numerous distant regulatory elements in the *Parhyale* genome, which may be important for gene regulation. Examining local DNA interactions using approaches such as Hi-C or other chromatin conformation capture approaches will be instrumental in identifying distant regulatory regions.

Together, these approaches have enabled the exploration of chromatin dynamics and transcription factor binding in an emerging model organism. Our work illustrates how a single, easily adapted protocol can yield data amenable to deep analysis. We would recommend that other researchers using ATAC-Seq or similar assays in emerging model organisms also take advantage of the additional information that could be gleaned from deep analysis of their data, which could provide further insights into how both local and global changes to nucleosome occupancy and transcription factor binding influence their biological processes of interest.

1.2.5 – Conclusion

By combining Omni-ATAC-Seq with RNA-Seq across a broad developmental time course, our work is able to identify and classify numerous candidate cis-regulatory elements in the genome of the amphipod crustacean *Parhyale hawaiensis*. We demonstrate how deep analysis of Omni-ATAC data can facilitate the identification of peaks with distinct accessibility, nucleosome occupancy, and transcription factor footprint enrichment. We further classify peaks as concordant or discordant regulatory elements by integrating differential accessibility and differential expression, revealing potential relationships between regulatory elements and nearby genes. Moreover, we show the potential to identify novel reporter genes using candidate promoters and enhancers from our data.

This work provides a substantial resource to the *Parhyale* community, and should accelerate the study of gene regulation in this emerging research organism. In addition, our work can serve as a framework for other researchers deploying ATAC-Seq and RNA-Seq approaches in emerging research organisms. Through deep analysis of ATAC-Seq data, combined short- and long-read sequencing, and integration of accessibility and gene expression, such researchers can identify regulatory elements with distinct biological functions and advance the study of gene regulation in their organisms of interest.

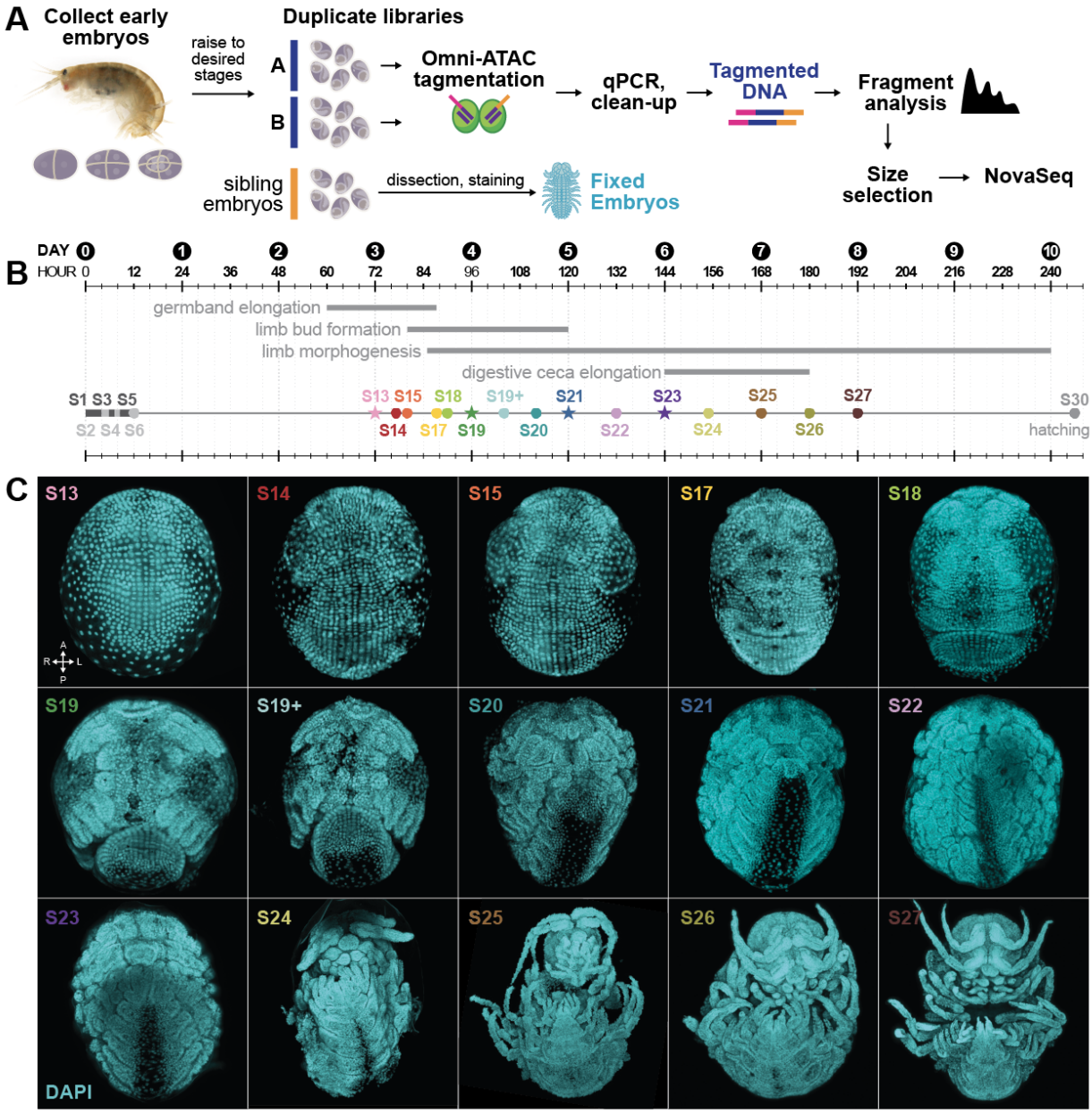


Fig. 1.1: Time-course ATAC-Seq in *Parhyale hawaiensis* embryos.

A) Overview of ATAC-Seq protocol. Embryos were collected at early developmental stages and raised to specific developmental timepoints. Duplicate libraries were generated by tagmenting 5 embryos each from a single clutch of sibling embryos. qPCR analysis was used to determine the optimum number of additional cycles of PCR. Tagmented DNA was cleaned and fragment analysis was performed to assess library quality. Final libraries were pooled in equal concentrations, size-selected, and sequence on an Illumina NovaSeq short read sequencer.

B) Timeline of developmental stages. RNA-Seq libraries were also generated for timepoints marked with a star.

C) Representative embryo images from clutches used for Omni-ATAC-Seq. Embryos are stained with DAPI and mounted ventral-side up. (A: anterior, P: posterior, L: left, R: right).

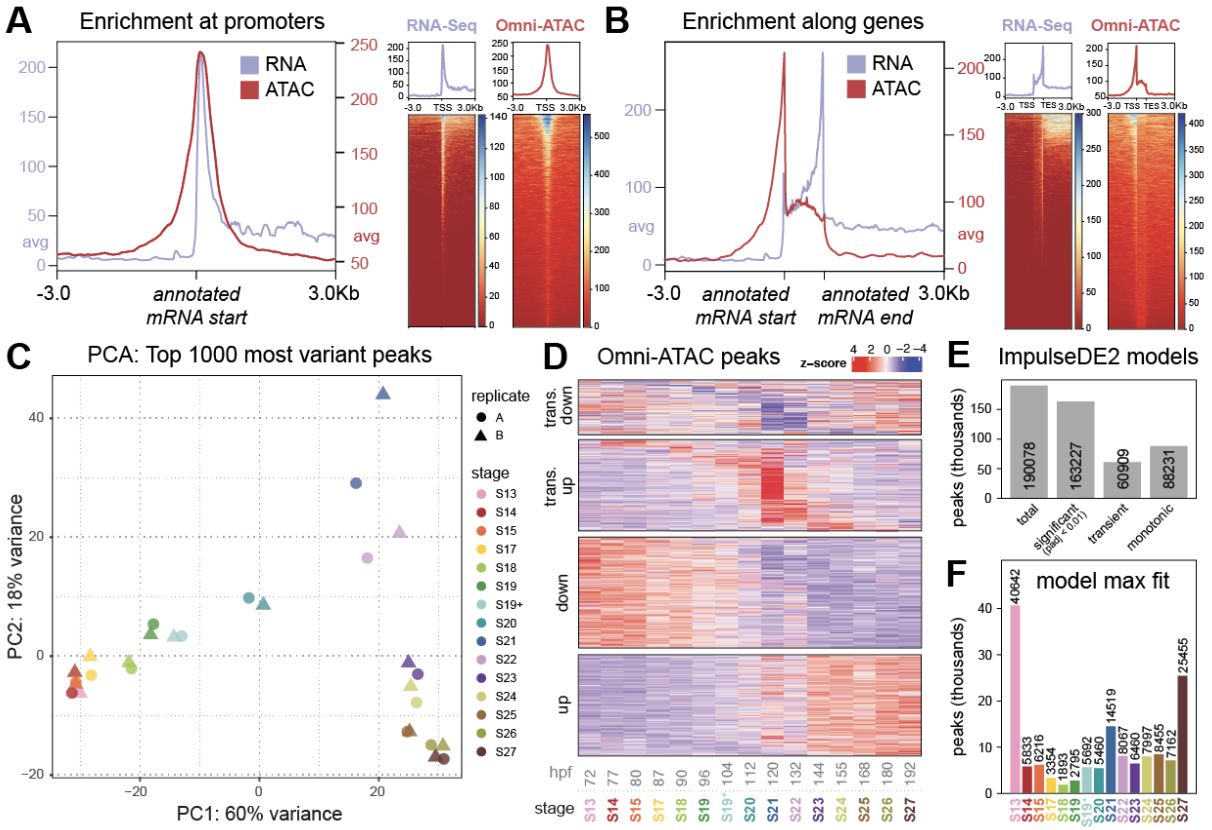


Fig. 1.2: Whole-genome analysis of ATAC-Seq reveals enrichment of promoter signals, dynamic accessibility over developmental time

A) Omni-ATAC and RNA-Seq signal enrichment at promoters. Annotated mRNA start positions are aligned, and average signal across genomic positions is plotted for either Omni-ATAC or RNA-Seq signal. Axes reflect the average signal of RNA-Seq or Omni-ATAC reads across all mRNA start positions in the genome. A strong enrichment of RNA-Seq signal is observed 3' of mRNA start sites, while enrichment of Omni-ATAC-Seq signal appears symmetric around mRNA start sites.

B) Omni-ATAC and RNA-Seq signal enrichment across gene bodies. Omni-ATAC signal is greatly enriched 5' of gene bodies and slightly enriched across genes, while RNA-Seq reads are enriched across gene bodies, with an increase in enrichment towards the 3' end of gene bodies.

C) PCA of Omni-ATAC libraries, generated using DESeq2. PCA loadings were assigned to the top 1000 variant peaks, where variance was measured as row variance across all libraries. PC1 (60% of variance) appears to correlate with developmental time. PC2 (18% variance) appears to be associated with middle developmental timepoints.

D) Heatmap of Omni-ATAC accessibility dynamics, generated from ImpulseDE2 model fits. Colors reflect a Z-score calculated from ImpulseDE2, where red (high Z) indicates increase in accessibility and blue (low Z) indicates decrease in accessibility.

E) Bar chart indicating number of peaks classified as significant, transient, and monotonic by ImpulseDE2.

F) Bar chart quantifying stage of maximum accessibility for significant ImpulseDE2 model fits. Peaks predominantly achieved maximum accessibility at the start and end of the time course.

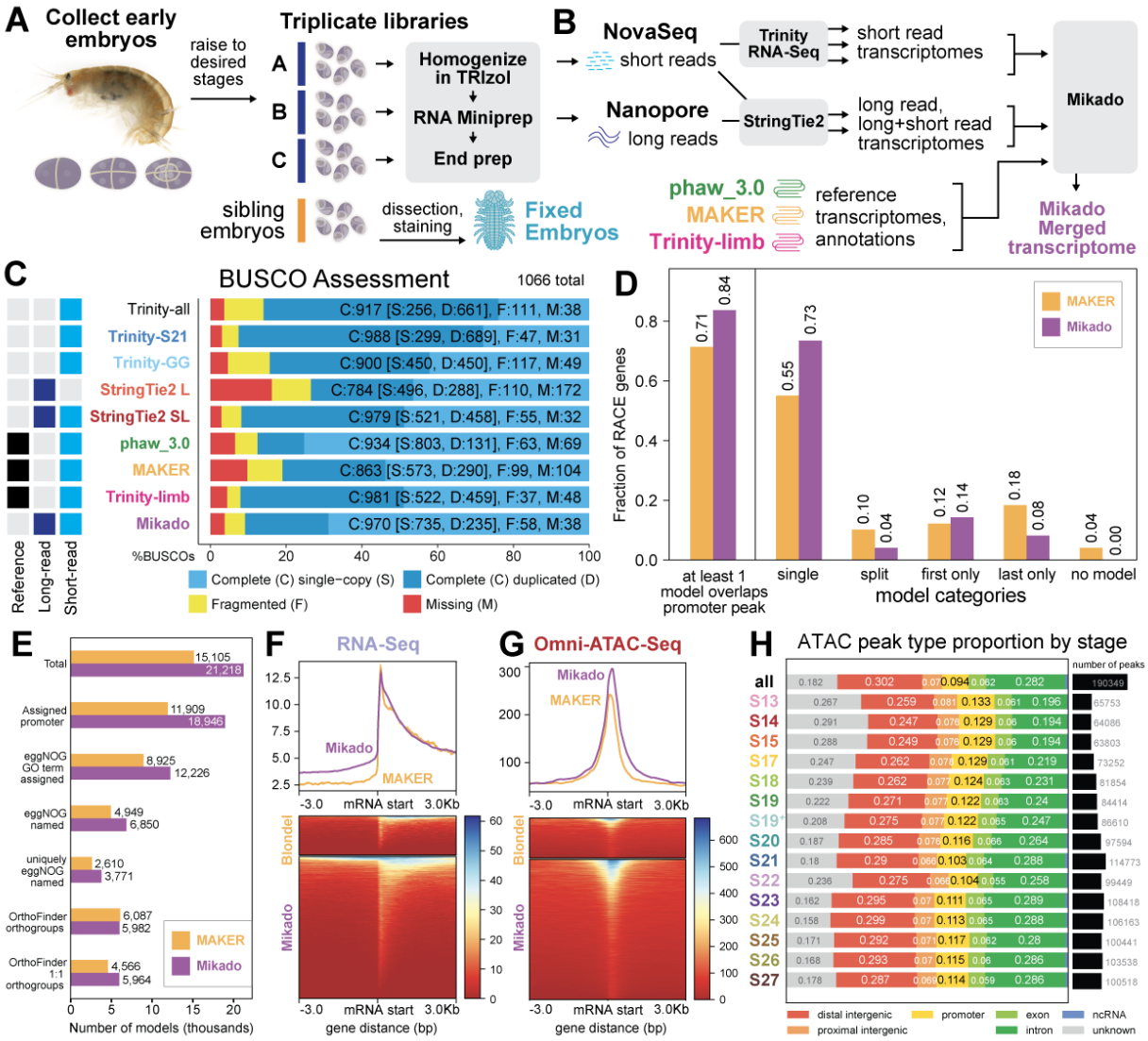


Fig. 1.3: Time-course RNA-Seq of Parhyale embryos

A) RNA-Seq protocol overview. Triplicate libraries were generated for four developmental stages.

B) RNA-Seq transcriptome assembly and merging pipeline. Short-read, long-read, and long- + short-read transcriptomes were generated for pooled developmental stages matched to Omni-ATAC timepoints (S13, S19, S21, S23) using Trinity and StringTie2. Transcriptomes were also assembled using Trinity for short reads from a limb developmental time course (see Methods). Assembled transcriptomes were then merged with additional transcripts from previous publications into a Mikado transcriptome.

C) BUSCO scores for representative datasets and Mikado transcriptome. Overall, the StringTie2 short + long -read transcriptome (StringTie2 SL) performed comparably to the best Trinity transcriptomes. The Mikado transcriptome appeared to have a comparable BUSCO complement to the other datasets with high BUSCO completeness.

D) Gene model completeness for Mikado transcriptome vs. MAKER genome annotation based on a dataset of 49 RACE genes. The Mikado transcriptome appears to have more complete gene models and less fragmentation.

E) Comparison of number of gene models of different functional annotation categories between MAKER genome annotation and Mikado transcriptome. The Mikado transcriptome appears to produce more annotated genes with a higher annotation quality.

F) Plot comparing RNA-Seq signal pileups between MAKER mRNA starts and Mikado mRNA starts. Axis represents mean signal at mRNA start sites for each dataset.

G) Plot comparing Omni-ATAC-Seq signal pileups between MAKER mRNA starts and Mikado mRNA starts.
H) Categorization of Omni-ATAC-Seq peaks in each stage-specific ATAC-Seq library by their position relative to Mikado gene models. Distal intergenic peaks are defined as those greater than 10kb away from the nearest gene, whereas unknown peaks were peaks that were located on a contig that did not have a gene model. A large majority of peaks (>71%) were intergenic peaks, and a majority of peaks (>51%) were distal intergenic peaks. The proportion of peaks of each type appeared relatively stable across developmental stages.

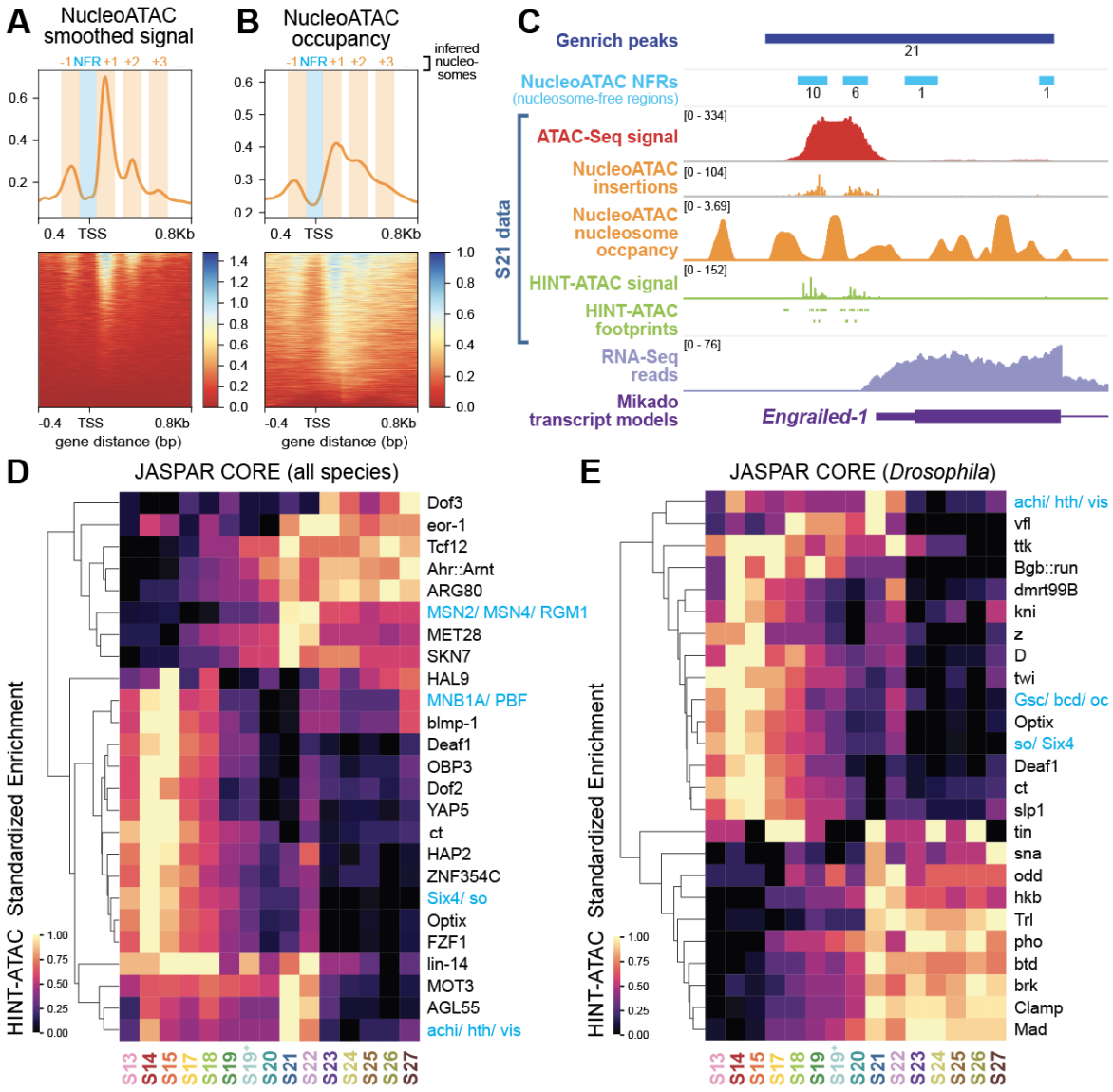


Fig. 1.4: Inference of nucleosome positioning and transcription factor binding from Omni-ATAC-Seq data

A) NucleoATAC smoothed signal at mRNA starts genome-wide. A clear pattern of peaks in the signal can be observed, suggestive of nucleosome positioning. Axis reflects mean signal across all mRNA starts at stage S21.

B) Inferred nucleosome occupancy at mRNA starts genome-wide. Axis reflects mean signal across all mRNA starts at stage S21. Inferred nucleosome annotations on A and B added based on observed signal.

C) Visualization of NucleoATAC, HINT-ATAC, Omni-ATAC, and RNA-Seq data at the inferred Engrailed-1 promoter at stage S21. Deeper inference on Omni-ATAC-Seq data allows insight into the broader chromatin landscape at developmental time points included in this dataset.

D) Summary of standardized enrichment of JASPAR CORE position frequency matrices (PFMs) among HINT-ATAC identified footprints across developmental stages. The top 25 most-enriched PFMs are displayed. Color is based on standardized enrichment, where minimum and maximum enrichment within each row are set to 0 and 1, respectively. PFMs marked in blue had identical enrichment ratios within each group at all timepoints, likely due to having highly similar PFMs. Overall, PFMs appeared enriched at early, late, or middle developmental stages.

E) Summary of top 25 most-enriched *Drosophila* PFMs from the JASPAR CORE database. Overall, PFMs appeared enriched at early, late, or middle developmental stages.

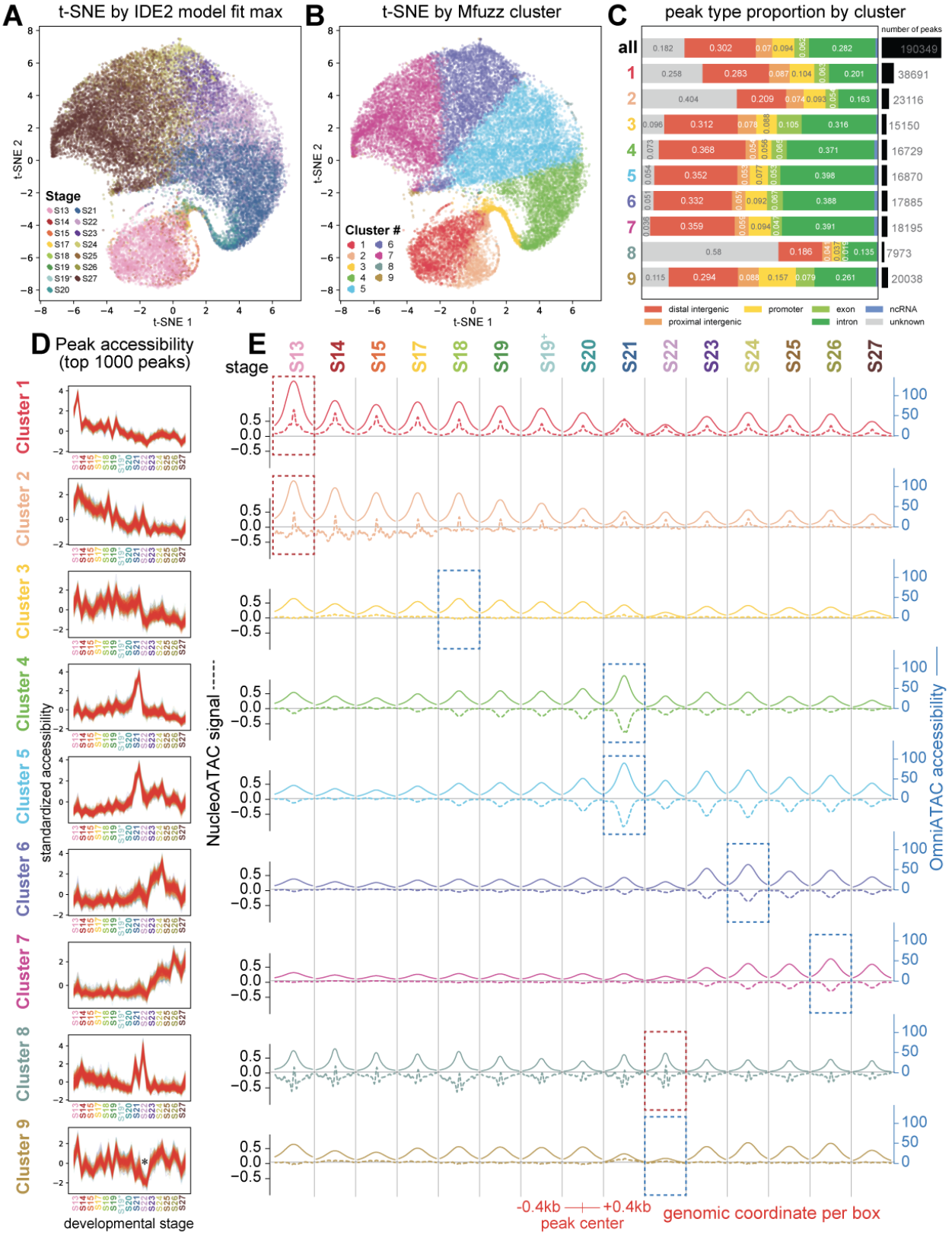

A
t-SNE by IDE2 model fit max
t-SNE 1
t-SNE 2
Stage
S13 S21
S14 S22
S15 S23
S17 S24
S18 S25
S19 S26
S19+ S27
S20
B
t-SNE by Mfuzz cluster
Cluster #
1 6
2 7
3 8
4 9
5
C
peak type proportion by cluster
number of peaks
all 0.182 0.302 0.07 0.094 0.066 0.282 190349
1 0.258 0.283 0.087 0.104 0.067 0.201 38691
2 0.404 0.209 0.074 0.093 0.064 0.163 23116
3 0.096 0.312 0.078 0.083 0.105 0.316 15150
4 0.073 0.368 0.054 0.058 0.082 0.371 16729
5 0.054 0.352 0.053 0.077 0.055 0.398 16870
6 0.051 0.332 0.057 0.092 0.067 0.388 17885
7 0.038 0.359 0.055 0.094 0.047 0.391 18195
8 0.58 0.186 0.047 0.013 0.135 7973
9 0.115 0.294 0.088 0.157 0.079 0.261 20038
distal intergenic
proximal intergenic
promoter
exon
intron
ncRNA
unknown
D
Peak accessibility
(top 1000 peaks)
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
standardized accessibility
developmental stage
E
stage
S13 S14 S15 S17 S18 S19 S19+ S20 S21 S22 S23 S24 S25 S26 S27
NucleoATAC signal
OmniATAC accessibility
-0.4kb
+0.4kb
peak center
genomic coordinate per box

Fig. 1.5: Identification and classification of regulatory element clusters

A) t-SNE plot of the top 10,000 most statistically significant peaks identified by ImpulseDE2 (IDE2), colored by IDE2 max fit.

B) t-SNE plot of the same points from A colored by Mfuzz cluster.

C) Distribution of peak position categories by Mfuzz cluster. Clusters, 1, 2, and 8 appeared to be enriched for "unknown" peaks, while other clusters appeared to be enriched for distal intergenic peaks.

D) Standardized accessibility plot for top 1000 peaks with strongest membership in each of the 9 clusters. Each library is plotted as a separate point along the line plot. Line color indicates the cluster membership value for each peak in each plot, with dark red indicating strong cluster membership, and blue indicating weaker cluster membership. An asterisk in the Cluster 9 panel marks the decrease in accessibility observed commonly in Cluster 9 peaks, more clearly shown in the IDE2 model fits in Supp. Fig. 5.1B.

E) Line plots showing Omni-ATAC accessibility (solid line, blue axis) and NucleoATAC signal (dashed line, black axis) for peaks in each cluster across time. Each box contains a line plot summarizing each signal at all peaks within that cluster with respect to a given developmental Omni-ATAC-Seq library. The center of each line plot reflects the average signal at the center of all Omni-ATAC-Seq peaks in that cluster at that developmental stage. Signal is visualized at 0.4kb upstream and downstream of peak centers. Within each signal type, the axes across line plots for each cluster are identical. Dotted lines around subsections of the plot highlight regions of interest. Dotted blue lines indicate the time point during which peaks in a given cluster achieved both maximum accessibility and decreased nucleosome occupancy. Dotted red lines indicate the time point during which peaks in a given cluster achieved both maximum accessibility and increased nucleosome occupancy.

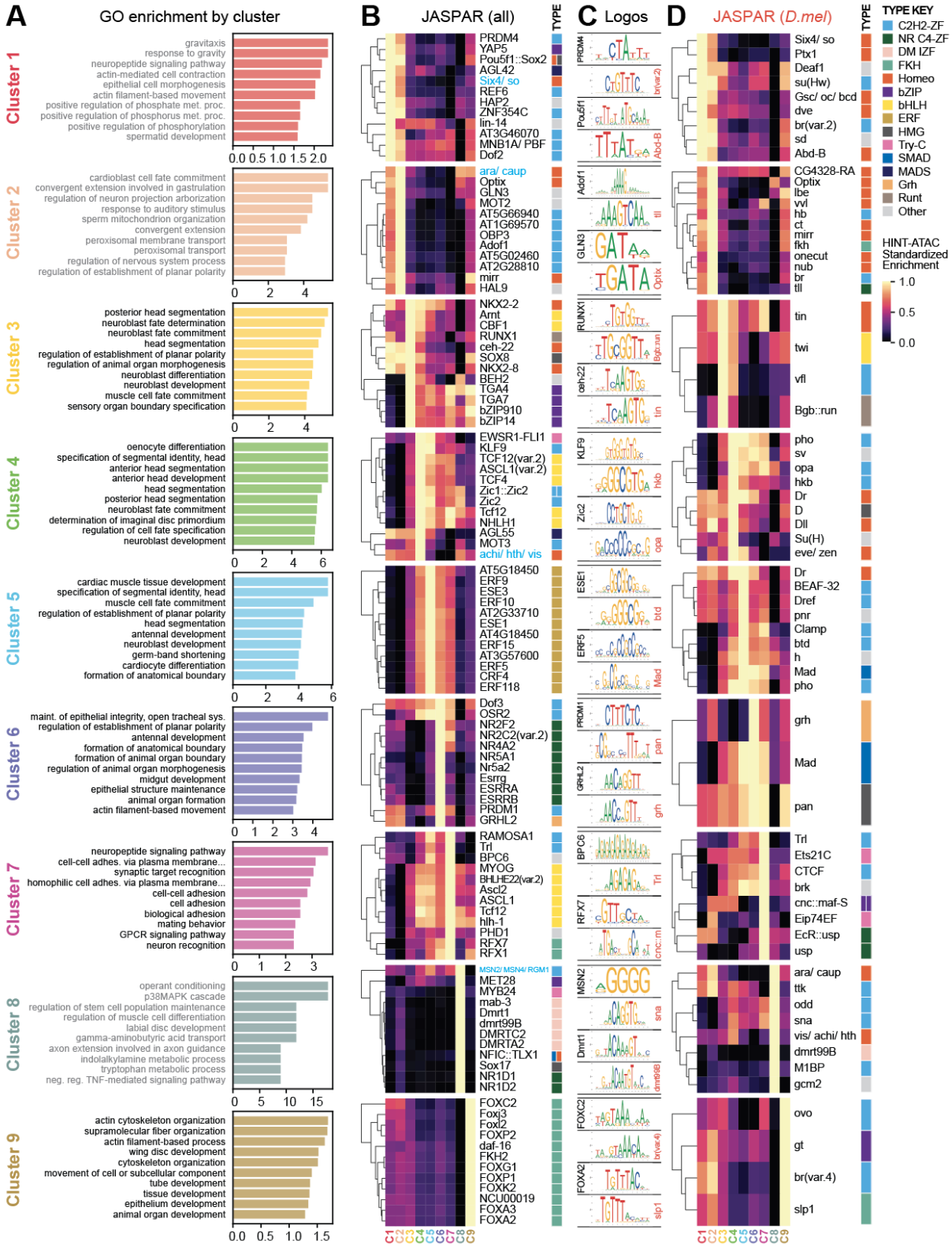

A
GO enrichment by cluster
Cluster 1
gravitaxis
response to gravity
neuropeptide signaling pathway
actin-mediated cell contraction
epithelial cell morphogenesis
actin filament-based movement
positive regulation of phosphate met. proc.
positive regulation of phosphorus met. proc.
positive regulation of phosphorylation
spermatid development
0.0 0.5 1.0 1.5 2.0
Cluster 2
cardioblast cell fate commitment
convergent extension involved in gastrulation
regulation of neuron projection arborization
response to auditory stimulus
sperm mitochondrion organization
convergent extension
peroxisomal membrane transport
peroxisomal transport
regulation of nervous system process
regulation of establishment of planar polarity
0 2 4
Cluster 3
posterior head segmentation
neuroblast fate determination
neuroblast fate commitment
head segmentation
regulation of establishment of planar polarity
regulation of animal organ morphogenesis
neuroblast differentiation
neuroblast development
muscle cell fate commitment
sensory organ boundary specification
0 2 4
Cluster 4
oenocyte differentiation
specification of segmental identity, head
anterior head segmentation
anterior head development
head segmentation
posterior head segmentation
neuroblast fate commitment
determination of imaginal disc primordium
regulation of cell fate specification
neuroblast development
0 2 4 6
Cluster 5
cardiac muscle tissue development
specification of segmental identity, head
muscle cell fate commitment
regulation of establishment of planar polarity
head segmentation
antennal development
neuroblast development
germ-band shortening
cardiocyte differentiation
formation of anatomical boundary
0 2 4 6
Cluster 6
maint. of epithelial integrity, open tracheal sys.
regulation of establishment of planar polarity
antennal development
formation of anatomical boundary
formation of animal organ boundary
regulation of animal organ morphogenesis
midgut development
epithelial structure maintenance
animal organ formation
actin filament-based movement
0 1 2 3 4
Cluster 7
neuropeptide signaling pathway
cell-cell adhes. via plasma membrane...
synaptic target recognition
homophilic cell adhes. via plasma membrane...
cell-cell adhesion
cell adhesion
biological adhesion
mating behavior
GPCR signaling pathway
neuron recognition
0 1 2 3
Cluster 8
operant conditioning
p38MAPK cascade
regulation of stem cell population maintenance
regulation of muscle cell differentiation
labial disc development
gamma-aminobutyric acid transport
axon extension involved in axon guidance
indolalkylamine metabolic process
tryptophan metabolic process
neg. reg. TNF-mediated signaling pathway
0 5 10 15
Cluster 9
actin cytoskeleton organization
supramolecular fiber organization
actin filament-based process
wing disc development
cytoskeleton organization
movement of cell or subcellular component
tube development
tissue development
epithelium development
animal organ development
0.0 0.5 1.0 1.5
B
JASPAR (all)
TYPE
PRDM4
YAP5
Pou5f1::Sox2
AGL42
Six4/ so
REF6
HAP2
ZNF354C
lin-14
AT3G46070
MNB1A/ PBF
Dof2
ara/ caup
Optix
GLN3
MOT2
AT5G66940
AT1G69570
OBP3
Adof1
AT5G02460
AT2G28810
mirr
HAL9
NKX2-2
Arnt
CBF1
RUNX1
ceh-22
SOX8
NKX2-8
BEH2
TGA4
TGA7
bZIP910
bZIP14
EWSR1-FLI1
KLF9
TCF12(var.2)
ASCL1(var.2)
TCF4
Zic1::Zic2
Zic2
Tcf12
NHLH1
AGL55
MOT3
achi/ hth/ vis
AT5G18450
ERF9
ESE3
ERF10
AT2G33710
ESE1
AT4G18450
ERF15
AT3G57600
ERF5
CRF4
ERF118
Dof3
OSR2
NR2F2
NR2C2(var.2)
NR4A2
NR5A1
Nr5a2
Esrrg
ESRRA
ESRRB
PRDM1
GRHL2
RAMOSA1
Trl
BPC6
MYOG
BHLHE22(var.2)
Ascl2
ASCL1
Tcf12
hlh-1
PHD1
RFX7
RFX1
MSN2/ MSN4/ RGM1
MET28
MYB24
mab-3
Dmrt1
dmrt99B
DMRTC2
DMRTA2
NFIC::TLX1
Sox17
NR1D1
NR1D2
FOXC2
Foxj3
Foxl2
FOXP2
daf-16
FKH2
FOXG1
FOXP1
FOXK2
NCU00019
FOXA3
FOXA2
C1 C2 C3 C4 C5 C6 C7 C8 C9
C
Logos
PRDM4
br(var.2)
Pou5f1
Abd-B
Adof1
ttl
GLN3
Optix
RUNX1
Bgb::run
ceh-22
tin
KLF9
hkb
Zic2
opa
ESE1
btd
ERF5
Mad
PRDM1
pan
GRHL2
grh
BPC6
Trl
RFX7
cnc::m
MSN2
sna
Dmrt1
dmrt99B
FOXC2
br(var.4)
FOXA2
slp1
D
JASPAR (D.mel)
TYPE
Six4/ so
Ptx1
Deaf1
su(Hw)
Gsc/ oc/ bcd
dve
br(var.2)
sd
Abd-B
CG4328-RA
Optix
lbe
vvl
hb
ct
mirr
fkh
onecut
nub
br
ttl
tin
twi
vfl
Bgb::run
pho
sv
opa
hkb
Dr
D
Dll
Su(H)
eve/ zen
Dr
BEAF-32
Dref
pnr
Clamp
btd
h
Mad
pho
grh
Mad
pan
Trl
Ets21C
CTCF
brk
cnc::maf-S
Eip74EF
EcR::usp
usp
ara/ caup
ttk
odd
sna
vis/ achi/ hth
dmrt99B
M1BP
gcm2
ovo
gt
br(var.4)
slp1
C1 C2 C3 C4 C5 C6 C7 C8 C9
TYPE KEY
C2H2-ZF
NR C4-ZF
DM IZF
FKH
Homeo
bZIP
bHLH
ERF
HMG
Try-C
SMAD
MADS
Grh
Runt
Other
HINT-ATAC
Standardized
Enrichment
1.0
0.5
0.0

Fig. 1.6: Functional annotation of clusters using GO enrichment and transcription factor binding prediction

A) GO enrichment per cluster based on the nearest gene for each peak. For each peak in each cluster, we identified the nearest gene and extracted the *Drosophila* OrthoFinder gene name for that gene. We performed GO enrichment using *Drosophila* GO terms, using the list of all OrthoFinder gene names found in our genome as the background for the analysis. Bar charts display fold enrichment for each GO term. Gray text represents GO terms that did not have a significant p-value ($p < 0.05$) after false discovery rate (FDR) correction.

B) HINT-ATAC enrichment per cluster for JASPAR CORE PFMs. For each cluster, we identified all unique TF footprints across all 15 timepoints found within peaks of that cluster and performed enrichment relative to a randomly-selected background. We show the top 12 PFMs with the greatest enrichment for each cluster that also had the highest enrichment in that cluster relative to other clusters. Colored boxes represent the transcription factor family of each transcription factor in the clustermap.

C) JASPAR motif logos for select pairs of transcription factors between non-*Drosophila* and *Drosophila* TFs with similar motifs as determined by STAMP alignment. For each cluster, two pairs of transcription factors are shown: a non-*Drosophila* transcription factor (labeled in black) and a *Drosophila* transcription factor (labeled in black). In some cases, a strong similarity is observed between the non-*Drosophila* and *Drosophila* PWMs.

D) HINT-ATAC enrichment per cluster for *Drosophila* JASPAR CORE PFMs. Some clusters had fewer than 12 PFMs for which that cluster had the highest enrichment.

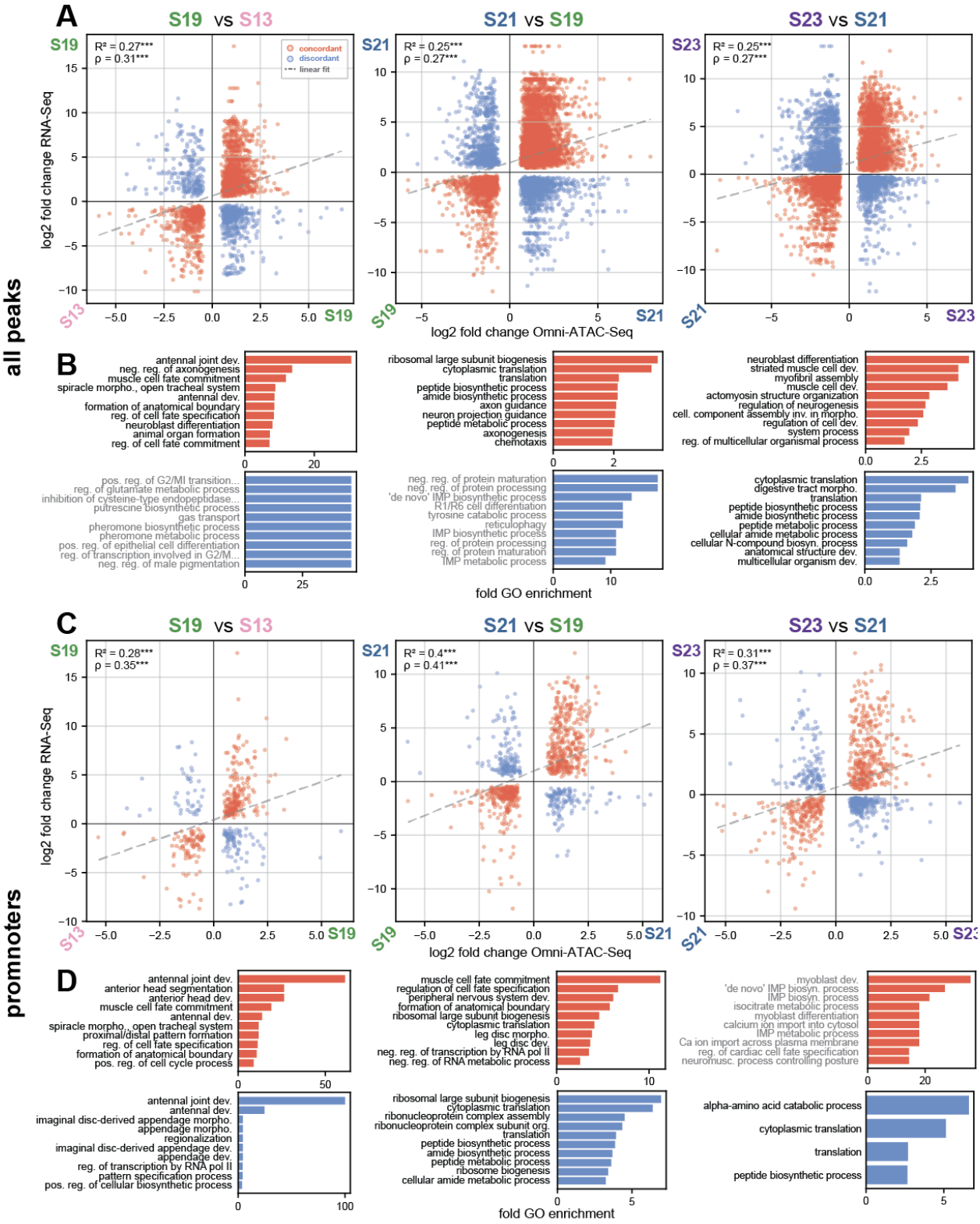

A
S19 vs S13
S21 vs S19
S23 vs S21
all peaks
log2 fold change RNA-Seq
log2 fold change Omni-ATAC-Seq
concordant
discordant
linear fit
B
fold GO enrichment
C
promnoters
D

Fig. 1.7: Correlation between RNA-Seq and Omni-ATAC-Seq

A) Correlation between log_2-fold change RNA-Seq and log_2-fold change Omni-ATAC-Seq at peak-gene pairs in sequential pairwise comparisons of developmental stages. Only peak-gene pairs with a significant difference in expression and accessibility between the two time points are plotted. Positive axis reflects higher expression or accessibility at the later developmental stage for each plot. Points colored in red reflect concordant relationships between log_2-fold change in RNA-Seq and Omni-ATAC-Seq; blue peaks reflect discordant relationships. Dotted line represents a linear fit to all data in the plot. Pearson correlation R^2 and Spearman correlation ρ for all points are displayed in each plot.

B) GO-term enrichment for gene-peak pairs with concordant or discordant expression and accessibility log_2-fold change, for all peaks. Gene lists were extracted based on the nearest Mikado gene to each peak.

C) Correlation between log_2-fold change RNA-Seq and log_2-fold change Omni-ATAC-Seq at all peaks in each developmental stage. Positive axis reflects higher expression at the later developmental stage for each plot.

D) GO-term enrichment for gene-peak pairs with concordant or discordant expression and accessibility log_2-fold change, for promoter peaks.

(*** = $p < 0.001$)

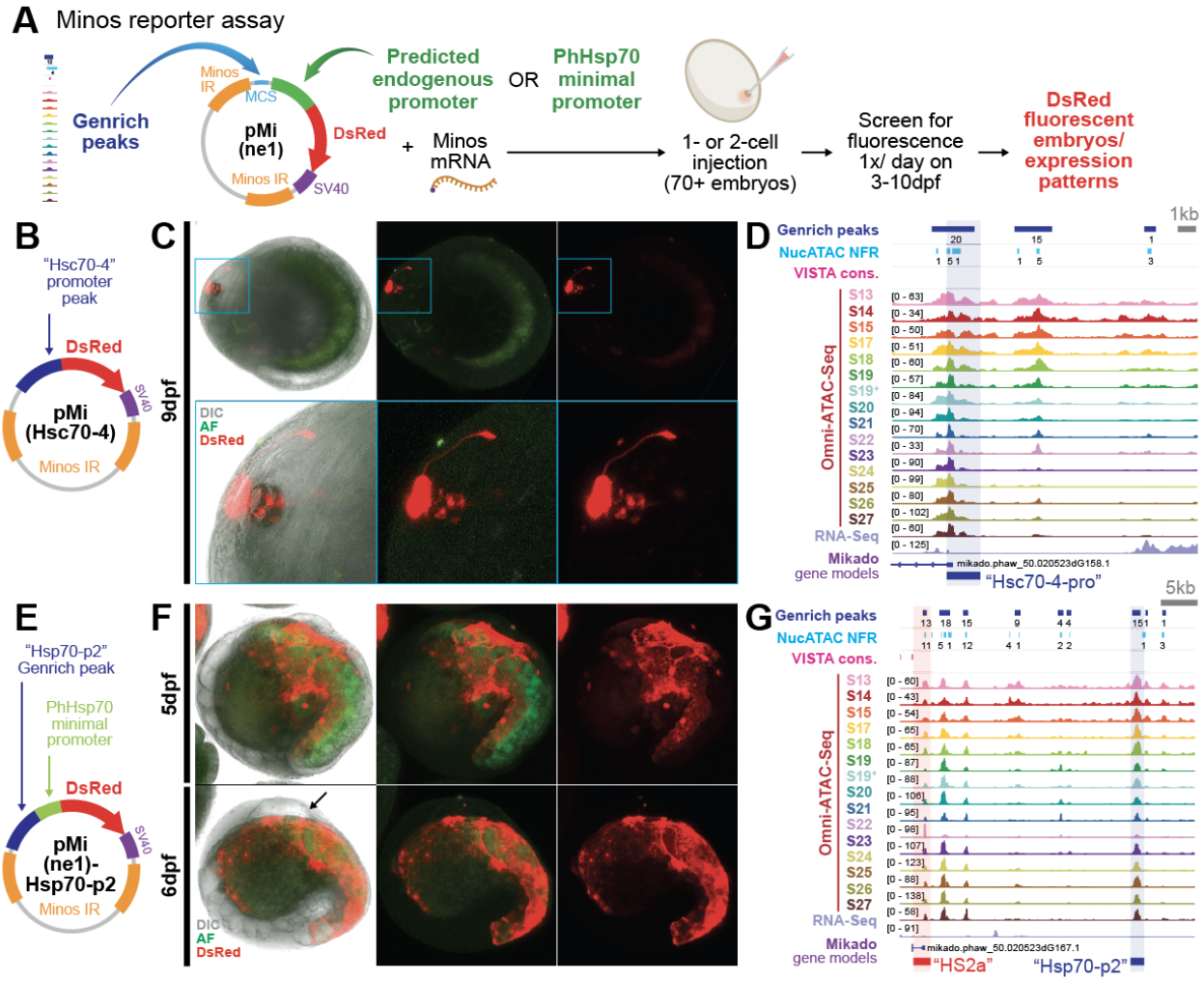


Fig. 1.8: Minos transposase reporter assay identification of novel regulatory elements

A) Minos transposase reporter assay and screening approach.

B) Plasmid schematic for the pMi(Hsc70-4) reporter.

C) Expression of the pMi(Hsc70-4) reporter in neurons associated with the eye of a 9hpf embryo. Bottom row shows a zoomed-in image of the head region. A clear projection of the neurons can be seen.

D) The genomic region of the Hsc70-4 promoter peak.

E) Plasmid schematic for the pMi(ne1)-Hsp70-p2 reporter.

F) Expression of pMi(ne1)-Hsp70-p2 reporter at 5 and 6dpf showing strong expression in the yolk. Arrow marks aberrant morphology at the dorsal side of the embryo observed consistently in all embryos with fluorescent expression.

G) The genomic region of the Hsp70-p2 peak relative to the HS2a element. The PhHsp70 minimal promoter used in this construct is located within the HS2a element.

Table 1.1: Plasmids tested using Minos transgenesis

Plasmid	Type	Gene – *Drosophila* expression	# Inj.	# Surv.	# DsRed+	Phenotype
pMi(gm)	promoter	N/A – *Parhyale* muscle reporter PhMS	109	39	24	Green muscle expression
pMi(retn)	promoter	*retained* – expressed in some neurons	240	104	5	No consistent phenotype
pMi(SVP)	promoter	*seven up* – expressed in ventral neuroblasts	122	31	0	None
pMi(su-s-2)	pro + peak	*suppressor of sable* – embryonic ubiquitous expression	127	57	7	No consistent phenotype
pMi(Rx)	pro + peak	*retinal homeobox* – expressed in various neural tissues	74	28	0	None
pMi(Hsc70-4)	promoter	*Heat shock 70 complex protein 4* – embryonic ubiquitous	212	99	22	Neurons associated with the eye
pMi(En1)	promoter	*engrailed* – parasegment stripes	71	33	0	None
pMi(En1)-p1	pro + peak	‘’	67	46	0	None
pMi(En1)-p1.5	pro + peak	‘’	70	51	0	None
pMi(En1)-p5	pro + peak	‘’	68	58	1	Faint expression domain
pMi(Sp69)	promoter	*Sp69* – leg gap gene	74	56	0	None
pMi(Sp69)-p1	pro + peak	‘’	50	30	0	None
pMi(Sp69)-p1.5	pro + peak	‘’	86	32	3	No consistent phenotype
pMi(Sp69)-p2.5	pro + peak	‘’	71	35	0	None
pMi(Sp69)-p3	pro + peak	‘’	241	121	8	Head, leg, gill
pMi(ne1)-Hsp70-p1	pro + peak	*Hsp70 complex*	45	26	0	None
pMi(ne1)-Hsp70-p2	pro + peak	*Hsp70 complex*	310	81	61	Yolk cells

1.3 – Methods, Chapter 1

Crustacean Cultures

Parhyale hawaiensis were raised at 25°C and fed a diet of carrots, shrimp pellets, and Spirulina flakes in Tropic Marin artificial seawater with a salinity between 31–35ppm in plastic tanks.

Embryo Cultures

Embryos were collected using previously described protocols (Rehm et al., 2009), staged using the Browne et al. 2003 staging guide, and raised at 27°C in a humidity-controlled incubator in filter-sterilized ASW (FASW). For ATAC-Seq and RNA-Seq experiments, clutches of 15 or more embryos were collected and staged between S1–S6, as these timepoints are among the shortest and most morphologically identifiable during early development.

Embryo Staging

Two groups of five embryos each from a single clutch were used in the ATAC-Seq experiments, and three groups of five embryos each from a single clutch were used in the RNA-Seq experiments. Embryos of the correct stage were selected based on morphological characteristics as described in the Browne et al. staging guide, and any embryos with abnormal or asynchronous morphology were discarded. Morphologically representative embryos from the same clutch as used for the ATAC-Seq were photographed on ventral and lateral positions within 10 minutes of beginning the tagmentation procedure. Any remaining embryos were boiled briefly and fixed using the protocol described in (Rehm et al., 2009) and DAPI stained to further confirm staging. The process of staging embryos, imaging embryos, boiling, and mashing embryos for Omni-ATAC tagmentation were all performed within a 30 minute time interval for each developmental stage.

RNA Sequencing

Embryos for RNA isolation were homogenized in TriZol using a DWK Life Sciences (Kimble) Biomasher II Closed System Disposable Tissue Homogenizer, and RNA was isolated using the Zymo Direct-Zol Miniprep Plus kit. RNA quality was assessed based on fragment analysis using a Bioanalyzer. cDNA was generated from RNA using the TaKaRa SMART-Seq v4 Ultra Low-Input kit. cDNA was then sequenced using the Illumina NovaSeq to generate short reads. Using the Nanopore Direct cDNA barcoding kit (SQK-LSK109), cDNA was also sequenced on a Nanopore MinION flow cell to generate long reads.

Limb development RNA Sequencing Methods

A pool of embryos consisting of embryonic stages 19, 20, 22, 23, 25, and 28 were homogenized with DWK Life Sciences Kimble Kontes Pellet Pestle Cordless Motor in Trizol and extracted using Trizol; PolyA+ libraries were prepared with the Truseq V1 kit (Illumina), starting with 0.6–3.5 mg of total mRNA, and sequenced on the Illumina HiSeq 2000 as paired-end 100 base reads, at the QB3 Vincent J. Coates Genomics Sequencing Laboratory.

Limb development RNA-Seq *de novo* Transcriptome assembly

For the Trinity-limb (old) assembly, transcripts were assembled *de novo* using Trinity r2013_08_14 (Grabherr et al., 2011)with parameters --JM 170G --CPU 10 --inchworm_cpu 6 --min_kmer_cov 2 --min_contig_length 49 --group_pairs_distance 700. Input RNA-Seq reads were treated as paired-ends. Output transcript assemblies shorter than 200 bp were discarded. The remaining assemblies were screened for contaminants with BLASTX (BLAST+ v2.2.26; parameters: -num_descriptions 50 -num_alignments 50 -evalue 1e-5 -lcase_masking -soft_masking true -seg yes)(Camacho et al., 2009) against a database of all bacterial proteins downloaded from NCBI (retrieved 2013-08-31), then against Swiss-Prot UniProt human sequences (retrieved 2013-08-31)(The UniProt Consortium, 2019). All hits were further required to cover a minimum of 40% of each assembled transcript sequence. When filtering for human contaminants, a 98% identity threshold was also required. For the Trinity-limb (new) assembly, transcripts were assembled using Trinity v2.5.1 using standard settings.

Transcriptome Assemblies

Short-read RNA-Seq data was used to generate both *de novo* and genome-guided transcriptomes using Trinity. For genome-guided assembly, reads were mapped using HISAT2. Long-read RNA-Seq data was mapped to the most recent *Parhyale* genome (phaw_5.0) using minimap2 and assembled using StringTie2. A combined transcriptome using both HISAT2-mapped short reads and minimap2-mapped long reads was also generated using StringTie2. See Supplementary Methods for additional information about assembly parameters.

Mikado Transcriptome

Short-read Trinity transcriptomes and the transcriptome from the Kao et al. manuscript were mapped to the phaw_5.0 genome using GMAP. Short-read transcriptomes and long-read StringTie2 transcriptomes were merged using the Mikado software along with a previous genome annotation generated by Leo Blondel (using MAKER). See supplementary methods for additional information about Mikado parameters.

Functional Annotation of Transcripts

The merged Mikado combined transcriptome was annotated using eggNOG and an Orthofinder comparison to *Drosophila*. The MAKER genome annotation was also annotated using eggNOG and OrthoFinder. See Supplementary Methods for additional information about functional annotation.

ATAC-Seq Data

We performed conventional ATAC-Seq as per Buenrostro et al. 2013 for stages S13–S22. Data were used for benchmarking in comparison to Omni-ATAC data, but were not used for downstream analyses.

Omni-ATAC-Seq Tagmentation

Tagmentation was performed using the reagents described in the Corces et al. 2017 Omni-ATAC-Seq paper, with the following modifications. Instead of the Illumina Nextera TD buffer, 2x Tagmentation

Buffer from Wang et al. 2013 was used (Corces et al., 2017; Wang et al., 2013). Homemade Tn5 enzymes were purified and received as a gift from Jase Gehring and assembled into Tn5 transposomes as per Picelli et al 2014, using adapters purchased from IDT with 5' phosphorylation (Picelli et al., 2014). Prior to adding RSB+D, embryos were washed 3x using 1X PBS. Embryos were mashed into a near-uniform solution using the tip of a low-retention p10 pipette, and the pipette was visually inspected for any remaining debris. We used the QIAGEN MinElute kit for purification and concentration of DNA.

Omni-ATAC-Seq Library Preparation

Saturation PCR conditions for Omni-ATAC-Seq libraries were performed using a Roche Lightcycler 480 as per the Corces et al. 2017 paper. Optimal conditions for additional amplification after pre-amplification were determined based on the number of additional PCR cycles corresponding to one-third of the maximum value, rounded up. Barcodes were added to samples using primers supplied by QB3-Berkeley using PCR as described in the Corces et al. 2017 Omni-ATAC-Seq paper. Following amplification, adapters were removed from libraries using Ampure bead purification and analyzed using a Bioanalyzer or Fragment Analyzer machine to assess library quality. Final libraries were pooled together to have relatively equal proportions based on additional qPCR quantification and size-selected using a Pippin Prep to remove fragments greater than 1.5kb in size. All libraries were sequenced on one lane of each of Illumina NovaSeq SP 150PE and NovaSeq S1 150PE. Adapters used for libraries are listed in Supplementary Table 1.

Omni-ATAC Sequencing Quality Control

Adapter trimming was performed using Trim Galore, which leverages Cutadapt and FASTQC (Andrews, 2010; Kreuger, 2017; Martin, 2011). FASTQC was also performed before and after adapter trimming to confirm removal of sequencing and Tn5 adapters. Percentage of reads remaining after deduplication was estimated based on FASTQC metrics. Library size was estimated by multiplying raw read count by percentage of reads remaining after deduplication for each lane, and then summing the two lanes. Reads were aligned to the phaw_5.0 genome as well as the *Parhyale* chrM using Bowtie2, and percentage read mapping was determined using Bowtie2 output (Langmead and Salzberg, 2012). After peak calling, merged peaks were used to evaluate correlation between replicates. Replicates were highly correlated, with a mean Pearson correlation of 0.988 and a mean Spearman correlation of 0.936 (Supp. Fig. 2 H, I).

Omni-ATAC Peak Calling

Aligned reads were quality-filtered with a Q value of 10 in samtools and used for Genrich analysis. Genrich analysis was run on both duplicate libraries simultaneously; Genrich performs peak calling on each peak individually, and then merges the p-values of the replicates using Fisher's method to generate a q-value, obviating the need to calculate an Irreproducible Discovery Rate (IDR). Output .bedgraph-like files from Genrich were reformatted using a custom Python script to be standard .bedgraph files, which were converted to bigWig files using the bedGraphToBigWig executable from

the UCSC Genome Tools software bundle. Output .narrowPeak files from Genrich were converted into .bed files for ease of visualization in IGV using a custom Python script.

Assigning spatial categories to Omni-ATAC peaks

We first assigned the nearest peak within 5kb of the first 200bp of each gene (mRNA and ncRNA) as a promoter peak using bedtools closest. We then assigned the remaining peaks into categories based on their position relative to mRNA and ncRNA annotations. Peaks that overlapped with mRNA and ncRNA annotations were assigned as exonic or intronic regulatory elements. The remaining peaks—those which had not been classified as promoters, and which did not overlap with genes—were classified as intergenic peaks. The intergenic peaks were divided into two categories: proximal and distal intergenic peaks. Proximal peaks were those less than 10kb away from the nearest gene, while distal intergenic peaks were those greater than 10kb away from the nearest gene.

Differential Expression and Accessibility Analyses

We performed differential expression analysis using ImpulseDE2 and DESeq2 using standard settings. To generate the Omni-ATAC-Seq read count matrix, we used bedtools multicov, using the merged Genrich peaks as our regions of interest. For our ImpulseDE2 analyses using the Omni-ATAC-Seq data, ImpulseDE2 model fits were extracted from the ImpulseDE2 output and used to visualize model expression using a custom Python script (see Supplementary Methods). To generate the RNA-Seq read count matrix for DESeq2, we generated a gene_trans_map file for the Mikado transcriptome, as would be available for the Trinity RNA-Seq analysis pipeline, and used the built-in Trinity differential expression pipeline (align_and_estimate_abundance.pl, abundance_estimates_to_matrix.pl) with Kallisto to generate a matrix of read counts. To comply with the requirement for integer counts in DESeq2 analysis, we rounded each value to the nearest whole number.

NucleoATAC Nucleosome Predictions

Quality-filtered reads from each biological duplicate were merged and analyzed using NucleoATAC, with genomic regions set as +/- 500bp windows around Genrich peaks.

HINT-ATAC Transcription Factor Footprinting

Quality-filtered reads from each biological duplicate were merged and analyzed using rgt-hint footprinting. We used the JASPAR2020 database and converted the position weight matrices from JASPAR format into a simple matrix format expected by RGT-HINT using the R package "universalmotif," and generated a ".mtf" file to store database information (see Supplementary Methods). For enrichment analyses, we used bedtools random to generate 13 million random 20bp sequences, as this was the average footprint size of genuine footprints detected by RGT-HINT in our data. This set of random sequences was used as background for our enrichment analyses. For cluster-specific enrichment analyses, we collated all unique transcription factor footprints from all developmental stages for each cluster (e.g. all footprints across S13, S14, etc. for all peaks in a given cluster) and compared enrichment levels to our randomly generated background.

Minos Transposon Cloning

Minos transposon reporter plasmids were cloned using Gibson homology-mediated cloning approaches and the NEB Gibson Assembly or NEBuilder kits. As a base plasmid, we used the pMi(ne1) plasmid, which contains the Hsp70 minimal promoter, a DsRed protein sequence, and an SV40 3'UTR sequence, as well as two Minos inverted repeats. For plasmids containing the Hsp70 minimal promoter, the insert was integrated between the EcoRV and BglII restriction sites. For plasmids containing an endogenous promoter, the insert replaced the sequence between the EcoRV and NcoI restriction sites, thereby removing the Hsp70 minimal promoter. Cloned plasmids were Sanger sequenced to confirm a full DsRed ORF and inclusion of desired genomic sequences.

Minos Transposase Assay

Minos transposase mRNA was generated using the ThermoFisher mMESSAGE mMACHINE T7 or T7 ULTRA kit using NotI-digested pBlueSK-MimRNA (Addgene #102535). mRNA and concentrated DNA were mixed into a final concentration of 1 μg/μL in a solution of 0.1% phenol red in nuclease-free water. 1- and 2-cell *Parhyale* embryos were injected with approximately 3-5 picoliters of injection mix using a borosilicate glass capillary needle pulled using a Sutter P-80 or P-85 instrument. Embryos were raised until hatching and examined once per day from 3dpf–10dpf using a Zeiss LSM780 confocal microscope to screen for DsRed fluorescence.

Data Availability

Omni-ATAC-Seq raw reads will be available at SRA, Bioproject: PRJNA765106. RNA-Seq raw reads will be available at SRA. Omni-ATAC and RNA-Seq downstream analyses will be available at GEO by time of future publication. Downstream analysis files for Omni-ATAC-Seq will include BigWig files for visualizing Omni-ATAC read pileups; stage-specific and combined Genrich peaks; NucleoATAC signal, occupancy, insert, v-plots, and other miscellaneous data; HINT-ATAC transcription factor footprinting predictions and enrichment statistics; JASPAR2020 converted database and .mtf file for HINT-ATAC analyses; and differential accessibility analysis tables. Downstream analysis files for RNA-Seq will include BigWig files for visualizing RNA-Seq read pileups; Nanopore alignment sequences; individual transcriptome GFF and FASTA files; eggNOG annotations and OrthoFinder annotations of transcripts; and RNA-Seq differential expression analysis tables. In addition, the GEO database will contain the updated Mikado genome annotation and assignment of peaks based on their position relative to genes, along with classification data for peaks as concordant or discordant in expression and accessibility. Code for analysis and visualizations will be curated and deposited in a GitHub repository prior to publication.

Contributions

D.A.S. conceived of, designed, and performed experiments and analyses. D.A.S. wrote and revised the manuscript. H.S.B. collected sequencing data for the limb development transcriptome and J.V.B. assembled it. N.H.P. provided guidance and supervised the project. J.V.B., N.H.P., and H.S.B. and provided feedback on the manuscript.

Acknowledgments
We thank Jase Gehring and Jenna Haines for providing tips and reagents for performing and troubleshooting ATAC-Seq, and Kasia Oktaba for suggesting the technique. Jenna Haines and Shaked Afik also provided helpful programming references for performing ATAC-Seq analyses. We are grateful to Aaron Pomerantz for helping to troubleshoot and perform Nanopore sequencing, and for tips on RNA-Seq analysis. This work would have been much more challenging without fantastic sequencing support from the QB3 Biosciences Functional Genomics Lab, particularly Shana McDevitt, Karen Lundy, and Justin Choi. We would also like to thank Gideon Harianja, Lauren Zane, and Thienkim Ho for their experimental support on early stages of this project that were ultimately not included in the manuscript. We are also grateful to Dan Rokhsar for helpful comments on the data analysis and manuscript. Thank you to the members of the Patel Lab, Craig Miller's lab, Iswar Hariharan's lab, and Hernan Garcia's lab for providing feedback and suggestions to the manuscript, particularly Craig Miller, Tyler Square, and Brandon Schlomann.

Funding
DAS was supported by the NSF Division of Graduate Education, GRFP 2016230010. This work was funded by an NSF Division of Integrative Organismal Systems Grant 1257379 to NHP.

Conflicts of Interest
The authors declare no conflicts of interest.

1.4 – References, Chapter 1

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data.

Borsari, B., Villegas-Mirón, P., Pérez-Lluch, S., Turpin, I., Laayouni, H., Segarra-Casas, A., Bertranpetit, J., Guigó, R., and Acosta, S. (2021). Enhancers with tissue-specific activity are enriched in intronic regions. Genome Res. *31*, 1325–1336.

Brogaard, K., Xi, L., Wang, J.-P., and Widom, J. (2012). A map of nucleosome positions in yeast at base-pair resolution. Nature *486*, 496–501.

Browne, W.E., Price, A.L., Gerberding, M., and Patel, N.H. (2005). Stages of embryonic development in the amphipod crustacean, Parhyale hawaiensis. Genesis *42*, 124–149.

Bruce, H.S., and Patel, N.H. (2020). Knockout of crustacean leg patterning genes suggests that insect wings and body walls evolved from ancient leg segments. Nat. Ecol. Evol. *4*, 1703–1712.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol *109*, 21.29.1-9.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

Cazet, J.F., Cho, A., and Juliano, C.E. (2021). Generic injuries are sufficient to induce ectopic Wnt organizers in Hydra. ELife *10*, e60562.

Chechik, G., and Koller, D. (2009). Timing of Gene Expression Responses to Environmental Changes. J. Comput. Biol. *16*, 279–290.

Christian, S.-E., Schultheis, D., Schwirz, J., Ströhlein, N., Troelenberg, N., Majumdar, U., Dao, V., Grossmann, D., Richter, T., Tech, M., et al. (2015). The iBeetle large-scale RNAi screen reveals gene functions for insect development and physiology. 7822.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962.

Costa-Silva, J., Domingues, D., and Lopes, F.M. (2017). RNA-Seq differential expression analysis: An extended review and a software tool. PLOS ONE *12*, e0190152.

Crawford, K., Diaz Quiroz, J.F., Koenig, K.M., Ahuja, N., Albertin, C.B., and Rosenthal, J.J.C. (2020). Highly Efficient Knockout of a Squid Pigmentation Gene. Curr. Biol. *30*, 3484-3490.e4.

Duan, Y., Zhang, W., Cheng, Y., Shi, M., and Xia, X.-Q. (2020). A systematic evaluation of bioinformatics tools for identification of long noncoding RNAs. RNA.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238.

Fischer, D.S., Theis, F.J., and Yosef, N. (2018). Impulse model-based differential expression analysis of time course sequencing data. Nucleic Acids Res. *46*, e119–e119.

Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D., et al. (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *48*, D87–D92.

Gaspar, J.M. Genrich.

Gatzmann, F., Falckenhayn, C., Gutekunst, J., Hanna, K., Raddatz, G., Carneiro, V.C., and Lyko, F. (2018). The methylome of the marbled crayfish links gene body methylation to stable expression of poorly accessible genes. Epigenetics Chromatin *11*, 57.

Gehrke, A.R., Neverett, E., Luo, Y.-J., Brandt, A., Ricci, L., Hulett, R.E., Gompers, A., Ruby, J.G., Rokhsar, D.S., Reddien, P.W., et al. (2019). Acoel genome reveals the regulatory landscape of whole-body regeneration. Science *363*, eaau6173.

Gilbert, D.G. (2019). Longest protein, longest transcript or most expression, for accurate gene reconstruction of transcriptomes? BioRxiv 829184.

Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R., and Lieb, J.D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res. *17*, 877–885.

Gisselbrecht, S.S., Palagi, A., Kurland, J.V., Rogers, J.M., Ozadam, H., Zhan, Y., Dekker, J., and Bulyk, M.L. (2020). Transcriptional Silencers in Drosophila Serve a Dual Role as Transcriptional Enhancers in Alternate Cellular Contexts. Mol. Cell *77*, 324-337.e8.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol *29*, 644–652.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc *8*, 1494–1512.

Halfon, M.S. (2020). Silencers, Enhancers, and the Multifunctional Regulatory Genome. Trends Genet. TIG *36*, 149–151.

Hinton, G.E., and Roweis, S. (2003). Stochastic Neighbor Embedding. In Advances in Neural Information Processing Systems, S. Becker, S. Thrun, and K. Obermayer, eds. (MIT Press), p.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. *47*, D309–D314.

JL, and G.-S., and Shubin, N. (2015). Deep conservation of wrist and digit enhancers in fish. Proc. ....

Kao, D., Lai, A.G., Stamataki, E., Rosic, S., Konstantinides, N., Jarvis, E., Donfrancesco, A., Natalia, P.-S., Semon, M., Grillo, M., et al. (2016). The genome of the crustacean Parhyale hawaiensis, a model for animal development, regeneration, immunity and lignocellulose digestion. Elife *5*, 065789.

Kissane, S., Dhandapani, V., and Orsini, L. (2021). Protocol for assay of transposase accessible chromatin sequencing in non-model species. STAR Protoc. *2*, 100341.

Konstantinides, N., and Averof, M. (2014). A common cellular basis for muscle regeneration in arthropods and vertebrates. Science *343*, 788–791.

Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi, R., Salzberg, S.L., and Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. *20*, 278.

Kreuger, F. (2017). TrimGalore.

Kumar, L., and E Futschik, M. (2007). Mfuzz: a software package for soft clustering of microarray data. Bioinformation *2*, 5–7.

Lai, Y.-T., Deem, K.D., Borràs-Castells, F., Sambrani, N., Rudolf, H., Suryamohan, K., El-Sherif, E., Halfon, M.S., McKay, D.J., and Tomoyasu, Y. (2018). Enhancer identification and activity evaluation in the red flour beetle, Tribolium castaneum. Development *145*.

Langmead, B., and Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods.

Legendre, P., and Legendre, L. (2012). Numerical ecology (Elsevier).

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100.

Li, Y., Chen, C., Kaye, A.M., and Wasserman, W.W. (2015). The identification of cis-regulatory elements: A review from a machine learning perspective. Biosystems *138*, 6–17.

Li, Z., Schulz, M.H., Look, T., Begemann, M., Zenke, M., and Costa, I.G. (2019). Identification of transcription factor binding sites using ATAC-seq. Genome Biol. *20*, 45.

Liu, B., Li, J., and Cairns, M.J. (2014). Identifying miRNAs, targets and functions. Brief. Bioinform. *15*, 1–19.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Ludwig, M., Patel, N., and Kreitman, M. (1998). Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change. Development.

Maaten, L. van der, and Hinton, G. (2008). Visualizing Data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

Mahony, S., and Benos, P.V. (2007). STAMP: a web tool for exploring DNA-binding motif similarities. Nucleic Acids Res. *35*, W253–W258.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal Vol 17 No 1 Gener. Seq. Data Anal. - 1014806ej171200.

Martin, A., Serano, J.M., Jarvis, E., Bruce, H.S., Wang, J., Ray, S., Barker, C.A., C, O., Liam, and Patel, N.H. (2015). CRISPR/Cas9 Mutagenesis Reveals Versatile Roles of Hox Genes in Crustacean Limb Specification and Evolution. Curr Biol *26*, 14–26.

Mattioli, K., Oliveros, W., Gerhardinger, C., Andergassen, D., Maass, P.G., Rinn, J.L., and Melé, M. (2020). Cis and trans effects differentially contribute to the evolution of promoters and enhancers. Genome Biol. *21*, 210.

Mito, T., Nakamura, T., Bando, T., Ohuchi, H., and Noji, S. (2011). The advent of RNA interference in Entomology. Entomol. Sci. *14*, 1–8.

Morton James T., Toran Liam, Edlund Anna, Metcalf Jessica L., Lauber Christian, Knight Rob, and Jansson Janet K. Uncovering the Horseshoe Effect in Microbial Analyses. MSystems *2*, e00166-16.

Ofran, Y., Mysore, V., and Rost, B. (2007). Prediction of DNA-binding residues from sequence. Bioinformatics *23*, i347–i353.

Paris, M., Wolff, C., Patel, N.H., and Averof, M. (2021). The Crustacean Model Parhyale hawaiensis. Preprints.

Pavlopoulos, A., and Averof, M. (2005). Establishing genetic transformation for comparative developmental studies in the crustacean Parhyale hawaiensis. Proc Natl Acad Sci USA *102*, 7888–7893.

Pavlopoulos, A., Kontarakis, Z., Liubicich, D.M., Serano, J.M., Akam, M., Patel, N.H., and Averof, M. (2009). Probing the evolution of appendage specialization by Hox gene misexpression in an emerging model crustacean. Proc Natl Acad Sci USA *106*, 13897–13902.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

Pérez-Zamorano, B., Rosas-Madrigal, S., Lozano, O.A.M., Castillo Méndez, M., and Valverde-Garduño, V. (2017). Identification of cis-regulatory sequences reveals potential participation of lola and Deaf1 transcription factors in Anopheles gambiae innate immune response. PloS One *12*, e0186435–e0186435.

Picelli, S., Björklund, Å.K., Reinius, B., Sagasser, S., Winberg, G., and Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively-scaled sequencing projects. Genome Res *24*, gr.177881.114.

Podani, J., and Miklós, I. (2002). Resemblance coefficients and the horseshoe effect in principal coordinates analysis. Ecology *83*, 3331–3343.

Poynton, H.C., Hasenbein, S., Benoit, J.B., Sepulveda, M.S., Poelchau, M.F., Hughes, D.S.T., Murali, S.C., Chen, S., Glastad, K.M., Goodisman, M.A.D., et al. (2018). The Toxicogenome of Hyalella azteca: A Model for Sediment Ecotoxicology and Evolutionary Toxicology. Environ. Sci. Technol. *52*, 6009–6022.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Radman-Livaja, M., and Rando, O.J. (2010). Nucleosome positioning: how is it established, and why does it matter? Dev. Biol. *339*, 258–266.

Ramos, A.P., Gustafsson, O., Labert, N., Salecker, I., Nilsson, D.-E., and Averof, M. (2019). Analysis of the genetically tractable crustacean Parhyale hawaiensis reveals the organisation of a sensory system for low-resolution vision. BMC Biol. *17*, 67.

Rasys, A.M., Park, S., Ball, R.E., Alcala, A.J., Lauderdale, J.D., and Menke, D.B. (2019). CRISPR-Cas9 Gene Editing in Lizards through Microinjection of Unfertilized Oocytes. Cell Rep. *28*, 2288-2292.e3.

Ratnere, I., and Dubchak, I. (2009). Obtaining Comparative Genomic Data with the VISTA Family of Computational Tools. Curr. Protoc. Bioinforma. *26*, 10.6.1-10.6.17.

Rehm, E.J., Hannibal, R.L., Chaw, R.C., Vargas-Vila, M.A., and Patel, N.H. (2009). Fixation and Dissection of Parhyale hawaiensis Embryos. Cold Spring Harb. Protoc. *2009*, pdb.prot5127.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinforma. Oxf. Engl. *26*, 139–140.

Rouhana, L., Weiss, J.A., Forsthoefel, D.J., Lee, H., King, R.S., Inoue, T., Shibata, N., Agata, K., and Newmark, P.A. (2013). RNA interference by feeding in vitro-synthesized double-stranded RNA to planarians: methodology and dynamics. Dev. Dyn. Off. Publ. Am. Assoc. Anat. *242*, 718–730.

Schep, A.N., Buenrostro, J.D., Denny, S.K., Schwartz, K., Sherlock, G., and Greenleaf, W.J. (2015). Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. Genome Res.

Serano, J.M., Martin, A., Liubicich, D.M., Jarvis, E., Bruce, H.S., La, K., Browne, W.E., Grimwood, J., and Patel, N.H. (2016). Comprehensive analysis of Hox gene expression in the amphipod crustacean Parhyale hawaiensis. Dev Biol *409*, 297–309.

Sharma, P.P., Schwager, E.E., Giribet, G., Jockusch, E.L., and Extavour, C.G. (2013). Distal-less and dachshund pattern both plesiomorphic and apomorphic structures in chelicerates: RNA interference in the harvestman Phalangium opilio (Opiliones). Evol. Dev. *15*, 228–242.

Signor, S.A., and Nuzhdin, S.V. (2018). The Evolution of Gene Expression in cis and trans. Trends Genet. *34*, 532–544.

Srivastava, M., L, M.-C., Kathleen, van Wolfswinkel, J.C., and Reddien, P.W. (2014). Whole-Body Acoel Regeneration Is Controlled by Wnt and Bmp-Admp Signaling. Curr Biol.

Stamataki, E., and Pavlopoulos, A. (2016). Non-insect crustacean models in developmental genetics including an encomium to Parhyale hawaiensis. Curr Opin Genet Dev *39*, 149–156.

Sun, D.A., and Patel, N.H. (2019). The amphipod crustacean Parhyale hawaiensis: An emerging comparative model of arthropod development, evolution, and regeneration. WIREs Dev. Biol. *8*, e355.

The UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. *47*, D506–D515.

Venturini, L., Caim, S., Kaithakottil, G.G., Mapleson, D.L., and Swarbreck, D. (2018). Leveraging multiple transcriptome assembly methods for improved gene structure annotation. GigaScience *7*.

Wang, L., and Brown, S.J. (2006). BindN: a web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. Nucleic Acids Res. *34*, W243–W248.

Wang, Q., Gu, L., Adey, A., Radlwimmer, B., Wang, W., Hovestadt, V., Bähr, M., Wolf, S., Shendure, J., Eils, R., et al. (2013). Tagmentation-based whole-genome bisulfite sequencing. Nat Protoc *8*, 2022–2032.

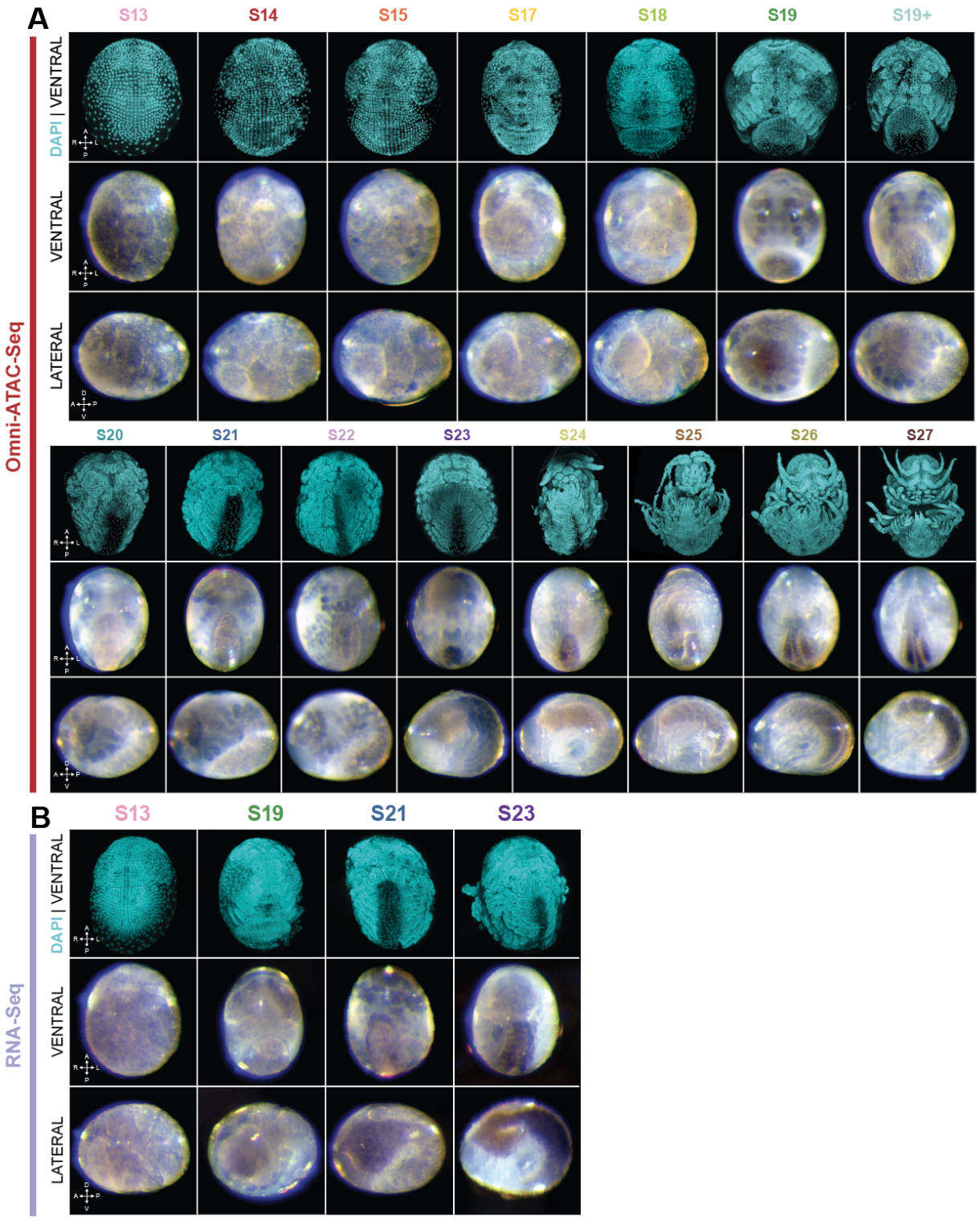
Wasserman, W.W., and Sandelin, A. (2004). Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet *5*, 276–287.

Wetterstramd, K. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP).

Wittkopp, P.J., Haerum, B.K., and Clark, A.G. (2004). Evolutionary changes in cis and trans gene regulation. Nature *430*, 85–88.

Yan, F., Powell, D.R., Curtis, D.J., and Wong, N.C. (2020). From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. Genome Biol. *21*, 22.

Yosef, N., and Regev, A. (2011). Impulse Control: Temporal Dynamics in Gene Transcription. Cell *144*, 886–896.

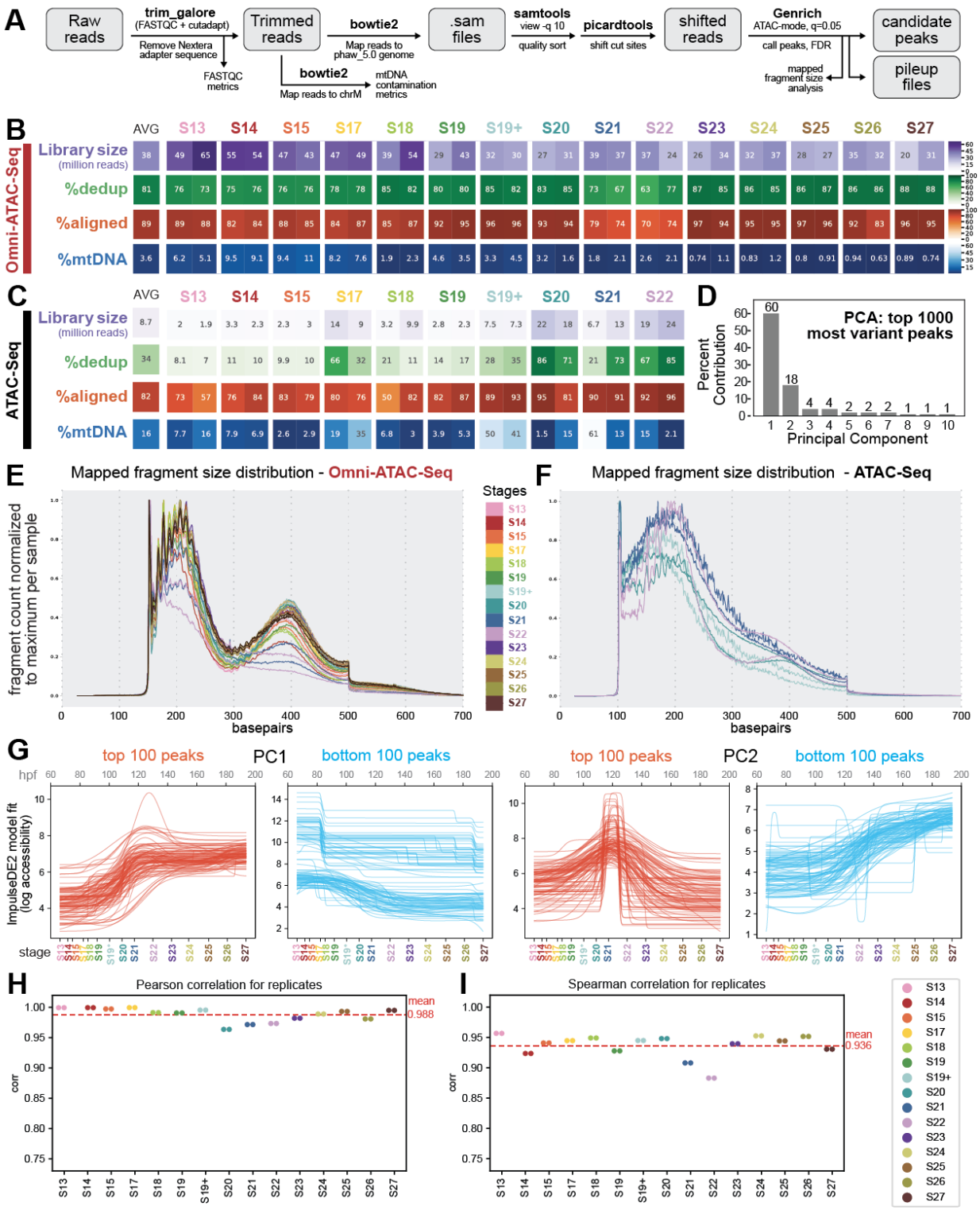


Supp. Fig. 1.1: Embryo images for time course Omni-ATAC-Seq and RNA-Seq

A) Representative embryo images from Omni-ATAC-Seq libraries. Top row shows ventral view of DAPI stained embryo. Bottom rows show ventral and lateral brightfield images of embryos shortly before tagmentation.

B) Representative embryo images from RNA-Seq libraries. Top row shows ventral view of DAPI stained embryo. Bottom rows show ventral and lateral brightfield images of embryos shortly before RNA extraction.

(A: anterior, P: posterior, L: left, R: right; D: dorsal, V: ventral).



Supp. Fig. 1.2.1: Omni-ATAC-Seq analysis pipeline and quality control metrics

A) Analysis pipeline for generating Omni-ATAC peaks and QC metrics from raw reads. Reads were quality and adapter-trimmed using trim_galore, and trimmed reads were mapped to either the entire genome or to only the mitochondrial chromosome (ChrM) to generate a conservative estimate of mtDNA contamination. Mapped reads were further filtered using samtools (Q = 10) and read ends were shifted using Picardtools to reflect the mechanism of Tn5

insertion. Shifted reads were passed to Genrich in ATAC mode using a q-value cutoff of 0.05. Files produced by Genrich included stage-specific peaks and bedgraph-like read pileup formats.

B) Quality control metrics for Omni-ATAC-Seq libraries. Library size, percentage of reads remaining after deduplication (%dedup), percentage of aligned reads (%aligned), and percentage of reads mapping to the mitochondrial genome (%mtDNA) are displayed for individual Omni-ATAC libraries.

C) Metrics from B visualized for conventional ATAC-Seq libraries. Overall, Omni-ATAC seq shows similar or improved performance on all metrics.

D) Percentage contribution to variance of each of the top 10 principal components from PCA in Fig. 2C.

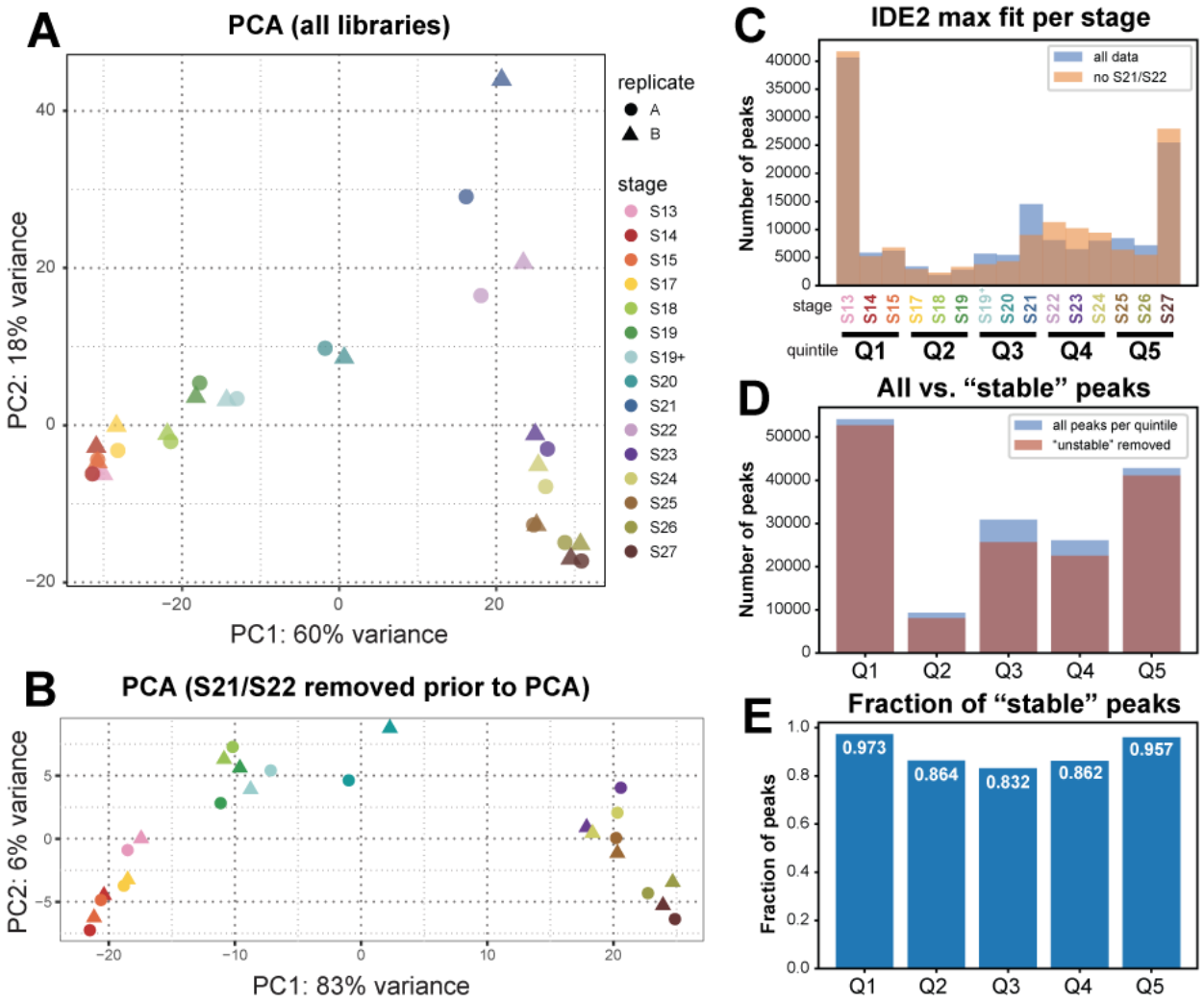
E) Mapped fragment size distribution of Omni-ATAC-Seq reads. Two clear peaks are seen in the fragment size distribution, reflecting proper sub-nucleosomal insertions.

F) Mapped fragment size distribution of conventional ATAC-Seq reads for 4 developmental stages: S19, S19+, S20, S21. Fragment size peaks are less distinct as compared to Omni-ATAC-Seq libraries.

G) Plots of ImpulseDE2 model fits to the top 100 and bottom 100 peaks loading for PC1 and PC2 in figure 1C. PC1 appears to be associated with developmental time, while PC2 appears to be associated with peaks that show an increase in accessibility in the middle of development.

H) Pearson correlation of mapped read counts in merged peaks between replicates for each developmental stage. Replicate libraries show a high average correlation value (0.988).

I) Spearman correlation of mapped read counts in merged peaks between replicates for each developmental stage. Replicate libraries show a high average correlation value (0.936). A lower correlation coefficient compared to Pearson may suggest that strong outliers contribute to the high Pearson correlation measure.



Supp. Fig. 1.2.2: Examining outlier effects of S21/S22 samples

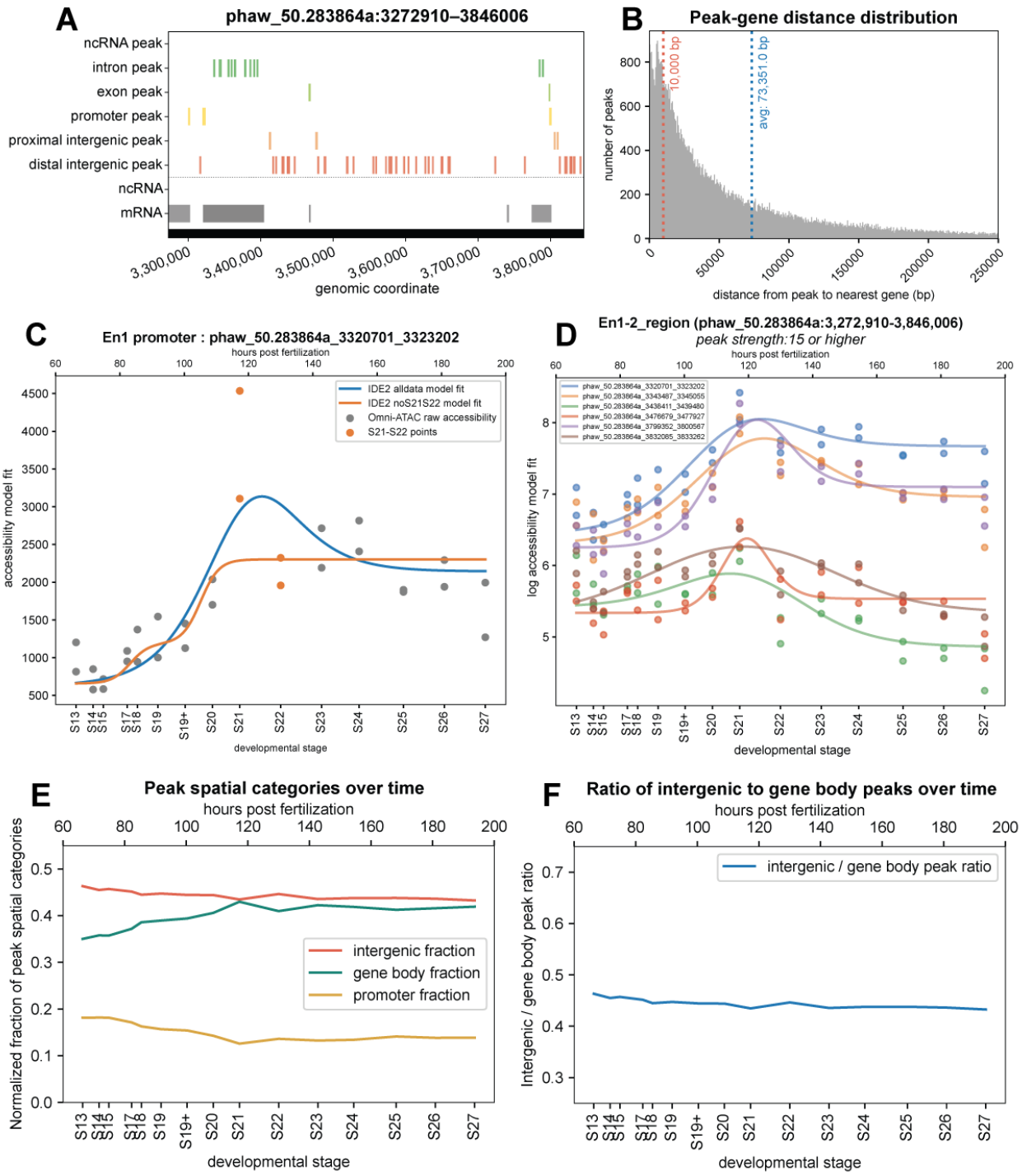
A) Principal Components Analysis (PCA) of all libraries. Samples S21 and S22 appear highly separated in PC2 relative to other samples.

B) PCA of all libraries except S21 and S22 samples. The relative position of libraries in PC2 appears largely unchanged, with early and late time points showing negative loading in PC2, and middle timepoints showing higher loading in PC2. This suggests PC2 is not driven entirely by differences in the S21/ S22 samples.

C) ImpulseDE2 model max fit for all data (alldata, blue) or data excluding S21-S22 (noS21S22, orange). For each peak, the developmental time of maximum accessibility was estimated, and the results are summarized as a histogram. In the absence of S21-S22 samples, the same overall trend of early, middle, and late peak maximum fit is observed.

D) Peak stability analysis. The 15 developmental stages were broken into 5 quintiles of 3 stages each (label in C). For each peak, the IDE2 model max fit stage was calculated for alldata or noS21S22, and the shift in the model max fit was evaluated. Peaks that moved more than 3 developmental stages (equivalent to 1 quintile shift) were considered "unstable". The majority of peaks from each quintile were considered "stable" in the absence of S21/S22 data.

E) Fraction of stable peaks from analysis in (D). For each quintile, >80% of peaks were considered "stable", suggesting that most peaks have the same trajectory regardless of the presence of S21/S22 data. This suggests that the remaining data are sufficient to reproduce the general accessibility trajectory of most peaks.



Supp. Fig. 1.2.3: Peak category analyses and data visualization tools

A) Visualization of the genomic region around the *Parhyale Engrailed-1* and *Engrailed-2* loci using the Region Type Viewer found in the ParhyaleATACExplorer Jupyter notebook. Peak types are colored based on their identity. A 1000bp padding has been added to the 5' and 3' widths of each element, for ease of visualization.

B) Distribution of peak-nearest gene distances for intergenic peaks. Most peaks are "distal intergenic", or >10kb away (red dotted line) and could not have been identified easily using previous Minos reporter techniques. The average distance of intergenic peaks from the nearest gene is 73,351bp (blue dotted line).

C) Visualization of ImpulseDE2 model fits for all data (blue line) or noS21S22 data (orange line), plotted alongside raw accessibility values (points; stage S21 and S22 points are highlighted in orange), shown for the promoter peak of the *Engrailed-1* gene. Using the ImpulseDE2 Model Comparisons section of the ParhyaleATACExplorer Jupyter notebook, users can visualize the two model fits for arbitrary peaks of interest.

D) Visualization of ImpulseDE2 model fits for all data within a selected genomic region (same as shown in A), filtered by "peak strength" = 15, or the number of peaks found within a merged peak, summed over time. IDE2 model fits can be visualized for all peaks within an arbitrary genomic address, or a specific set of labeled peaks when provided as a Python dictionary.

E) Fraction of intergenic, gene body, and promoter peaks across developmental time, normalized for the number of peaks after removing "unknown" peaks. The fraction of intergenic and promoter peaks appears to slightly decline, while the fraction of gene body peaks increases over time.

F) Ratio of gene body peaks to intergenic peaks over time. The ratio appears to decline very slightly over time.


A
~28,300,000*
Parhyale Hox cluster (phaw_50.282695a:24,500,000-28,300,000)
~24,500,000*
*distances not to scale
RACE
lab complete
pb 2 fragments
Hox3 Homeodomain only
Dfd complete
Scr complete
Antp complete
Ubx-i1 complete
abd-A-i1 complete
Abd-B-i1 complete
Ubx-i2 complete
abd-A-i2 complete
Abd-B-i2 complete
B
MAKER
lab 2 fragments
pb 3' end only
Hox3 3' end only
Dfd complete
Scr complete
Antp fusion to Ubx
Ubx-i1 5' end only
abd-A-i1 split 5' UTR
Abd-B-i1 extended 5' UTR
Ubx-i2 missing
abd-A-i2 missing
Abd-B-i2 missing
C
Mikado
lab 2 fragments
pb complete
Hox3 complete
Dfd complete
Scr complete
Antp complete
Ubx-i1 missing
abd-A-i1 complete
Abd-B-i1 extended 5' UTR
Ubx-i2 complete
abd-A-i2 missing
Abd-B-i2 missing
D
49 RACE genes
manual annotation of Promoter peak First exon Last exon
transcriptomes
GMAP
reference annotations
mapped models
bedtools closest
model categories
First exon
Last exon
Promoter peak
Genrich peaks
Omni-ATAC
RNA-Seq
Single at least one model overlaps both first and last exon
Split models overlap first/ last exon, but none overlap both
First only Models only overlap first exon
Last only Models only overlap last exon
E
Fraction of RACE genes
1.0
0.8
0.6
0.4
0.2
0.0
RACE
Trinity-S21
Trinity-GG
StringTie2 L
StringTie2 SL
phaw_3.0
MAKER
Trinity-limb
Mikado
Reference
at least 1 model overlaps promoter peak
single
split
first only
last only
no model
model categories
F
BUSCO Assessment
Trinity-all C:917 [S:256, D:661], F:111, M:38
Trinity-S13 C:962 [S:335, D:627], F:49, M:55
Trinity-S19 C:989 [S:336, D:653], F:36, M:41
Trinity-S21 C:988 [S:299, D:689], F:47, M:31
Trinity-S23 C:1003 [S:305, D:698], F:36, M:27
Trinity-GG C:900 [S:450, D:450], F:117, M:49
StringTie2 L C:784 [S:496, D:288], F:110, M:172
StringTie2 SL C:979 [S:521, D:458], F:55, M:32
phaw_3.0 C:934 [S:803, D:131], F:63, M:69
MAKER C:863 [S:573, D:290], F:99, M:104
Trinity-limb (new) C:977 [S:353, D:642], F:59, M:30
Trinity-limb (old) C:981 [S:522, D:459], F:37, M:48
Mikado C:970 [S:735, D:235], F:58, M:38
Reference
Long-read
Short-read
%BUSCOs 0 20 40 60 80 100
Complete (C) single-copy (S)
Complete (C) duplicated (D)
Fragmented (F)
Missing (M)
G
Contribution of gene models from different sources to Mikado merged transcriptome
Percentage of gene models
25
20
15
10
5
0
25.0
15.6
10.8
10.6
8.3
6.7
5.7
4.6
4.4
3.5
3.2
1.4
0.1
StringTie2 SL
phaw_3.0
Trinity-all
Trinity-GG
MAKER
Trinity-S23
StringTie2 L
Trinity-S21
Trinity-limb (new)
Trinity-S19
Trinity-S13
Trinity-limb (old)
RACE

Supp. Fig. 1.3.1: Additional transcriptome and gene model evaluation metrics

A) Summarized completeness of RACE *Hox* gene models. Hox gene models for most genes are complete, including two isoforms each for *Ubx*, *abd-A*, and *Abd-B*. For *Hox3*, only a homeodomain sequence was isolated.

B) Summarized completeness of MAKER (phaw_5.0) *Hox* gene models. Many gene models show deviations from RACE sequences, including fragmentation, artefactual fusion of gene models, and extension of 5' UTRs.

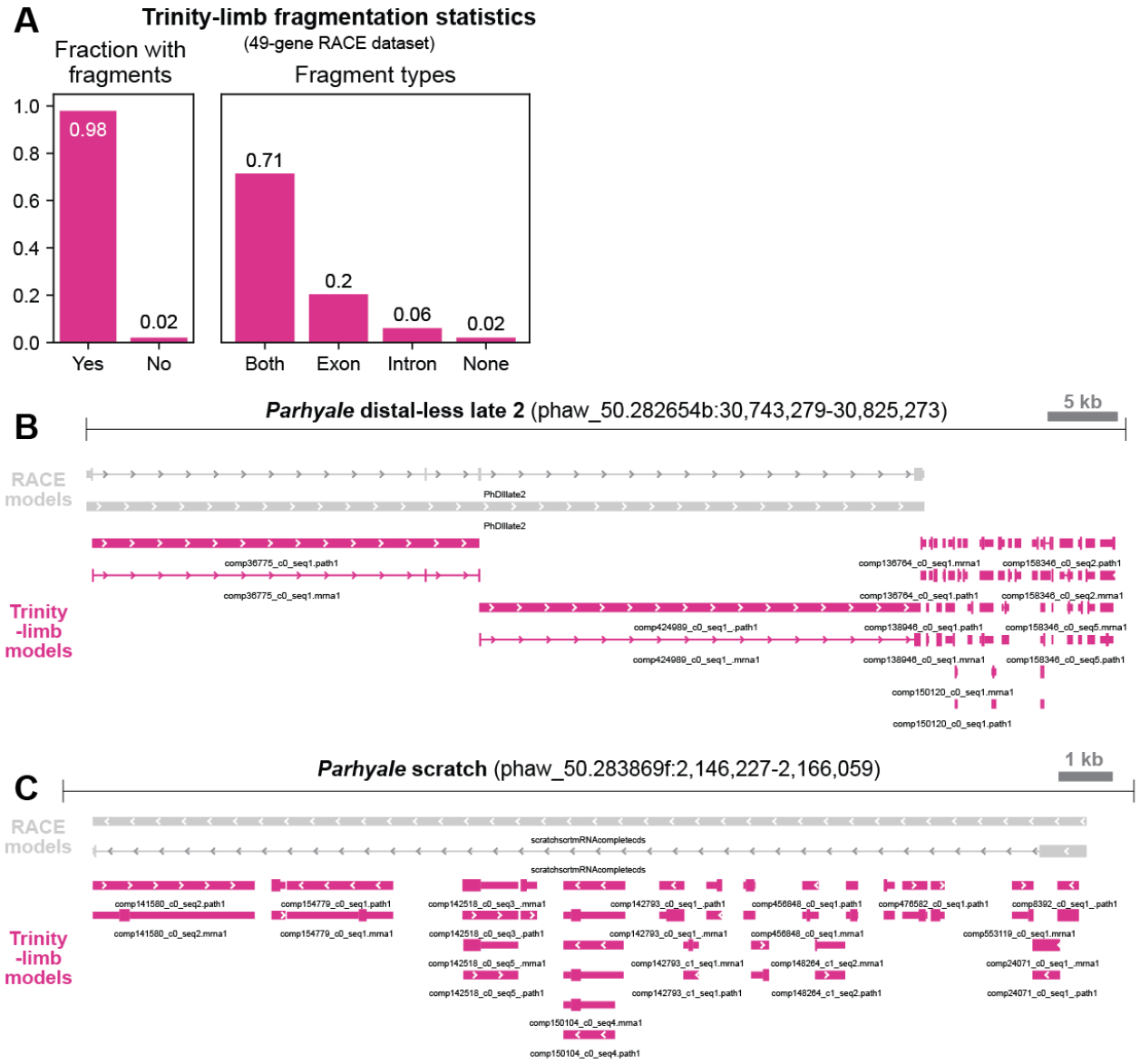
C) Summarized completeness of *Hox* gene models from Mikado transcriptome. The Mikado transcriptome appears to have more complete *Hox* gene models than the MAKER genome annotation. In addition, a complete *Hox3* sequence was generated.

D) Model evaluation methods for gene model fragmentation. 49 RACE genes were selected based on the ability to unambiguously annotate a first exon, last exon, and promoter peak based on examination of all gene models, Omni-ATAC-Seq peaks, and Omni-ATAC-Seq and RNA-Seq read pileups. For each transcriptome or genome annotation, for each RACE gene, all models were compared to manual annotation windows using bedtools. If a single model from all models in a given transcriptome overlapped with the first and last exon of the manual annotation, that RACE gene was classified as a "single" model for that transcriptome. If no single model overlapped both first and last exons, but models existed that overlapped either, the RACE gene was classified as "split" for that transcriptome. If gene models only overlapped with manually annotated first or last exons, then the RACE gene was classified as "first" or "last". Finally, if no gene models overlapped, the RACE gene was classified as "no model".

E) Summary of gene model evaluations from D for select transcriptomes. Overall, the Trinity-S21 transcriptome had the highest number of models that overlapped the Omni-ATAC promoter peak, and the Mikado transcriptome had the second highest. The StringTie2 SL transcriptome had the highest number of single gene models, and the Mikado transcriptome had the second highest. For both metrics, the Mikado transcriptome outperformed the MAKER genome annotation.

F) BUSCO evaluation for additional transcriptomes included in the dataset.

G) Fraction of gene models attributed to each transcriptome in the final Mikado transcriptome. Overall, the StringTie2 SL transcriptome produced the greatest number of transcripts evaluated as "best" by the Mikado pipeline.

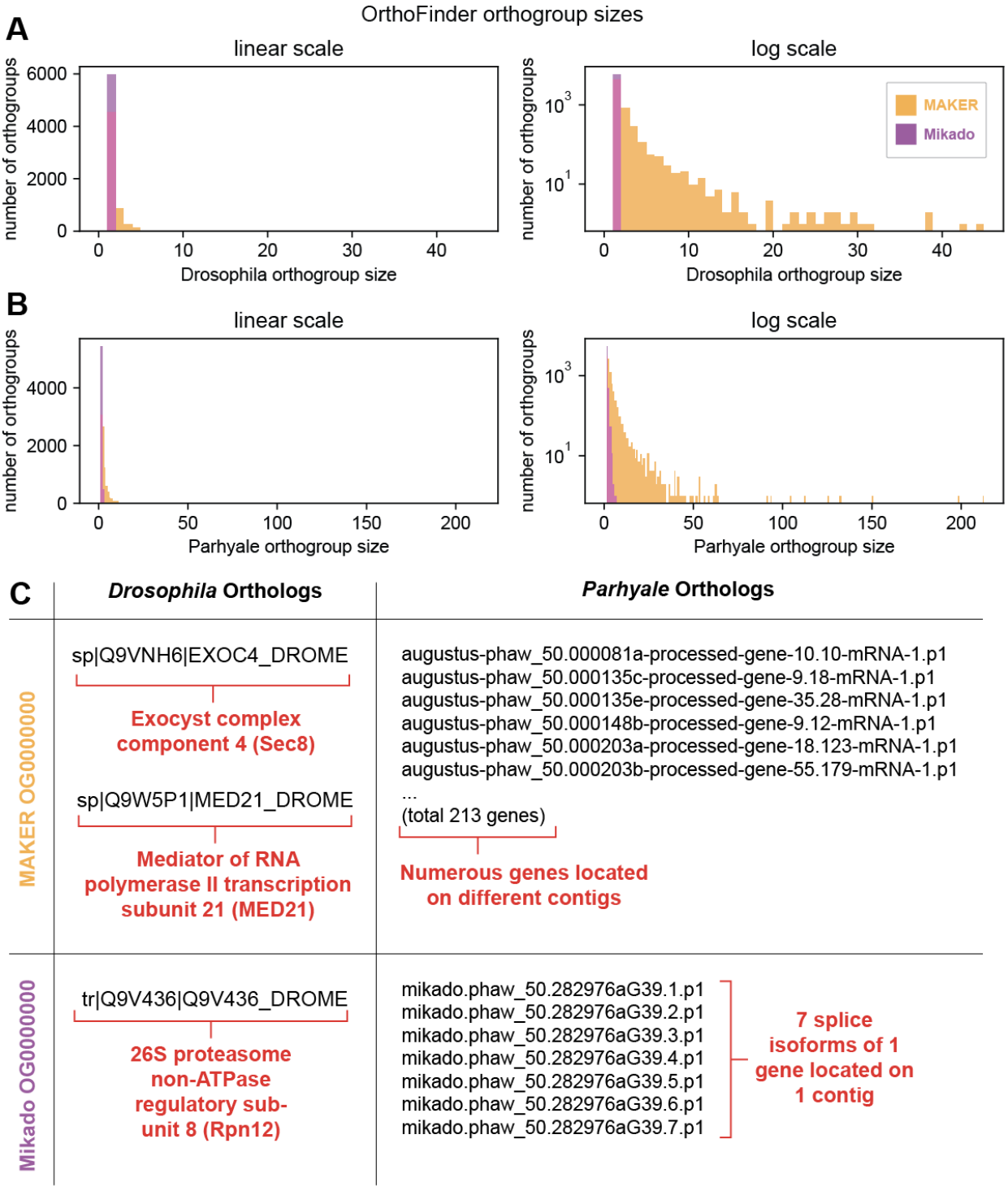


Supp. Fig. 1.3.2: Additional transcriptome and gene model evaluation metrics

A) Fraction of Trinity-limb (old) gene models with spurious transcript fragments when compared with RACE data. Fragments were observed overlapping with introns, exons, or both; some transcript fragments appeared to contain erroneous splice junctions joining exons and introns, while others were strictly intronic.

B) Example of transcript fragments compared to RACE data at the *Parhyale distal-less late 2* locus. This gene model had two transcript fragments overlapping with known exons, but no single gene model spanning the entire locus. Moreover, a large number of small 3'UTR-mapping fragments were observed, a signal that was frequently observed across genes in the RACE dataset.

C) Example of transcript fragments compared to RACE data at the *Parhyale scratch* locus. Transcript fragments that overlap with both intronic and exonic sequences are observed, as well as strictly intronic fragments. This gene model was a particularly extreme example of transcript fragmentation observed among the genes in the RACE dataset.



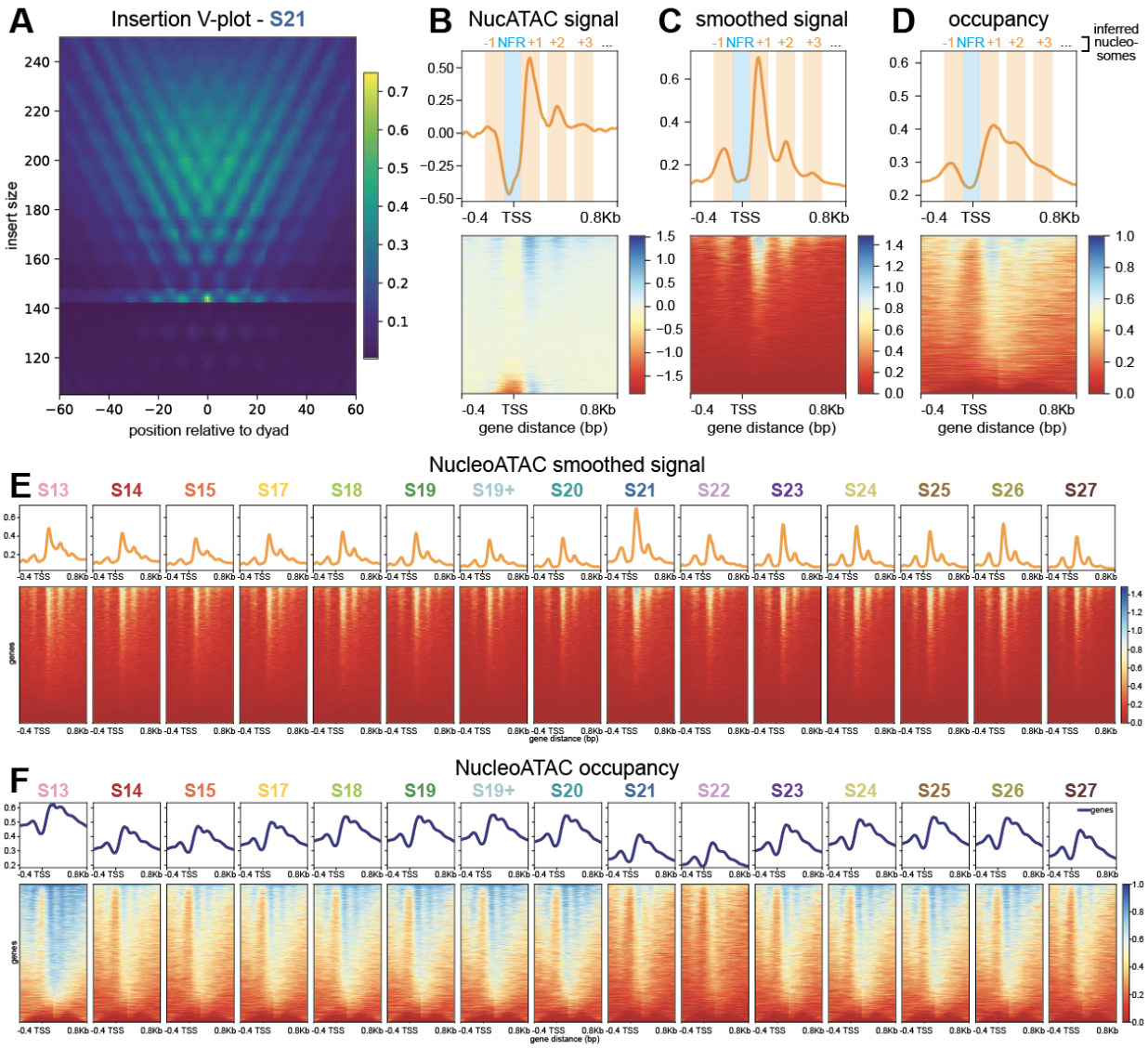
Supp. Fig. 1.3.3: Additional transcriptome and gene model evaluation metrics

A) Orthogroup sizes grouped by number of *Drosophila* orthologs in each orthogroup for the MAKER genome annotation and the Mikado transcriptome. The MAKER genome annotation, when annotated using OrthoFinder, produces many gene models that are grouped into orthogroups with large numbers of putative orthologs when compared to the results of using OrthoFinder on the Mikado transcriptome.

B) Orthogroup sizes grouped by number of *Parhyale* orthologs in each orthogroup for the MAKER genome annotation and the Mikado transcriptome. As with the *Drosophila* measure, the number of orthogroups with

large numbers of orthologs is greater for the MAKER genome annotation.

C) Example of peptides spuriously classified as orthologs when comparing the MAKER genome annotation to the Mikado transcriptome. OG0000000 is the single largest orthogroup identified by OrthoFinder when comparing the *Drosophila* UNIPROT database to either the MAKER genome annotation or the Mikado transcriptome. In the MAKER annotation, the largest orthogroup contains two unrelated *Drosophila* proteins (Sec8 and MED21), which are grouped together along with 213 different *Parhyale* gene models scattered across the genome. Examination of other orthogroups revealed a similar result, in which unrelated *Drosophila* proteins were grouped along with numerous *Parhyale* gene models, impeding precise identification of orthologs between the species, and suggesting problems in the quality of the MAKER gene models. In the Mikado annotation, the largest orthogroup contains a single *Drosophila* protein (Rpn12) grouped along with 7 splice isoforms of a single *Parhyale* gene model located on a single contig. Examination of other orthogroups derived from the Mikado transcriptome revealed fewer obviously incorrect groupings than observed for the MAKER annotation.



Supp. Fig. 1.4: NucleoATAC quality control and signal over time

A) Insertion V-plot generated from NucleoATAC using reads from S21 libraries. A clear V-shaped pattern is observed, as expected from read data fragmented along a nucleosomal distribution.

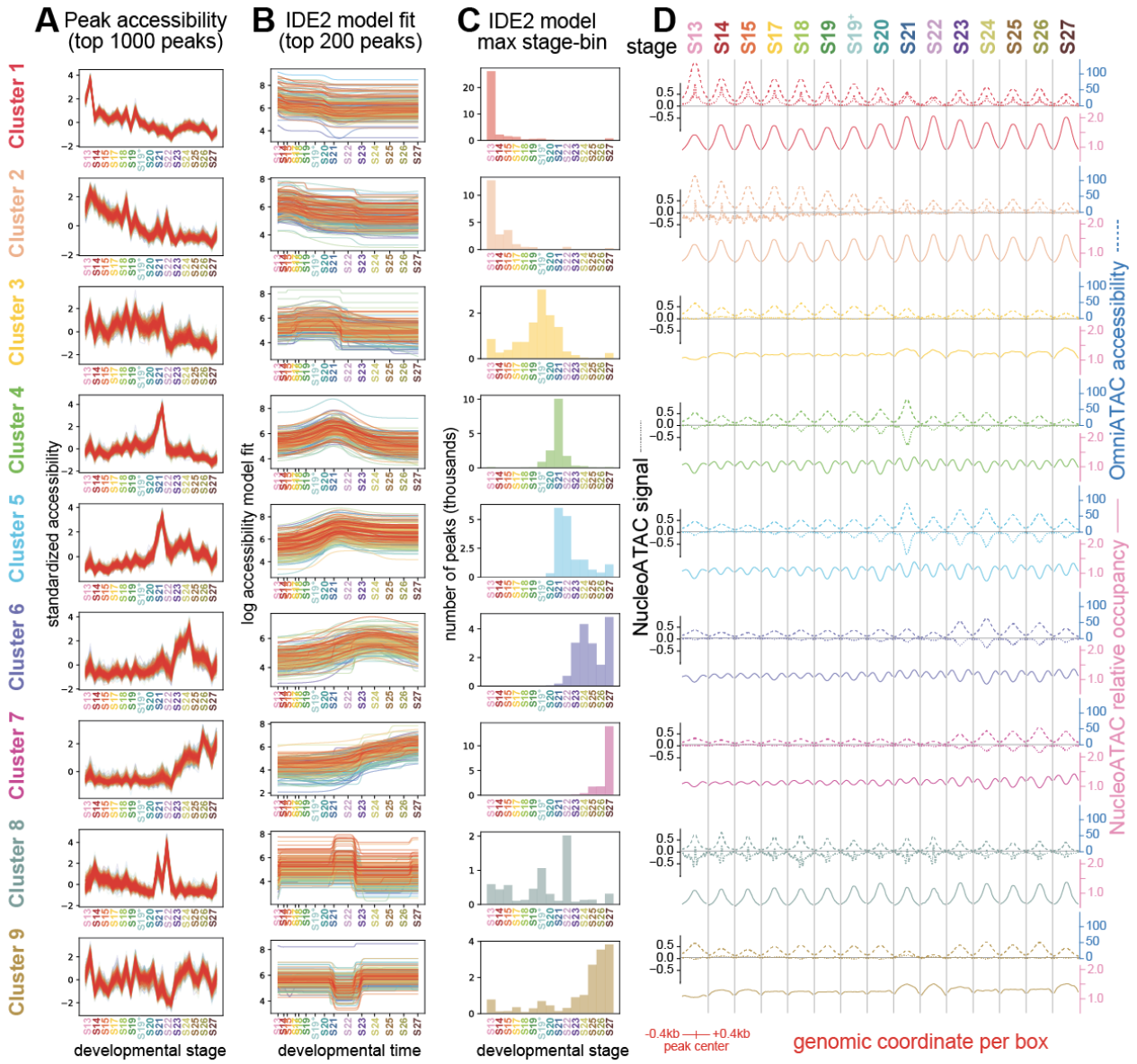
B) NucleoATAC raw signal at mRNA starts genome-wide. A strong +1 and -1 nucleosome position signal is observed. A strong negative signal is observed at the expected nucleosome-free region (NFR). Overlaid windows in B-D were drawn manually based on observed signal.

C) NucleoATAC smoothed signal at mRNA starts genome-wide. A strong +1 and -1 nucleosome position signal is observed.

D) NucleoATAC occupancy at mRNA starts genome-wide. A strong +1 and -1 nucleosome position signal is observed.

E) NucleoATAC smoothed signal at mRNA starts genome-wide for all developmental stages.

F) NucleoATAC occupancy signal at mRNA starts genome-wide for all developmental stages. The occupancy value calculated by NucleoATAC varies based on library size; in this plot, values are not normalized based on library size.



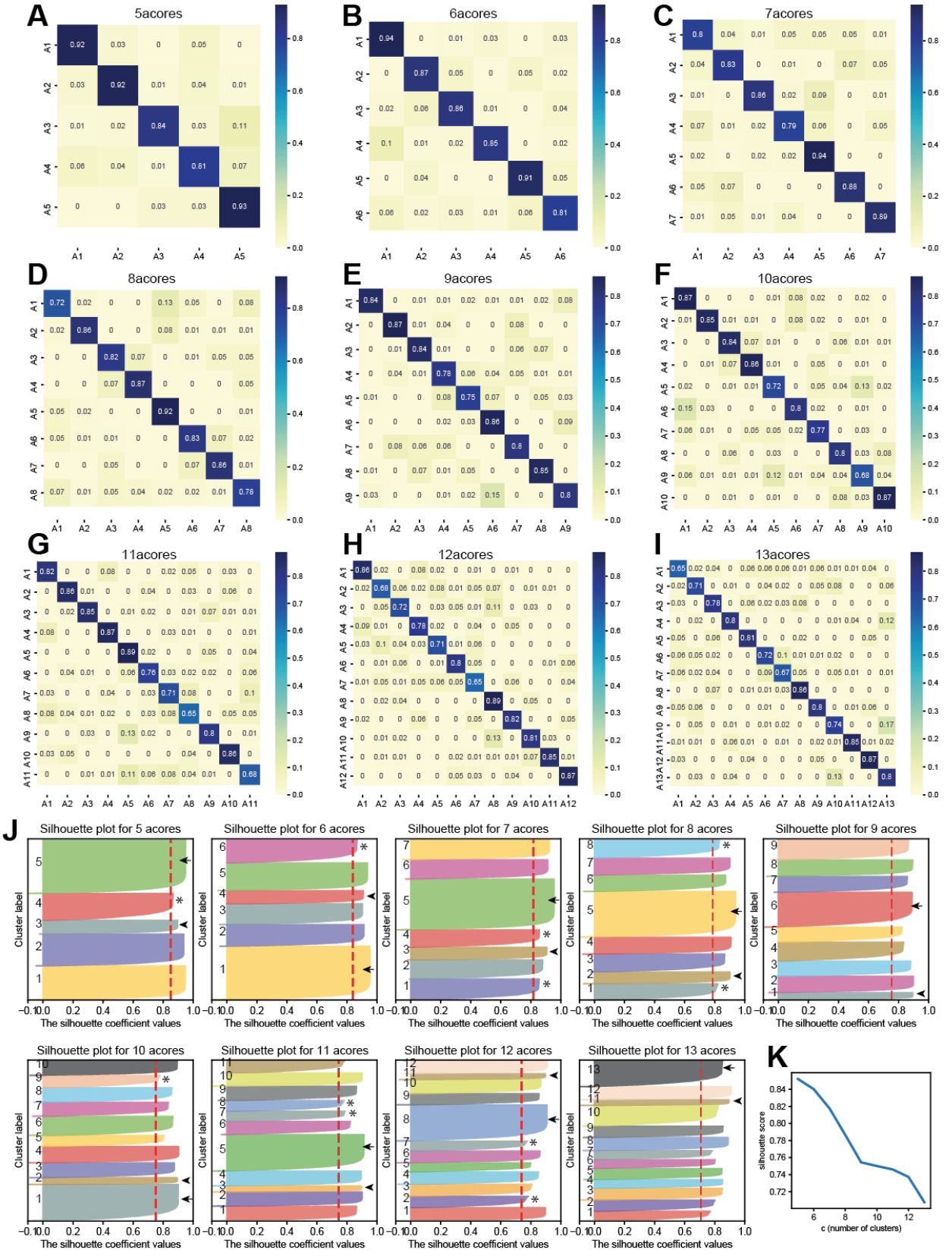
Supp. Fig. 1.5.1: Identification and classification of regulatory element clusters

A) Standardized accessibility plot for top 1000 peaks with strongest membership in each of the 9 clusters. Each library is plotted as a separate point along the line plot.

B) ImpulseDE2 model fits for the top 200 peaks with strongest membership in each of the 9 clusters.

C) Histogram of when all peaks in each cluster achieve their maximum accessibility as calculated by IDE2 max fit.

D) Line plots showing Omni-ATAC accessibility (dashed line, blue axis), NucleoATAC signal (dotted line, black axis), and relative NucleoATAC histone occupancy (solid line, pink axis) for peaks in each cluster across time. Each box contains a line plot summarizing each of the three signals at all peaks within that cluster with respect to a given developmental Omni-ATAC-Seq library. The center of each line plot reflects the average signal at the center of all Omni-ATAC-Seq peaks in that cluster at that developmental stage. Signal is visualized at 0.4kb upstream and downstream of peak centers. Within each signal type, the axes across line plots for each cluster are identical.


A
5acores
B
6acores
C
7acores
D
8acores
E
9acores
F
10acores
G
11acores
H
12acores
I
13acores
J
Silhouette plot for 5 acores
Silhouette plot for 6 acores
Silhouette plot for 7 acores
Silhouette plot for 8 acores
Silhouette plot for 9 acores
Silhouette plot for 10 acores
Silhouette plot for 11 acores
Silhouette plot for 12 acores
Silhouette plot for 13 acores
Cluster label
The silhouette coefficient values
K
silhouette score
c (number of clusters)

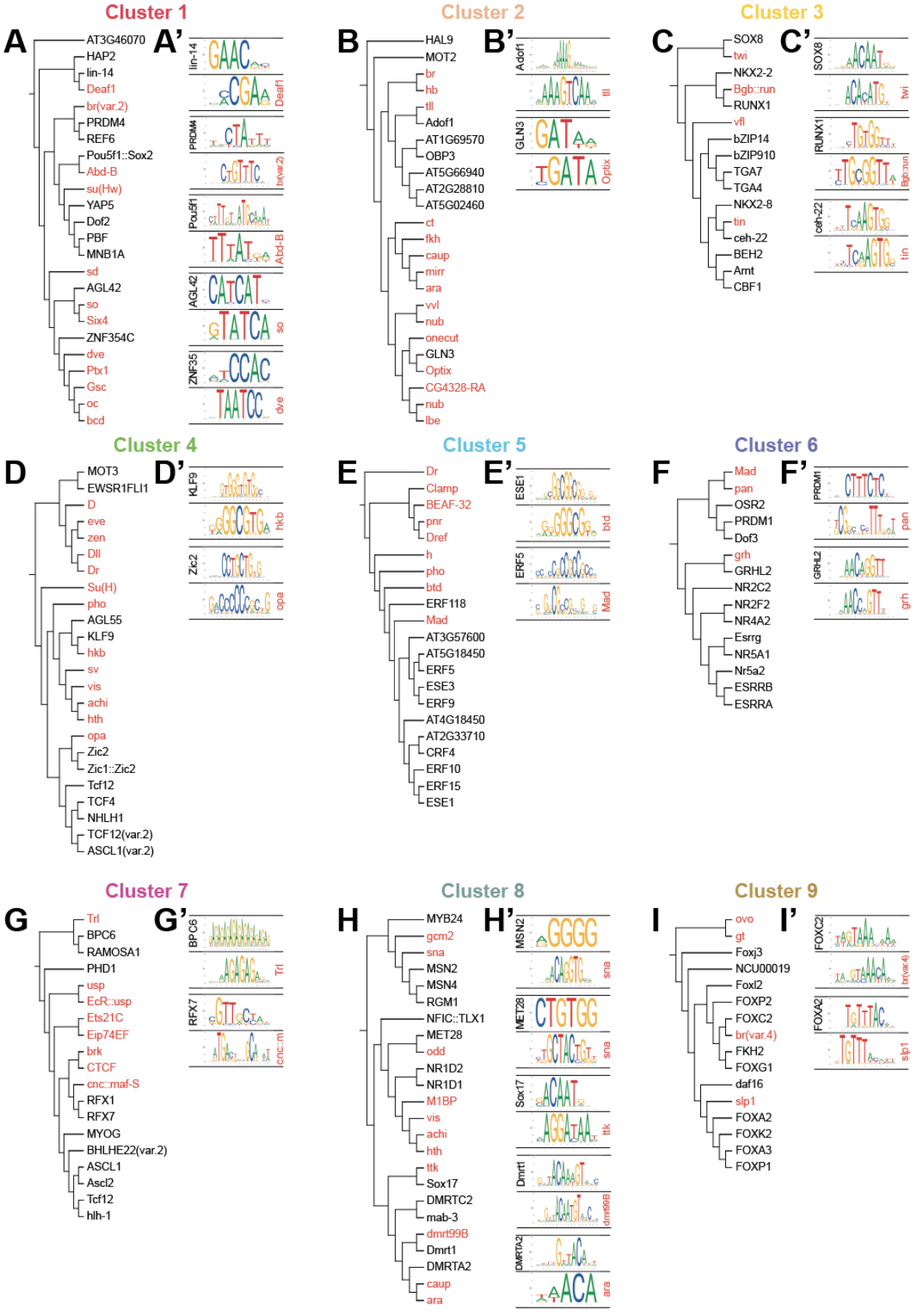
Supp. Fig. 1.5.2: Mfuzz clustering assessment

Note: cluster labels in this plot are for assessment purposes, and do not correspond to cluster labels used in other figures.

A–I) Cluster overlap matrices generated my Mfuzz indicating percentage overlap between clusters.

J) Silhouette plots for different numbers of clusters. Dashed red line marks the average silhouette score. Asterisks mark clusters with marginal silhouettes (silhouettes at or below the average). An arrowhead marks the thinnest silhouette, while an arrow marks the thickest silhouette within each plot. All cluster numbers other than 9 and 13 showed marginal clusters and large contrast between the thinnest and thickest silhouette widths.

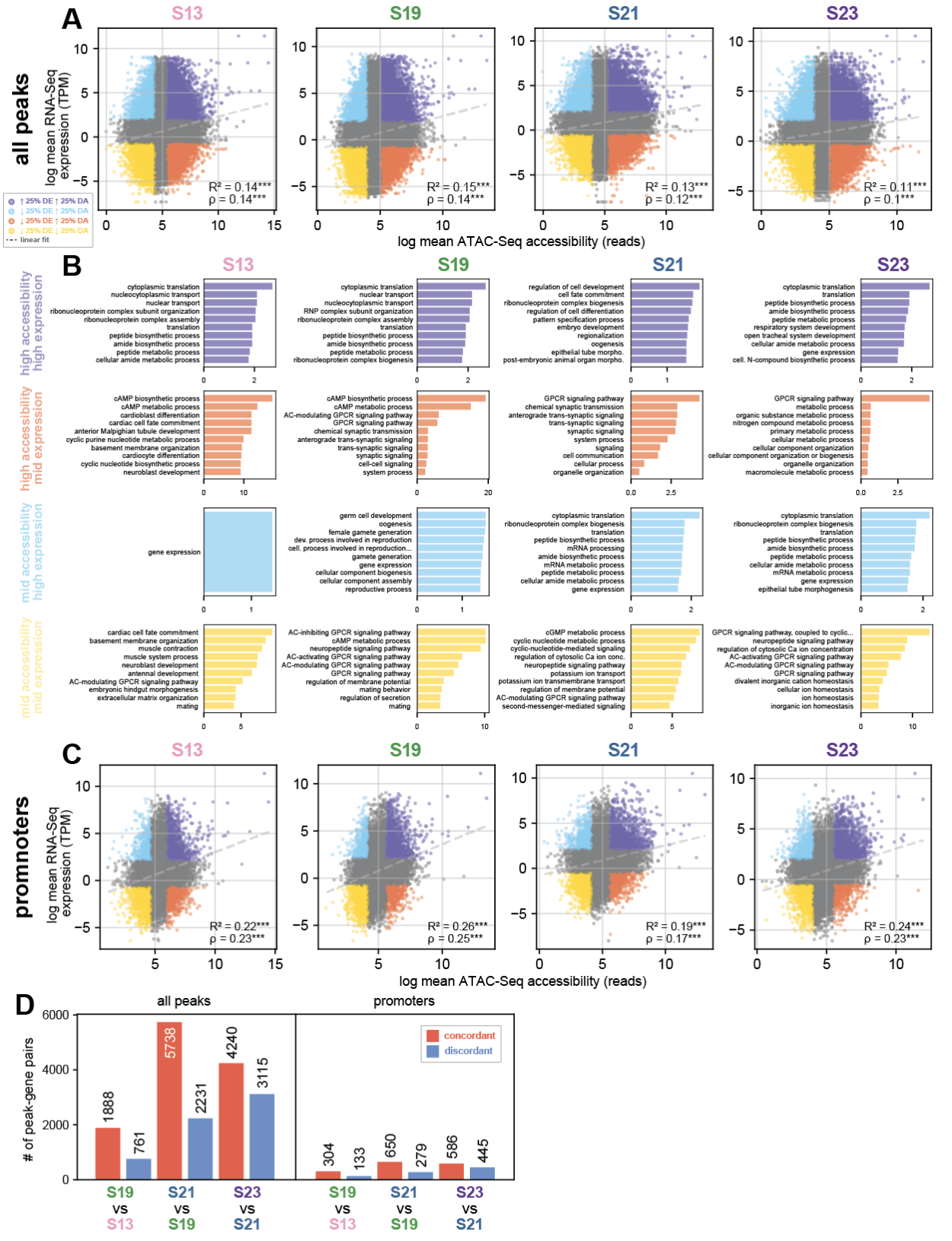
K) Mean silhouette score by number of clusters.


Cluster 1
A
AT3G46070
HAP2
lin-14
Deaf1
br(var.2)
PRDM4
REF6
Pou5f1::Sox2
Abd-B
su(Hw)
YAP5
Dof2
PBF
MNB1A
sd
AGL42
so
Six4
ZNF354C
dve
Ptx1
Gsc
oc
bcd
A'
lin-14
Deaf1
PRDM4
br(var.2)
Pou5f1
Abd-B
AGL42
so
ZNF354C
dve
Cluster 2
B
HAL9
MOT2
br
hb
tll
Adof1
AT1G69570
OBP3
AT5G66940
AT2G28810
AT5G02460
ct
fkh
caup
mirr
ara
vvl
nub
onecut
GLN3
Optix
CG4328-RA
nub
lbe
B'
Adof1
tll
GLN3
Optix
Cluster 3
C
SOX8
twi
NKX2-2
Bgb::run
RUNX1
vfl
bZIP14
bZIP910
TGA7
TGA4
NKX2-8
tin
ceh-22
BEH2
Arnt
CBF1
C'
SOX8
twi
RUNX1
Bgb::run
ceh-22
tin
Cluster 4
D
MOT3
EWSR1FLI1
D
eve
zen
Dll
Dr
Su(H)
pho
AGL55
KLF9
hkb
sv
vis
achi
hth
opa
Zic2
Zic1::Zic2
Tcf12
TCF4
NHLH1
TCF12(var.2)
ASCL1(var.2)
D'
KLF9
hkb
Zic2
opa
Cluster 5
E
Dr
Clamp
BEAF-32
pnr
Dref
h
pho
btd
ERF118
Mad
AT3G57600
AT5G18450
ERF5
ESE3
ERF9
AT4G18450
AT2G33710
CRF4
ERF10
ERF15
ESE1
E'
ESE1
btd
ERF5
Mad
Cluster 6
F
Mad
pan
OSR2
PRDM1
Dof3
grh
GRHL2
NR2C2
NR2F2
NR4A2
Esrrg
NR5A1
Nr5a2
ESRRB
ESRRA
F'
PRDM1
pan
GRHL2
grh
Cluster 7
G
Trl
BPC6
RAMOSA1
PHD1
usp
EcR::usp
Ets21C
Eip74EF
brk
CTCF
cnc::maf-S
RFX1
RFX7
MYOG
BHLHE22(var.2)
ASCL1
Ascl2
Tcf12
hlh-1
G'
BPC6
Trl
RFX7
cnc::m
Cluster 8
H
MYB24
gcm2
sna
MSN2
MSN4
RGM1
NFIC::TLX1
MET28
odd
NR1D2
NR1D1
M1BP
vis
achi
hth
ttk
Sox17
DMRTC2
mab-3
dmrt99B
Dmrt1
DMRTA2
caup
ara
H'
MSN2
sna
MET28
sna
Sox17
ttk
Dmrt1
dmrt99B
DMRTA2
ara
Cluster 9
I
ovo
gt
Foxj3
NCU00019
Foxl2
FOXP2
FOXC2
br(var.4)
FKH2
FOXG1
daf16
slp1
FOXA2
FOXK2
FOXA3
FOXP1
I'
FOXC2
br(var.4)
FOXA2
slp1

Supp. Fig. 1.6: Alignment of motifs in fuzzy clusters

A–I) Dendrograms generated from sequence alignment of JASPAR motifs using the online STAMP tool visualized using Interactive Tree of Life (iTOL). Top 12 most-enriched transcription factor binding sites from the entire JASPAR CORE database are in black for each dendrogram; up to top 12 most-enriched transcription factor binding sites from the *Drosophila* transcription factors in JASPAR CORE are in red for each dendrogram.

A'–I') Select pairs of transcription factor binding sites with apparently similar sequences based on STAMP alignment. All JASPAR CORE TFs are in black; *Drosophila* TFs are in red.


A
S13
S19
S21
S23
all peaks
log mean RNA-Seq expression (TPM)
↑ 25% DE ↑ 25% DA
↓ 25% DE ↑ 25% DA
↓ 25% DE ↓ 25% DA
linear fit
R² = 0.14*** ρ = 0.14***
R² = 0.15*** ρ = 0.14***
R² = 0.13*** ρ = 0.12***
R² = 0.11*** ρ = 0.1***
log mean ATAC-Seq accessibility (reads)
B
high accessibility high expression
high accessibility mid expression
mid accessibility high expression
mid accessibility mid expression
C
promnoters
R² = 0.22*** ρ = 0.23***
R² = 0.26*** ρ = 0.25***
R² = 0.19*** ρ = 0.17***
R² = 0.24*** ρ = 0.23***
log mean ATAC-Seq accessibility (reads)
D
all peaks
promoters
concordant
discordant
# of peak-gene pairs
1888
761
5738
2231
4240
3115
304
133
650
279
586
445
S19 vs S13
S21 vs S19
S23 vs S21

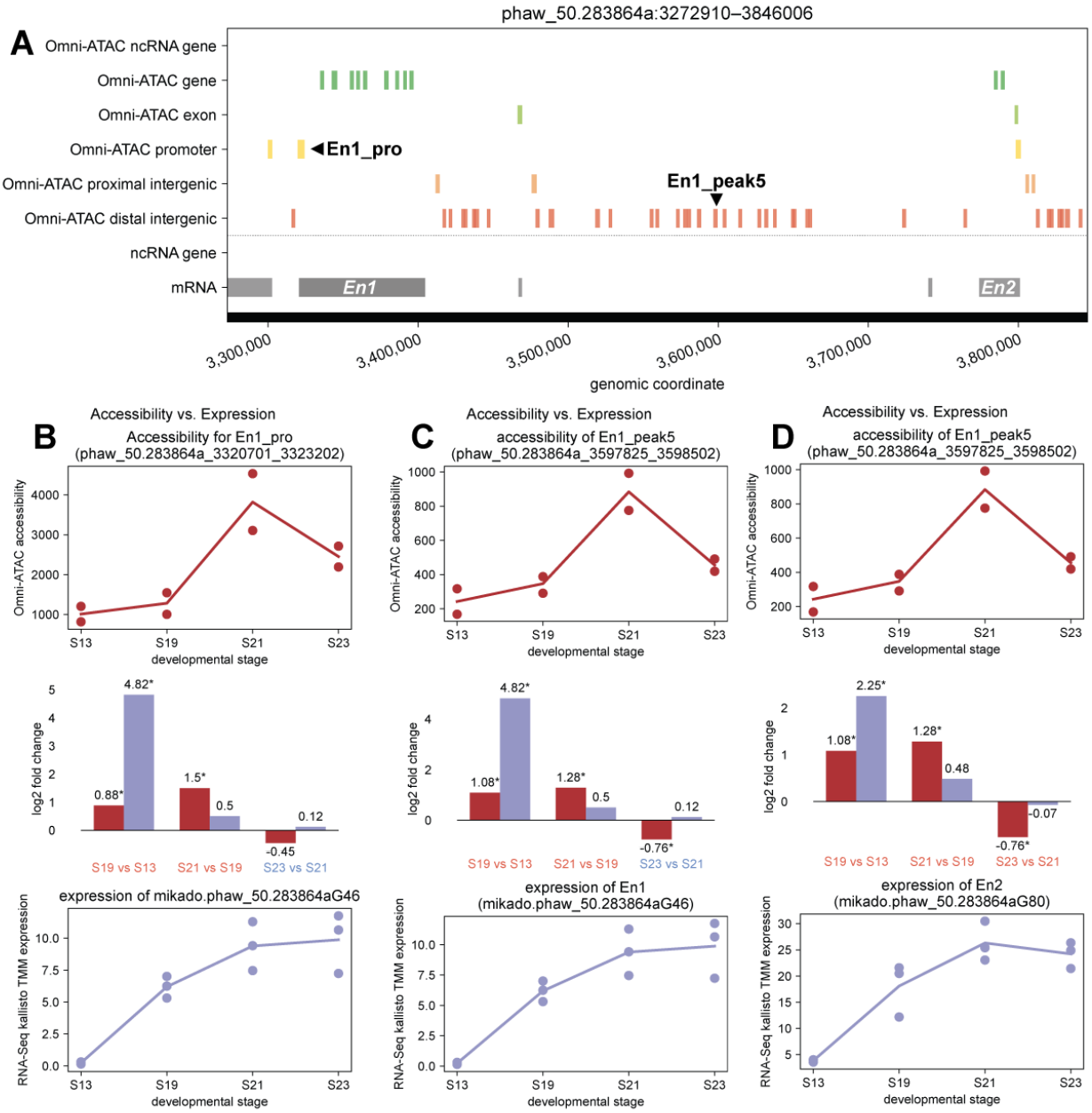
Supp. Fig. 1.7.1: Correlation between accessibility and gene expression at individual timepoints

A) Correlation between log mean RNA-Seq TPM and log mean Omni-ATAC-Seq accessibility at all peaks in each developmental stage. Peaks colored based on highest and lowest quartiles of expression and accessibility. Dotted line represents a linear fit to all data in the plot. Pearson correlation R^2 and Spearman correlation ρ for all points are displayed in each plot.

B) GO-term enrichment for gene-peak pairs with high accessibility and high expression (purple), high accessibility and medium expression (coral), medium accessibility and high expression (blue), and medium accessibility and medium expression (yellow), for all peaks. The threshold for "high" expression and accessibility were all values above the 75th percentile, while "mid" expression and accessibility were all values below the 25th percentile. Gene lists were extracted based on the nearest Mikado gene to each peak, and orthologous *Drosophila* names were assigned to genes using OrthoFinder.

C) Correlation between log mean RNA-Seq TPM and log mean Omni-ATAC-Seq accessibility at promoter peaks.

D) Quantification of number of significant peak-gene pairs at each time point comparison in Fig. 7. Left plot quantifies number of concordant and discordant peaks across all peak-gene pairs, while right plot quantifies peak-gene pairs that were assigned as promoters.



Supp. Fig. 1.7.2: Visualization of accessibility versus expression for arbitrary peak-gene pairs

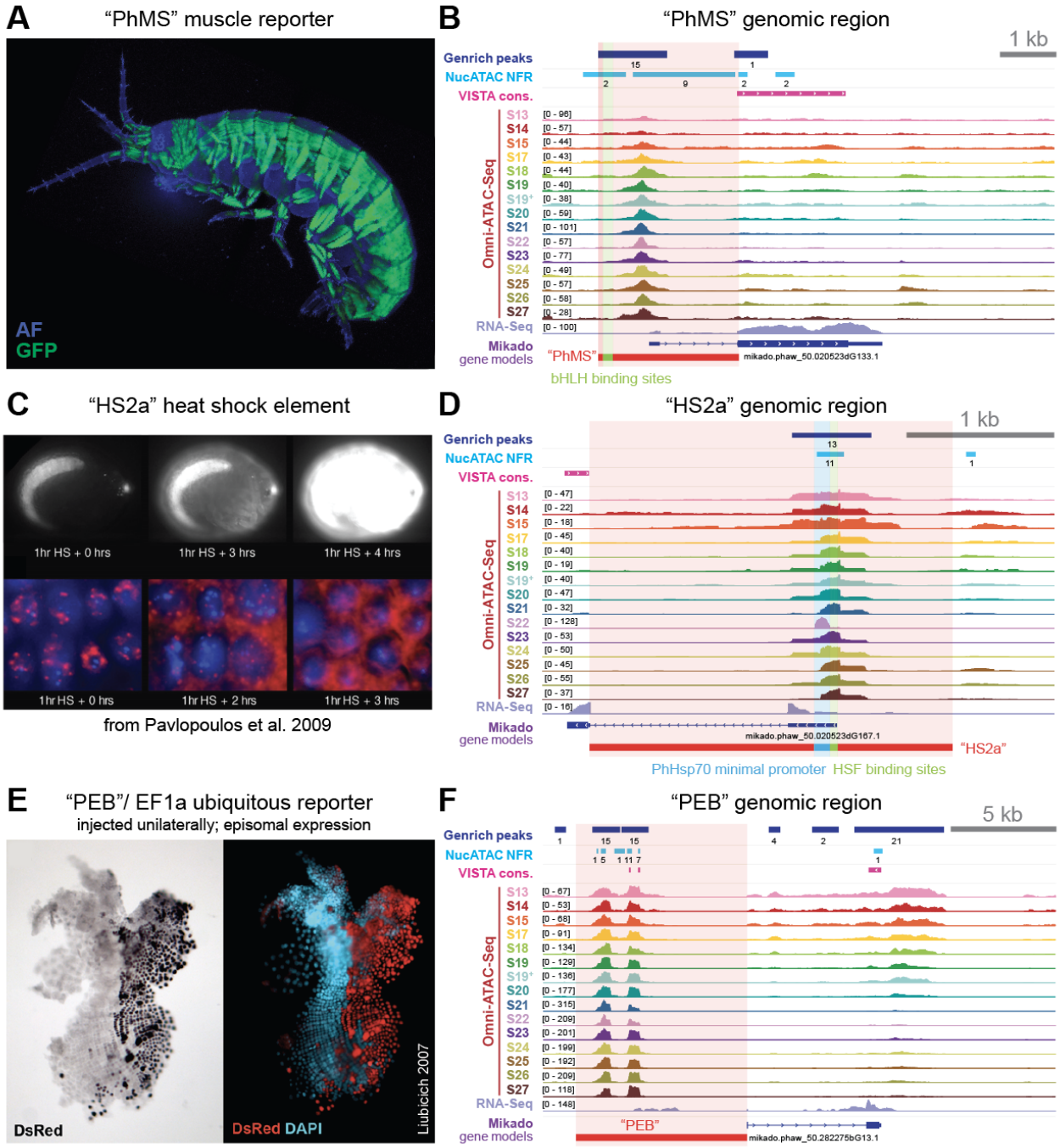
A) Visualization of the genomic region around the Parhyale Engrailed-1 and Engrailed-2 loci. Text labels in plot region were added manually after plotting.

B) Visualization of the average accessibility (red line) and raw accessibility values (red dots) for the promoter peak of the *Engrailed-1* gene (top). Visualization of the average gene expression (lavender line) and raw expression values (lavender dots) of the nearest gene to the *Engrailed-1* promoter peak (in this case, *Engrailed-1*, or mikado.phaw_50.283864aG46)(bottom). Visualization of log2 fold change in accessibility (red bars) and gene expression (lavender bars) for each pair of timepoint comparisons (S19 vs. S13, S21 vs. S19, and S23 vs. S21) (middle). Asterisks in bar charts mark log2 fold changes called as statistically significant by DESeq2 (padj < 0.05). Color of comparison labels indicates concordant or discordant relationships between accessibility and gene expression (red for concordant, blue for discordant). When interpreting this visualization, it is important to note

that some comparisons may show high log2 fold change that is non-significant; usually, this is when the measured values for a feature are very low. Additionally, statements about concordance and discordance should consider whether both accessibility and expression log2 fold changes are significant. In the case of this example, the comparisons at S19 vs. S13 would be considered truly "concordant" in our dataset, while the comparisons at S21 vs. S19 and S23 vs. S21 would not be considered concordant or discordant, as one or more features are not significantly changing over time.

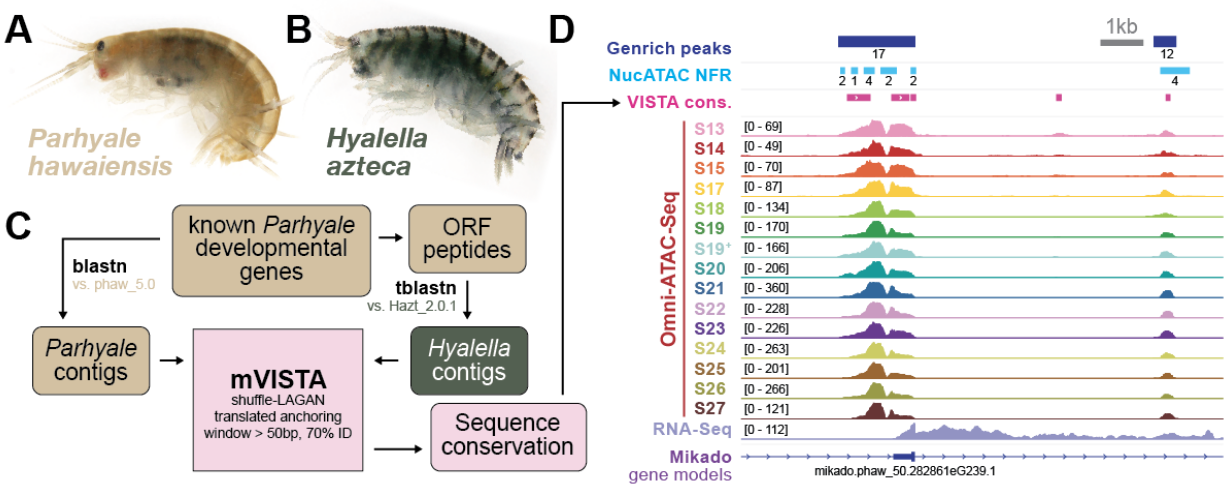
C) Visualizations as described in (B) for the *Engrailed-1* gene and a distal intergenic peak located between *Engrailed-1* and *Engrailed-2*, nicknamed "En1_peak5", position labeled in (A).

D) Visualizations as described in (B) for the *Engrailed-2* gene and "En1_peak5".



Supp. Fig. 1.8.1: Omni-ATAC-Seq recovers previously identified regulatory elements

A) Expression of the PhMS reporter visualized using GFP in a juvenile *Parhyale*.
B) The PhMS genomic region. A strong Genrich peak overlaps with a predicted cluster of b-HLH binding sites.
C) Heat shock expression images from a transgenic *Parhyale* embryo carrying the HS2a element.
D) The HS2a genomic region. A peak covers the PhHsp70 minimal promoter and heat shock factor binding sites.
E) Expression of the PEB reporter, visualized in a dissected embryo using a DsRed antibody. Injected embryos showed nearly full expression of this reporter in each germ layer (Liubicich 2007 PhD thesis).
F) The PEB genomic region. Two strong peaks overlap with the 5' end of this reporter.



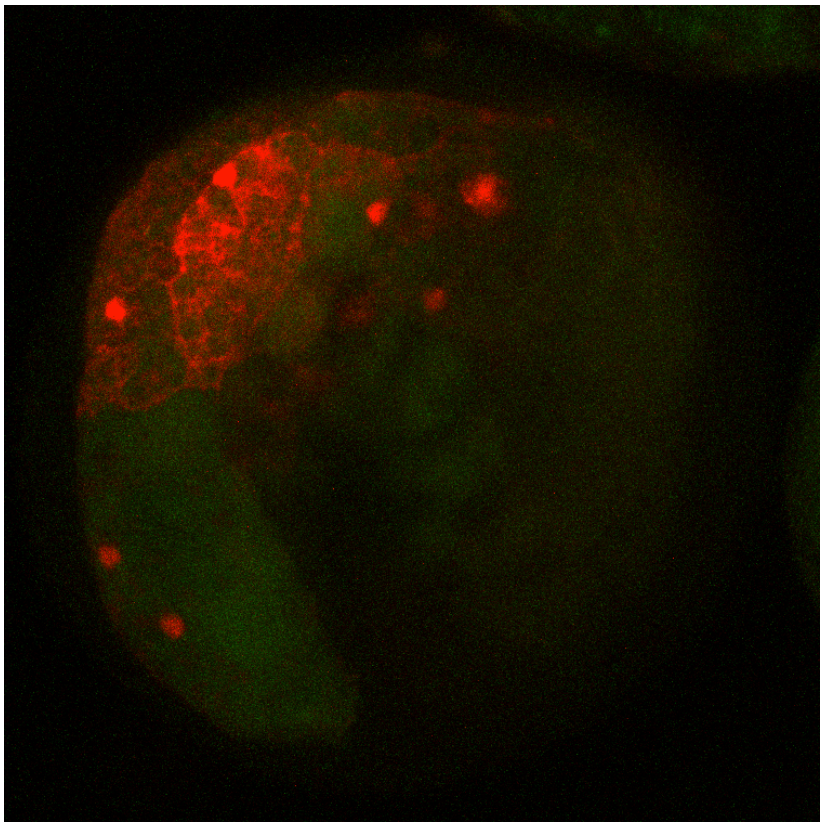
Supp. Fig. 1.8.2: VISTA homology identification between *Parhyale* and *Hyalella*

A) Adult female *Parhyale hawaiensis*.

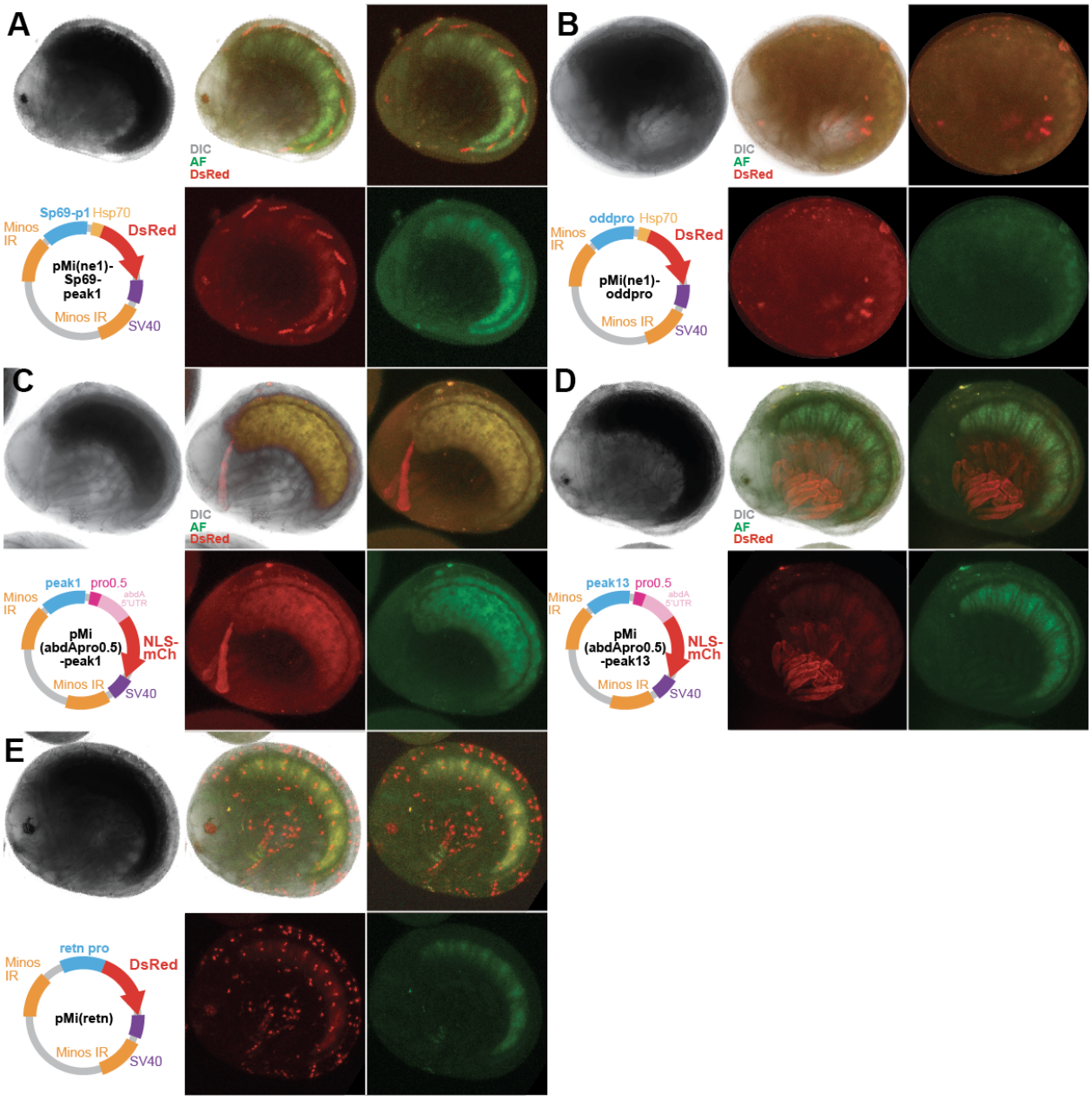
B) Adult female *Hyalella azteca*.

C) Pipeline for identification of orthologous genomic regions between *Hyalella* and *Parhyale* for VISTA sequence conservation analysis.

D) Visualization of Genrich peaks, NucleoATAC nucleosome-free regions (NFRs), VISTA conservation, Omni-ATAC-Seq signal, RNA-Seq signal, and Mikado gene models at the putative Sp69 promoter peak.



Supplementary Video 1.1: Yolk cells labeled by the Hsp70-p2 reporter
Max intensity projection; frames taken 5 minutes apart.



Supp. Fig. 1.9: Spontaneous, non-reproducible expression patterns observed in Minos assays

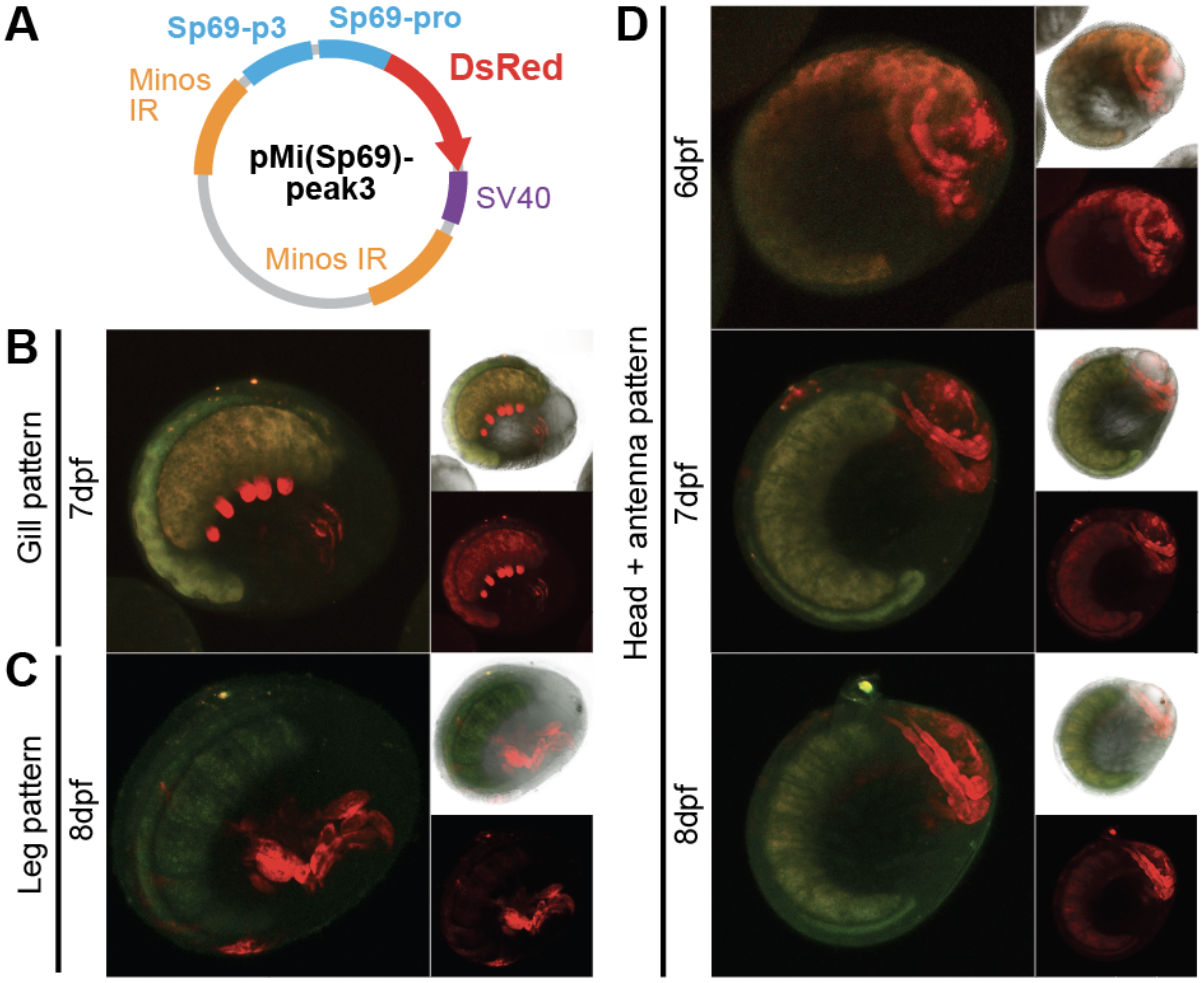
A) Expression observed in one embryo injected with a plasmid containing the PhHsp70 minimal promoter and the Sp69-p1 peak.

B) Expression observed in one embryo injected with a plasmid containing proximal promoter peak for the odd-skipped gene along with an Hsp80 minimal promoter.

C) Expression observed in one embryo injected with a plasmid containing a putative minimal promoter for the abd-A gene and an upstream peak (abd-A-peak1).

D) Expression observed in one embryo injected with a plasmid containing a putative minimal promoter for the abd-A gene and an upstream peak (abd-A-peak13).

E) Expression observed in one embryo injected with a plasmid containing the putative *Parhyale-retn* promoter.



Supp. Fig. 1.10: Expression observed from Sp69-p3 reporter

A) Plasmid map of the pMi(Sp69)-peak3 construct. The construct contains both the putative promoter peak of Sp69, along with a distal peak (peak3).

B) Gill and claw expression observed in one embryo injected with this plasmid.

C) Leg expression observed in one embryo injected with this plasmid.

D) Head and antennal expression patterns observed in 5 embryos injected with this plasmid; embryo with strongest expression pattern shown.

Supplementary Table 1.1: Illumina sequencing indices for Omni-ATAC libraries

Sample ID	i7 (4000)	i5 (4000) PE	i5 (4000) SR Flowcell/MiSeq
S13A	TACAGAGC	CTTGGATG	CATCCAAG
S13B	AAGCGTTC	GGAAGGAT	ATCCTTCC
S14A	AGTGACCT	TTCTCTCG	CGAGAGAA
S14B	ACCATAGG	AAGTCCGT	ACGGACTT
S15A	CTCGAACA	ATGGTCCA	TGGACCAT
S15B	TGCGTAAC	TGTCCAGA	TCTGGACA
S17A	AACAGTCC	GATTGCTC	GAGCAATC
S17B	CTAAGACC	GTTGTAGC	GCTACAAC
S18A	TGTTCCGT	GTGAATCC	GGATTCAC
S18B	ACTCAACG	ACTCCATC	GATGGAGT
S19A	GCCTTCTT	TGAGGTGT	ACACCTCA
S19B	TGCTCTAC	TCTTGACG	CGTCAAGA
S19plusA	GTACCACA	GATACTGG	CCAGTATC
S19plusB	GCATTGGT	ATCTTCGG	CCGAAGAT
S20A	CTGTGGTA	ACATTGCG	CGCAATGT
S20B	TTACCGAC	GTTGTTCG	CGAACAAC
S21A	TGACCGTT	CATGGAAC	GTTCCATG
S21B	GCTAAGGA	AACGTGGA	TCCACGTT
S22A	TAGCCATG	CTGCACTT	AAGTGCAG
S22B	ACTTGGCT	GTAGCATC	GATGCTAC
S23A	TGTCACAC	GTGCTTAC	GTAAGCAC
S23B	GACACAGT	CAACCTAG	CTAGGTTG
S24A	GTGGTATG	ATCTCGCT	AGCGAGAT
S24B	ATGCGCTT	GGACCTAT	ATAGGTCC
S25A	AGGAACAC	ACGTTACC	GGTAACGT
S25B	TGGAAGCA	TACTGCGT	ACGCAGTA
S26A	GTCGTTAC	GGTACTAC	GTAGTACC
S26B	GAACCTTC	AGCTTGAG	CTCAAGCT
S27A	TATGACCG	TGTGAAGC	GCTTCACA
S27B	AACGCACA	CGTTATGC	GCATAACG

Supplementary Table 1.2: Manual annotation coordinates of 49 selected RACE genes

gene_name	IGV_address	Strand	pro_peak_address	exon1_address	exon-1_address	Trinity fragments?	Fragment type
AbdominalBisoform1cloneJ7b11	phaw_50.282695a:24623698-24688129	+	phaw_50.282695a:24622599-24624450	phaw_50.282695a:24,623,624-24,624,479	phaw_50.282695a:24,687,725-24,692,898	Yes	Both
abdomminalAisoform2cloneK313	phaw_50.282695a:25629814-25736948	+	phaw_50.282695a:25628608-25630296	phaw_50.282695a:25,627,088-25,630,363	phaw_50.282695a:25,736,469-25,741,275	Yes	Both
AntennapediaSracecloneL18	phaw_50.282695a:26532519-26613138	+	phaw_50.282695a:26532243-26532820	phaw_50.282695a:26,532,505-26,533,152	phaw_50.282695a:26,611,100-26,613,267	Yes	Both
ash26F	phaw_50.283875a:11478705-11507717	+	phaw_50.283875a:11477924-11479069	phaw_50.283875a:11,478,677-11,480,112	phaw_50.283875a:11,507,570-11,508,560	Yes	Both
betacateninmRNAcompletecds	phaw_50.282639b:6484601-6538431	+	phaw_50.282639b:6451035-6452617	phaw_50.282639b:6,451,925-6,452,777	phaw_50.282639b:6,537,263-6,541,062	Yes	Both
col	phaw_50.283865b:130292-293921	+	phaw_50.283865b:127661-128803	phaw_50.283865b:130,178-130,591	phaw_50.283865b:296,755-299,392	Yes	Both
DeformedcDNAcloneFLb2	phaw_50.282695a:27062438-27152191	+	phaw_50.282695a:27061893-27063246	phaw_50.282695a:27,062,404-27,063,447	phaw_50.282695a:27,151,053-27,152,222	Yes	Exon
deltaproteinmRNAcompletecds	phaw_50.015400b:11871425-11889172	+	phaw_50.015400b:11870888-11871925	phaw_50.015400b:11,871,436-11,871,564	phaw_50.015400b:11,890,317-11,891,420	Yes	Intron
distallessEarlymRNAcompletecds	phaw_50.282654b:30641259-30681459	+	phaw_50.282654b:30640553-30642307	phaw_50.282654b:30,641,130-30,641,834	phaw_50.282654b:30,685,403-30,687,854	Yes	Both
dpp	phaw_50.015400a:5773544-5790898	+	phaw_50.015400a:5772703-5774732	phaw_50.015400a:5,773,547-5,774,013	phaw_50.015400a:5,789,594-5,799,808	Yes	Both
engrailed1	phaw_50.283864a:3322422-3404220	+	phaw_50.283864a:3320702-3323202	phaw_50.283864a:3,321,545-3,323,280	phaw_50.283864a:3,403,484-3,411,112	Yes	Both
engrailed2	phaw_50.283864a:3774988-3800052	-	phaw_50.283864a:3799353-3800567	phaw_50.283864a:3,798,531-3,800,059	phaw_50.283864a:3,757,132-3,758,228	Yes	Intron
eve2	phaw_50.283028b:9254180-9262623	-	phaw_50.283028b:9262111-9262983	phaw_50.283028b:9,262,071-9,262,616	phaw_50.283028b:9,250,289-9,256,607	Yes	Exon
extradenticleproteinexdgene	phaw_50.283468a:8063095-8076889	+	phaw_50.283468a:7799208-7800212	phaw_50.283468a:7,799,472-7,800,056	phaw_50.283468a:8,097,516-8,101,622	Yes	Both
forkheadORF	phaw_50.000135a:2643865-2644971	-	phaw_50.000135a:2649799-2650196	phaw_50.000135a:2,649,559-2,649,662	phaw_50.000135a:2,642,441-2,645,164	Yes	Exon
hes4	phaw_50.282654b:31719607-31734841	+	phaw_50.282654b:31719410-31720007	phaw_50.282654b:31,719,607-31,719,748	phaw_50.282654b:31,732,919-31,736,371	Yes	Both
homothoraxproteinhthgene	phaw_50.283815b:16274653-16574076	-	phaw_50.283815b:16604672-16607172	phaw_50.283815b:16,605,602-16,606,372	phaw_50.283815b:16,168,073-16,170,044	Yes	Both
KNIRPS1kni1mRNAcompletecds	phaw_50.283866:9026088-9035313	+	phaw_50.283866:9025431-9026438	phaw_50.283866:9,026,082-9,026,813	phaw_50.283866:9,034,022-9,037,291	Yes	Both
KNIRPS2kni2mRNAcompletecds	phaw_50.283866:9982901-10018092	-	phaw_50.283866:10018846-10019492	phaw_50.283866:10,017,879-10,018,098	phaw_50.283866:9,982,950-9,984,746	Yes	Exon
Kruppel5raceclone5	phaw_50.004430:12895048-13017154	+	phaw_50.004430:12893620-12894503	phaw_50.004430:12,893,830-12,895,369	phaw_50.004430:13,014,676-13,017,950	Yes	Both
notchproteinmRNAcompletecds	phaw_50.283815c:31626720-31812600	-	phaw_50.283815c:31812314-31813325	phaw_50.283815c:31,812,341-31,812,602	phaw_50.283815c:31,625,065-31,629,214	No	None
odd1	phaw_50.000289b:10469078-10470818	-	phaw_50.000289b:10470131-10471455	phaw_50.000289b:10,470,089-10,470,823	phaw_50.000289b:10,469,064-10,470,057	Yes	Exon
odd3	phaw_50.000289b:9790462-9792439	+	phaw_50.000289b:9789912-9790964	phaw_50.000289b:9,790,460-9,790,970	phaw_50.000289b:9,791,530-9,795,846	Yes	Exon
odd5	phaw_50.000289b:10003623-10007578	-	phaw_50.000289b:10007157-10007887	phaw_50.000289b:10,007,151-10,007,894	phaw_50.000289b:10,001,868-10,005,067	Yes	Both
opa1consensus	phaw_50.000135f:28969563-29076765	+	phaw_50.000135f:28968467-28970375	phaw_50.000135f:28,969,544-28,970,470	phaw_50.000135f:29,072,172-29,076,827	Yes	Both
optixmRNAcompletecds	phaw_50.283817f:8745320-8792652	+	phaw_50.283817f:8744687-8746467	phaw_50.283817f:8,745,412-8,746,075	phaw_50.283817f:8,788,522-8,794,388	Yes	Both
Par6	phaw_50.283823c:3497885-3548486	-	phaw_50.283823c:3548118-3548740	phaw_50.283823c:3,548,186-3,548,556	phaw_50.283823c:3,492,657-3,499,147	Yes	Both
Pax371proteinmRNAcompletecds	phaw_50.000214e:5044320-5078499	-	phaw_50.000214e:5078053-5078983	phaw_50.000214e:5,078,120-5,078,455	phaw_50.000214e:5,042,664-5,047,700	Yes	Both
pdm	phaw_50.283869d:4505837-4506191	+	phaw_50.283869d:4420193-4421142	phaw_50.283869d:4,420,800-4,421,678	phaw_50.283869d:4,506,146-4,509,682	Yes	Both
PhDIllate1	phaw_50.282654b:30431942-30563945	+	phaw_50.282654b:30429539-30431011	phaw_50.282654b:30,430,577-30,432,821	phaw_50.282654b:30,563,597-30,569,176	Yes	Both
PhDIllate2	phaw_50.282654b:30746554-30806154	+	phaw_50.282654b:30745371-30747232	phaw_50.282654b:30,746,377-30,747,054	phaw_50.282654b:30,805,338-30,813,233	Yes	Both
pnt3RACE	phaw_50.282861b:41123311-41125508	-	phaw_50.282861b:41129249-41130109	phaw_50.282861b:41,128,992-41,129,723	phaw_50.282861b:41,112,188-41,123,884	Yes	Exon
proboscipedia5raceloneI192	phaw_50.282695a:27659540-27817538	+	phaw_50.282695a:27657466-27658309	phaw_50.282695a:27,657,840-27,659,994	phaw_50.282695a:27,818,947-27,821,498	Yes	Both
prosperoproteinmRNApartialcds	phaw_50.283875a:6013042-6035299	+	phaw_50.283875a:5998087-5999786	phaw_50.283875a:5,998,945-6,001,387	phaw_50.283875a:6,034,691-6,038,907	Yes	Intron
rho5RACEvariant1	phaw_50.000203b:25395477-25496923	+	phaw_50.000203b:25395456-25395780	phaw_50.000203b:25,395,472-25,395,662	phaw_50.000203b:25,496,406-25,503,322	Yes	Both
rho5RACEvariant2	phaw_50.000203b:25265811-25496923	+	phaw_50.000203b:25265169-25265874	phaw_50.000203b:25,264,762-25,265,904	phaw_50.000203b:25,496,405-25,503,316	Yes	Both
runt2	phaw_50.283028b:1783456-1810535	-	phaw_50.283028b:1809649-1811397	phaw_50.283028b:1,809,475-1,810,690	phaw_50.283028b:1,776,166-1,784,663	Yes	Exon
scallopedproteinsdgene	phaw_50.007301a:763757-786495	-	phaw_50.007301a:920698-921986	phaw_50.007301a:921,178-921,537	phaw_50.007301a:762,031-762,973	Yes	Both
scratchscrtmRNAcompletecds	phaw_50.283869f:2146785-2165201	-	phaw_50.283869f:2270072-2271735	phaw_50.283869f:2,270,060-2,271,025	phaw_50.283869f:2,146,761-2,150,484	Yes	Both
Sexcombsreduced5raceloneE17c1	phaw_50.282695a:26769068-26862063	+	phaw_50.282695a:26768486-26770008	phaw_50.282695a:26,861,434-26,862,081	phaw_50.282695a:26,768,947-26,770,207	Yes	Both
shortgastrulationproteinmRNAcompletecds	phaw_50.015400b:7523582-7867524	+	phaw_50.015400b:7522832-7524249	phaw_50.015400b:7,523,573-7,525,453	phaw_50.015400b:7,866,925-7,867,874	Yes	Both
slp1	phaw_50.283811:7718478-7720485	+	phaw_50.283811:7718255-7718898	phaw_50.283811:7,718,433-7,720,788	phaw_50.283811:7,725,774-7,731,670	Yes	Both
snail2proteinmRNApartialcds	phaw_50.283864b:718097-721779	+	phaw_50.283864b:713312-713863	phaw_50.283864b:713,632-713,751	phaw_50.283864b:717,219-721,810	Yes	Both
snail3proteinmRNApartialcds	phaw_50.283864b:30526-33704	+	phaw_50.283864b:25948-26939	phaw_50.283864b:26,453-26,990	phaw_50.283864b:29,154-33,680	Yes	Both
Sp69protein	phaw_50.282861e:14546239-14663576	+	phaw_50.282861e:14544573-14546375	phaw_50.282861e:14,545,862-14,546,358	phaw_50.282861e:14,667,910-14,675,477	Yes	Both
spi5RACElargefragment3RACE	phaw_50.000081a:2326730-2344027	+	phaw_50.000081a:2243466-2244527	phaw_50.000081a:2,243,861-2,244,179	phaw_50.000081a:2,347,070-2,349,001	Yes	Both
SuH	phaw_50.283826:2280509-2340463	-	phaw_50.283826:2416373-2417483	phaw_50.283826:2,416,066-2,416,895	phaw_50.283826:2,270,773-2,276,731	Yes	Exon
unc4NPhsequence	phaw_50.000135e:7542170-7621463	+	phaw_50.000135e:7541550-7542771	phaw_50.000135e:7,542,144-7,542,554	phaw_50.000135e:7,644,646-7,649,912	Yes	Exon
wnty	phaw_50.283866:2773623-3090108	-	phaw_50.283866:3089459-3090541	phaw_50.283866:3,089,558-3,090,096	phaw_50.283866:2773623-3090108	Yes	Both

Supplementary Table 1.3: Software version numbers

Software name	Version	Usage
bamtools	2.5.1	Bam file manipulations
bedtools	2.28.0 or 2.30.0	Bed file manipulations
Bowtie2	2.3.0 or 2.3.4.1	Read alignment
BUSCO	3.0.2	Transcriptome and genome completion evaluation
cutadapt	2.4	Removing sequencing adapters from reads
deeptools	3.3.1	Visualization of data genome-wide
DESeq2 (R)	1.34.0	Differential accessibility/ expression analyses in pairwise comparisons
eggNOG-mapper	2.0.5	Automated gene function assignment
FASTQC	0.11.7	Library quality assessment
Genrich	0.6	Omni-ATAC peak calling
GMAP	2020-11-20	Aligning transcriptomes to genome for Mikado analyses
HISAT2	2.1.0	Aligning RNA_Seq reads to genome
igvtools	2.3.98	File format conversions for viewing in IGV
ImpulseDE2 (R)	0.99.10	Performing IDE2 analyses
JASPAR2020 (R)	0.99.10	Extracting PWMs from JASPAR
kallisto	0.43.1	Read abundance estimation for RNA-Seq
Mfuzz	2.54.0	Fuzzy clustering of peak accessibility
Mikado	2.3.0	Mikado transcriptome merging
minimap2	2.18-r1052-dirty	Aligning long-read Nanopore sequences to *Parhyale* genome
NucleoATAC	0.2.1	Prediction of nucleosome positions using Omni-ATAC data
OrthoFinder	2.5.4	Orthology assignment between *Parhyale* and *Drosophila*
PicardTools	2.9.0	Deduplication of reads
Portcullis	1.1.2	Identifying valid splice junctions from RNA-Seq read data
Python	3.7 or 3.8.8	Interfacing with Jupyter notebooks and running Python scripts on a SLURM scheduler
R	4.1.1 or 4.0.3 or 3.8	Performing IDE2, DESeq2 analyses; converting JASPAR files to RGT-HINT format
RGT-HINT	0.13.1	Inference of transcription factor footprints
samtools	1.8	Quality filtering (q=10) and sorting of bowtie2-mapped reads
StringTie	2.1.7	Assembly of StringTie2 L and StringTie2 SL transcriptomes
Transdecoder	5.0.2	Generating peptide sequences for OrthoFinder, Mikado, and other analyses
trim_galore	0.4.4	Trimming reads used in all analyses
Trinity	r2013_08_14	Assembling Trinity-limb (old) transcriptome
Trinity	2.5.1	Assembling Trinity-all, Trinity-gg, Trinity-S13, Trinity-S19, Trinity-S21, Trinity-S23, Trinity-limb (new) transcriptomes
universalmotif (R)	1.12.1	Converting JASPAR PWMs to RGT-HINT compatible format
VISTA	web VISTA	Identifying sequence homology between *Parhyale* and *Hyalella*

1.5 – Supplementary Methods, Chapter 1

S21-S22 sample outlier properties

For several metrics, the data we generated for samples at stages S21 and S22 appeared to show outlier effects, such as strong separation in PC2 of PCA analysis (Fig. 2C), a lower percentage of mapped reads (Supp. Fig. 2.1B), and lower inter-sample Spearman correlation (Supp. Fig. 2.1I). These data could be a reflection of variability in library quality, or could represent genuine biological differences with these samples. The shape of the PCA curve in Fig. 2C is consistent with a previously-described "horseshoe effect", "Guttman effect", or "arch effect" that is observed in microbial ecology datasets where a single strong gradient drives separation between samples in the dataset (Legendre and Legendre, 2012; Morton James T. et al.; Podani and Miklós, 2002). In the case of our dataset, we expect to find a single strong gradient – developmental time – which should separate out our samples.

To assess the degree to which PC2 in our data is driven by separation of S21/S22 samples, we removed the S21/S22 samples from our dataset and performed PCA. We observed that the relative positions of the remaining libraries remains largely similar (Supp. Fig. 2.2A, B), with time points closer to the middle of development showing a positive loading in PC2, and time points closer to the beginning and end of development showing a negative loading. This suggests that the second major axis of variation in our dataset would remain a contrast between middle vs. terminal timepoints, even without the S21/S22 data, and that the relative positioning of the remaining data within PC-space is not entirely driven by S21/S22.

To further assess the degree of the S21/S22 samples' outlying effects, we also performed ImpulseDE2 analysis to generate model fits without S21/S22 data. Doing so allowed us to determine to what degree the S21/S22 stages are necessary for driving the accessibility trajectory of individual peaks, and of the data more broadly. We performed IDE2 with either all data, or the S21/S22 data removed prior to input into IDE2. This generated two sets of model fits to the "cloud" of accessibility vs. time measurements: one that included the S21/S22 data, and one without. We evaluated, for each peak in our dataset, the time point at which the IDE2 model achieved maximum accessibility (the "IDE2 max fit"), and plotted both the "all" and "noS21S22" data as a histogram (Supp. Fig. 2.2C). The presence of peaks that achieve predicted maximum accessibility in the S21/S22 stages in the "no S21/S22" data is a result of how we calculate "max fit", which does not require that there is a known accessibility value at a given timepoint; only that the time point during which the model fit is maximum is closest to the timing of that developmental stage. Overall, we still observed early, middle, and late enrichment of IDE2 max fit even when the S21/S22 data are removed. We do see a rightward shift in the middle timepoint histogram in the direction of later stages, although this may be expected given the absence of concrete accessibility values at S21/S22 in the "no S21/S22" data. This indicates that our data globally retain the general trends of early, middle, and late enrichment of accessibility in the absence of the S21/S22 data. Moreover, this suggests that, even without the S21/S22 data, the remaining data from early and late stages result in a model fit that still predicts maximum accessibility at middle developmental stages for many peaks.

To further measure the influence of the S21/S22 data in IDE2 model fit, we also evaluated the degree of change in the general accessibility of a peak when the S21/S22 stages were removed. This analysis aimed to assess whether removing S21/S22 data resulted in an IDE2 model with the same

general trajectory as with all data, as opposed to the more stringent requirement of evaluating whether the exact developmental stage of the peak was changed. To perform this analysis, we grouped developmental stages into five quintiles, each representing three stages of development. We asked, for each peak in our dataset, whether that peak's IDE2 max fit was "stable" when the S21/S22 data were removed; that is, if the quintile of the IDE2 max fit was altered when the S21/S22 data were removed (i.e. if a peak moved more than 3 developmental stages away from its original position), a peak was considered "unstable" (Supp. Fig. 2.2D, E). We observed that over 80% of peaks in each quintile remained "stable" after removing the S21/S22 data, suggesting that the vast majority peaks show the same general trajectory of accessibility even without the S21/S22 data. Peaks within the middle time points appeared to be more unstable than peaks at the terminal timepoints, which could be expected given that the S21/S22 timepoints constituted the middle-most timepoints in our dataset.

It is difficult to determine whether a change in accessibility trajectory for a given peak caused by the removal of S21/S22 data is indicative of technical differences in sample preparation, such as batch effects; biological variation, such as a potentially unknown mutant or sick embryo; or due to genuine wildtype biological processes that occur at the S21/S22 stages. These caveats acknowledged, a comparative analysis of the data in the absence of the S21/S22 stages suggests that much of the global picture of development remains the same.

To facilitate future users of this dataset, we have included the model parameters calculated from IDE2 using both the full dataset and the data with S21/S22 removed, as well as a Jupyter notebook (ParhyaleATACExplorer.ipynb) that allows users to plot the raw accessibility data and IDE2 model fits with respect to time, so that downstream experiments can consider the potential differences with the S21/S22 samples.

Trinity transcriptome assembly parameters

Trinity *de novo* transcriptome (all stages) was assembled with --max_memory 1500G and --CPU 32. Illumina NovaSeq short reads were supplied as a sample file, where each sample had two read files for read1 and read2. Read files were generated by first trimming transcripts using trim_galore, and then combining the val_1 reads with unpaired_1, and val_2 reads with unpaired_2 for each stage using cat. The same merged read files were used for stage-specific assemblies. For the genome-guided assembly, *in silico* normalized reads generated by Trinity for read1 and read2 across all stages were aligned to the genome using hisat2 --dta and then sorted using samtools. Trinity was run as Trinity --genome_guided_bam <bam file> --genome_guided_max_intron 300000.

StringTie2 transcriptome assembly parameters

Nanopore reads, when generated, are partitioned into chunks in different folders in the output. Reads were compiled by barcode. All reads were combined together into a single file and mapped to the genome using minimap2 -ax splice using the phaw_5.0.fa genome file as a target. StringTie2 L assembly was run using stringtie -L <nanopore_reads.bam>, while StringTie2 SL assembly was run using stringtie -mix <illumina_reads.bam> <nanopore_reads.bam>. Both StringTie2 assemblies used default settings.

Mikado setup parameters

Trinity transcriptomes were aligned to the phaw_5.0.fa genome using GMAP with --max-intronlength-ends=300000 --format=gff3_gene. All HISAT2 aligned read BAM files from Illumina RNA-Seq were fed into Portcullis using portcullis full --force --copy --verbose -t 32. Due to yet-to-be-resolved errors in the Portcullis package, we manually converted some assignments of the portcullis output using the following commands:

```
$ cat portcullis_out/2-junc/portcullis_all.junctions.tab |awk '!(($11=="?" && $14=="NA") ||
    ($11=="?" && $13=="NA"))' > portcullis_out/2-junc/portcullis_all.junctions.cleaned.tab
$ portcullis filter --verbose portcullis_out/1-prep
    portcullis_out/2-junc/portcullis_all.junctions.cleaned.tab
```

The scoring parameters and other variables for the Mikado configure file are found in the "20210630_list.txt" file. We applied penalties to the Trinity-based transcriptomes and bonuses to the StringTie2 transcriptomes based on our observations that Trinity transcripts, when aligned to the genome, showed many apparently spurious short transcripts at the RACE genes we examined.

For mikado configure, we used --mode permissive --scoring mammalian.yaml --copy-scoring mammalian.yaml -bt uniprot_sprot.fasta. We reasoned that the large genome size of *Parhyale* would be appropriate to analyze using the mammalian settings as opposed to the other available settings.

For the Transdecoder step, we ran the analysis with default settings. We omitted BLAST from our Mikado pipeline, as the extremely large file size of our mikado_prepared.fasta file proved intractable to straightforward BLASTX analysis. We estimated that it would take around 5 months to run using our current computing approaches. An updated Mikado transcriptome could be generated by breaking up the mikado_prepared.fasta file into many smaller BLASTX analyses, which could run much faster. The required files for further Mikado analyses are available in the GEO accession data.

eggNOG annotation

We generated a .pep file of the final Mikado transcriptome or the MAKER genome annotation using Transdecoder, and used emapper.py with default settings and -m diamond.

OrthoFinder annotation

Using a .pep file generated from the final Mikado transcriptome or the MAKER genome annotation using Transdecoder, we ran OrthoFinder to compare each transcript source to the database of all *Drosophila melanogaster* proteins found in the UNIPROT database.

DESeq2 parameters

For differential expression and accessibility analyses, we performed pairwise DESeq2 runs for adjacent pairs of developmental stages (e.g. S19 vs. S13, S21 vs. S19). Analysis was performed using standard settings. Peaks or genes were considered differentially expressed when padj < 0.05.

ImpulseDE2 analysis

We performed ImpulseDE2 analysis using standard settings, using the matrix of reads generated from bedtools multicov and used for DESeq2 analysis. For the Time parameter, we used the number of

hours for each developmental stage, as indicated in the Browne et al. *Parhyale* staging guide (Browne et al., 2005). The time points were: 72, 77, 80, 87, 90, 96, 104, 112, 120, 132, 144, 155, 168, 180, and 192 hours, representing stages 'S13', 'S14', 'S15', 'S17', 'S18', 'S19', 'S19+', 'S20', 'S21', 'S22', 'S23', 'S24', 'S25', 'S26', 'S27', respectively. We extracted the IDE2 model fits for each peak (available at the GEO accession) as a .tsv file, used a custom Python function to plot accessibility trajectories (will be available at the GitHub repository). To generate such plots, we recreated the accessibility plotting function from the ImpulseDE2 package in R as two Python functions, shown below:

```
def ImpulseFunction(t, vecImpulseParam):
  return (1/vecImpulseParam[3]) * (vecImpulseParam[2] + (vecImpulseParam[3] -
    vecImpulseParam[2]) * (1/(1 + np.exp(-vecImpulseParam[1] * (t - vecImpulseParam[5])))) *
    (vecImpulseParam[4] + (vecImpulseParam[3] - vecImpulseParam[4]) * (1/(1 +
    np.exp(vecImpulseParam[1] * (t - vecImpulseParam[6])))))

def ImpulseEvaluate(vecTimepoints, vecImpulseParam):
  vecImpulseValue = [ImpulseFunction(i, vecImpulseParam) for i in vecTimepoints]
  vecImpulseOutput = [i if i > 10^(-10) else 10^(-10) for i in vecImpulseValue]
  return vecImpulseOutput
```

JASPAR2020 database

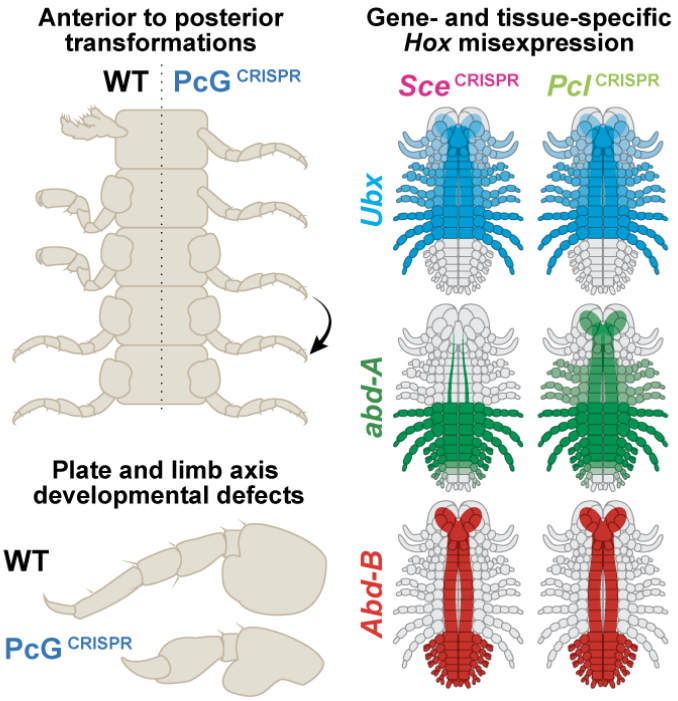
For RGT-HINT analyses, the software expects each motif as a simple matrix representing position weights at each base for A, T, G, and C. This format is not readily available for the JASPAR2020 database, which utilizes the JASPAR format. We downloaded the full JASPAR2020 CORE non-redundant database, split the database into individual JASPAR format files for each transcription factor, and then converted that JASPAR format tile into a plain matrix file using the "universalmotif" R package. We also utilized the metadata extracted from this package to generate an .mtf file as required by RGT-HINT. All files for this type of analysis will be available on GEO at time of publication.

2 – Distinct regulation of *Hox* genes by Polycomb Group genes in a crustacean

Dennis A Sun[1], Yuri Takahashi[1], Rebecca J Chang[1], and Nipam H Patel[2,3*]
1. Department of Molecular and Cell Biology, University of California, Berkeley
2. Marine Biological Laboratory, Woods Hole, MA
3. Department of Organismal Biology and Anatomy, University of Chicago, Chicago, USA
* corresponding author

Abstract

Non-insect crustaceans exhibit tremendous body plan diversity. The evolution of diverse patterns of Hox gene expression has been implicated as a primary driver of body plan evolution between crustacean groups, but the mechanisms underlying Hox regulatory evolution remain unknown. We identify Polycomb and Trithorax Group proteins, crucial for proper Hox regulation across bilaterians, in the genome of the amphipod crustacean Parhyale hawaiensis, and demonstrate their essential functions in crustacean Hox regulation and embryonic development using CRISPR-Cas9 mutagenesis. Examination of Hox misexpression patterns between individual Hox genes with respect to timing, tissue, and PcG mutant gene in mutant embryos reveals that each crustacean Hox gene follows its own idiosyncratic regulatory mechanism. These results suggest a distinct regulation of Hox genes that may have enabled crustacean body plan evolution.



Primary Findings

- The genome of the amphipod crustacean *Parhyale hawaiensis* contains all core Polycomb Group (PcG) and Trithorax Group (TrxG) proteins
- CRISPR-Cas9 mutagenesis of PcG proteins induces homeotic transformations and misexpression of *Hox* genes that differ from similar experiments in insects
- PcG knockout embryos show proper initiation of *Hox* expression boundaries at early developmental stages

- Each of the three posterior *Hox* genes in *Parhyale* displays distinct patterns of misexpression in response to PcG knockout
- *Hox* regulation appears to occur via different mechanisms in the nervous system vs. limbs
- PcG phenotypes reveal the potential for distinct layers of *Hox* regulation in crustaceans

2.0 – Introduction

Non-insect crustaceans are a group of organisms with tremendous variation in body organization. Unlike hexapods or chelicerates, which stereotypically only have head and thoracic appendages, crustaceans generally have one appendage for every body segment (Angelini and Kaufman, 2005; Hughes and Kaufman, 2002). Moreover, unlike the limbs found in myriapods (millipedes and centipedes), which are largely serially identical along the body axis, crustaceans have a tremendous variety of limbs for different functions, such as feeding, grasping, walking, or swimming.

Individual groups of crustaceans have evolved different arrangements of these appendages, which are referred to as bauplans, or body plans (Averof and Akam, 1995; Deutsch and Mouchel-Vielh, 2003). Some of these body plans are familiar to most people, such as the decapod body plans of shrimp, crabs, and lobsters, or the isopod body plan of pill bugs and woodlice (Schwentner et al., 2017, 2018; Wolfe et al., 2019). Beyond those with familiar body plans, many other crustacean groups exist, including remipedes, fairy shrimp, and shield shrimp, each with its own unique arrangement of limbs along the anterior-posterior axis. As a result of these Swiss Army knife-like bodies, crustaceans are able to survive in numerous aquatic and terrestrial environments.

The evolution of crustacean body plan diversity is hypothesized to be a consequence of variations in the expression of Homeotic (Hox) genes, a family of genes crucial for establishing identity along the anterior-posterior axis in diverse animal groups (Averof and Akam, 1995; Averof and Patel, 1997; Hrycaj and Wellik, 2016; Krumlauf, 1994; Panganiban et al., 1995). For example, shifts in the expression domain of the Hox gene Ultrabithorax have been correlated with the emergence of novel identities across the crustacean tree (Averof and Patel, 1997). These results have suggested that crustacean Hox regulatory network evolution may be a driving force in crustacean body plan evolution.

Recent experiments in the amphipod crustacean Parhyale hawaiensis have begun to directly reveal the function of Hox genes in patterning crustacean body plans. Using CRISPR-Cas9 mutagenesis, RNA interference, and heat shock-induced misexpression, previous work has characterized the expression (Liubicich et al., 2009; Serano et al., 2016) and function (Liubicich et al., 2009; Martin et al., 2015; Pavlopoulos et al., 2009) of each of the Parhyale Hox genes (summarized in Supp. Fig. 1.1D).

For example, the Hox gene Ultrabithorax (Ubx) is expressed in most of the thoracic appendages (T2-8) in Parhyale: at lower levels in the claws (T2/3) and higher levels in the walking legs (T4/5) and jumping legs (T6-8). RNAi knockdown and CRISPR knockout of Ubx reveal that this gene is necessary for the forward walking leg (T4/5) identity, and overexpression experiments have revealed that ectopic Ubx expression can induce transformations of anterior body segments towards T4/5 walking leg identity (Supp Fig. 1.1F). These results indicate that Ubx expression is both necessary and sufficient for establishing T4/5 forward walking leg identity.

Moreover, knockout experiments examining multiple Hox genes in Parhyale have indicated that Hox genes also regulate one another (Jarvis et al., 2022). For example, upon CRISPR-Cas9 knockout of the Hox gene Abdominal-B, normally found in the abdominal appendages (A1-6), expression of Ubx expands into the previous Abd-B domain, suggesting that Abd-B represses Ubx (summarized in Supp. Fig. 1.1E). Knockout of both Abd-B and Ubx results in expansion of Scr expression towards the posterior, inducing homeotic transformations to T1 identity in the regions lacking abd-A expression; thus, Ubx and Abd-B repress Scr. These results suggest that Hox cross-regulatory mechanisms are essential for proper body organization in Parhyale.

While these studies have revealed the function of Hox genes in establishing regional identity and in cross-regulation, the upstream mechanisms of crustacean Hox regulation remain unknown. In insects, early Hox regulation appears to be governed at two phases: early establishment of Hox boundaries by upstream transcription factors, and maintenance of proper Hox expression through later stages of development (Akam, 1987; Denell, 1978; Irish et al., 1989; Maeda and Karch, 2006, 2015; Puro and Nygrén, 1975; Wedeen et al., 1986). For example, in Drosophila, the gap and pair-rule gene networks activate or repress Hox genes through special cis-regulatory elements known as initiator elements, which establish active or repressive chromatin in specific regions of the genome for each body segment (Maeda and Karch, 2006, 2015; Peifer et al., 1987)(Supp. Fig. 1A). These chromatin marks are laid down and maintained by the activity of Polycomb Group (PcG) and Trithorax Group (TrxG) genes, two families of general transcriptional repressors and activators, respectively (Geisler and Paro, 2015; Kassis et al., 2017; Schuettengruber et al., 2017).

PcG and TrxG genes were first identified as Hox regulators in Drosophila due to the homeotic transformations induced by perturbation of PcG/TrxG function. For example, maternal-zygotic null embryos for many PcG genes exhibit homeotic transformations of all thoracic segments towards the most posterior identity (A8)(Breen and Duncan, 1986; Martin and Adler, 1993)(Supp. Fig. 1B-C). Within these mutants, most of the Drosophila Hox genes (such as Scr, Dfd, Antp, Ubx, abd-A, and Abd-B) appear to be broadly misexpressed anterior to their wildtype boundary and in a variety of different embryonic tissues, including the embryonic ectoderm, nervous system, and mesoderm (Simon et al., 1992; Wedeen et al., 1986). Despite this broad misexpression of many genes, it is proposed that the cross-regulatory repression of more anterior Hox genes by Abd-B results in a uniform transformation of all body segments to the most posterior identity (Breen and Duncan, 1986; Simon et al., 1992; Struhl and White, 1985).

In contrast, some TrxG genes were observed to induce homeotic transformations in the opposite direction – posterior to anterior – as a result of decrease or loss of proper Hox expression (Breen and Harte, 1991, 1993)(Supp. Fig. 1B). Others TrxG genes were identified through their dominant suppressor effects on PcG mutants, indicating their competing roles in Hox regulation (Fauvarque et al., 2001; Kennison and Tamkun, 1988).

These two families of genes appear to have globally similar functions in vertebrate Hox regulation. Loss of PcG or TrxG function in mice often results in Hox misexpression and homeotic transformations (Akasaka et al., 2001; Brinkmeier et al., 2015; Guenther et al., 2005; Isono Kyo-ichi et al., 2005; del Mar Lorente et al., 2000; Wang et al., 2007; Yu Benjamin D. et al., 1998), suggesting that the ancestral bilaterian Hox cluster was regulated by PcG/TrxG function. However, mouse Hox

regulation differs from that of Drosophila in that Hox expression is first initiated in a temporal sequence within a single tissue region (the posterior primitive streak), rather than being initiated in spatially separate regions (Deschamps and van Nes, 2005; Deschamps and Wijgerde, 1993; Kmita and Duboule, 2003; Mallo and Alonso, 2013; Wang et al., 2004). This early activation is then followed by an expansion of the expression domains towards their final expression patterns, a process not governed by cell migration, but instead by progressive changes to histone marks associated with PcG/TrxG function (Chambeyron and Bickmore, 2004; de Graaff Wim et al., 2003; Deschamps and van Nes, 2005; Deschamps and Wijgerde, 1993; Rastegar Mojgan et al., 2004; Soshnikova and Duboule, 2009). Thus, in vertebrates, the final expression pattern of Hox genes appears to be both established and maintained by PcG/TrxG function.

The molecular functions of PcG/TrxG complexes across organisms are diverse, but primarily occur via the chromatin modifications they establish. Multiple PcG complexes exist, but are broadly broken down into three groups: Polycomb Repressive Complex 1 (PRC1), Polycomb Repressive Complex 2 (PRC2), and Pleiohomeotic Repressive Complex (PhoRC) (Fig. 1A, adapted from Erokhin et al., 2018; Schuettengruber et al., 2017). Within these broader groups, additional sub-complexes have been characterized, such as PRC1 vs. dRAF, and PRC2.1 vs. PRC2.2, each of which has its own particular biochemical function. TrxG complexes are more diverse both in function and number, including BAP/PBAP, TAC1, COMPASS and COMPASS-like, and other complexes (Fig. 1A, adapted from Kassis et al., 2017). Many of the TrxG complexes directly reverse the biochemical function of PcG complexes; as such, the two groups of genes are often considered to function in opposition (Geisler and Paro, 2015).

In this paper, we reveal the role of PcG and TrxG genes in the regulation of Hox genes in the amphipod crustacean Parhyale hawaiensis. We demonstrate that all core members of these complexes are present in the Parhyale genome and expressed during embryonic development. Using CRISPR-Cas9 mutagenesis, we ablate PcG and TrxG function and identify homeotic transformations induced by PcG knockout, indicating that these genes are essential for Parhyale Hox regulation. In addition, we examine the effects of PcG knockout on the expression of Parhyale Hox genes using in situ hybridization chain reaction and immunohistochemistry. These data reveal that the Parhyale Hox genes appear to have idiosyncratic regulatory mechanisms with respect to developmental timing, tissue type, and specific PcG complex components. Our results provide the first mechanistic insights into the upstream regulatory architecture of the Parhyale Hox complex, and suggest that Hox genes in crustaceans may have evolved distinct regulation. Such distinct regulation offers a mechanistic explanation for how different crustacean groups were able to evolve distinct patterns of Hox expression to generate their dramatic body plan diversity.

2.1 – Results

2.1.1 – The *Parhyale* genome contains all core PcG/TrxG components

To identify candidate genes of interest for CRISPR-Cas9 mutagenesis, we examined the current literature on Drosophila melanogaster and Mus musculus PcG/TrxG function. We cataloged phenotypes of interest for a number of genes (Supp. Tables 1 and 2), and used reciprocal BLAST to

search for all core PcG/TrxG complex components in several Parhyale hawaiensis transcriptomes and genome annotations (see Methods). The data sources for each transcript, along with the gene ID in that transcript source, are listed in Supp. Tables 3 and 4.

We were able to identify all of the core PcG/TrxG genes we examined in the Parhyale genome. We observed that Parhyale has a single copy of extra sex combs and pleiohomeotic, in contrast to Drosophila, where additional gene duplications with redundant functionality (extra sex combs-like (esc-l) and pleiohomeotic-like (pho-l) appear to have evolved (Figure 1B)(Brown et al., 2003; Kurzhals et al., 2008; Ohno et al., 2008). In addition, we found four Parhyale genes that appeared to be related to the Drosophila genes Psc, Su(z)2, and l(3)73Ah.

Psc and Su(z)2 are two genes in Drosophila that are located adjacent to each other in the genome and which have semi-redundant functionality (Brunk et al., 1991; Lo Stanley M. et al., 2009). In Parhyale, we found four potential candidates for Psc and Su(z)2 (Fig. 1C, Supp. Fig. 2.1F, G). One of these candidates, which we named Phaw Psc-1, appeared to be the top BLAST hit for both Dmel Psc and Dmel Su(z)2. A second candidate, which we named Phaw Psc-2, appeared to be the second-best BLAST hit for both Dmel Psc and Dmel Su(z)2. For both Dmel Psc and Su(z)2, we also found a third BLAST hit, which we identified as Phaw l3(73)Ah based on BLAST to Dmel l3(73)Ah. Finally, the fourth-best BLAST hit to both Dmel Psc and Dmel Su(z)2 appeared to be another protein with a RING-HC PCGF and RAWUL PCGF2-like domain, similar to Psc (Supp. Fig. 2.1F, G). We named this gene Phaw Psc-like.

To determine whether each of these PcG/TrxG genes were expressed during embryonic development in Parhyale, we leveraged the recently described Mikado transcriptome, which also serves as an improved genome annotation (Sun et al., 2021). We identified the best Mikado transcript for each PcG/TrxG gene, and examined the expression of that transcript at four developmental stages (S13, S19, S21, S23; three replicates per stage) from a previously described transcriptome (Supp. Fig. 1.2)(data from Sun et al. 2021). Based on these data, all PcG/TrxG genes we examined appeared to be expressed at these four stages of development.

Based on the potential to induce homeotic phenotypes previously described in Drosophila, we selected a number of PcG and TrxG genes to perform CRISPR-Cas9 mutagenesis. For PcG, we designed CRISPR guide RNAs to the PRC1 genes Sce, Pc, Psc-1, Psc-2; the PRC2 genes Su(z)12, E(z), Esc, and Pcl; and the PhoRC gene pho. For TrxG, we designed CRISPR guide RNAs to the BAP/PBAP gene Brm; Kis; the COMPASS and COMPASS-like genes Ash2 and Trx; and three additional genes categorized as TrxG members: Fs1h, kto, and skd. For each gene, we designed guide RNAs targeting the 5'-most region of the RNA sequence for which BLAST evidence suggested homology to either Drosophila or another arthropod. We injected pairs of guide RNAs into 1- or 2-cell embryos and screened embryos for phenotypes using confocal and widefield microscopy.

2.1.2 – PcG KO induces homeotic transformations of anterior segments to T4/T5 walking leg identity

When we examined knockouts of PcG genes, we consistently observed striking homeotic transformations of anterior body segments towards T4/5 walking leg identity. Representative hatchlings from select CRISPR experiments are shown in Fig. 2, illustrating the homeotic

transformations we observed. We observed this phenotype in hatchlings for 7/9 of the genes we targeted using CRISPR (*Sce*, *Pc*, *Su(z)12*, *E(z)*, *Esc, Pcl*, and *pho*, see Table 1). We most frequently observed transformations of T2/3 claw appendages to T4/5 walking leg identity, but we also observed for some genes, such as *Pcl* and *Su(z)12*, transformations of antennae towards T4/5 walking leg identity. We also observed another class of apparent transformations of A1-3 swimming leg identities to A4-6 anchoring leg identities in *Pc* mutants, but at frequencies much lower than transformations to T4/5 walking leg identities. Finally, among all mutants exhibiting transformations towards T4/5 identity, we also observed defects in mouthpart morphology, where mouthparts appeared to exhibit "stubby" morphology not obviously analogous to any known wild type morphology. A summary of the phenotypes we observed can be found in Table 1.

The transformations of anterior segments to T4/5 identity are notable in that they appear identical to the homeotic transformations observed by heat shock misexpression of *Parhyale Ultrabithorax* (Pavlopoulos et al., 2009). In that previous study, anterior segments were frequently transformed towards T4/5 walking leg identities in heat-treated embryos and hatchlings. Based on our understanding of the role of each *Hox* gene in patterning the *Parhyale* body plan, we hypothesized that the homeotic phenotypes we observed in our PcG knockouts would be a result of misexpression of *Ubx*. These results suggest that PcG genes are necessary for proper *Hox* regulation in *Parhyale*.

2.1.3 – PcG genes are essential for proper limb proximal-distal axis development and plate formation

In addition to the homeotic phenotypes we observed, we also observed other types of developmental defects in CRISPR-targeted embryos. In particular, for 8/9 genes we examined (excluding *Psc-1*), we observed various non-homeotic defects in limb morphogenesis (see Table 1). For example, in addition to homeosis of T2/3 -> T4/5 in *Sce* mutant embryos, we also observed fusions or loss of leg segments (Fig. 2C, Fig. 3B, B'). This proximal-distal defect phenotype varied from hatchling to hatchling, and from limb to limb across the body plan. To assess the potential for off-target effects of our knockouts, we performed thorough phenotype analyses of hatchlings from CRISPR knockouts using single guides of three genes:*Sce*, *Pcl*, *Su(z)12*. For all three genes, we were able to recover all of our cataloged phenotypes for each guide RNA (Supp. Figs. 2.2, 2.3, 2.4).

In addition, we observed seemingly improper cuticle deposition, resulting in wrinkled-looking appendages, bulbous limbs, twisted limbs, and other general morphological defects (Fig. 3, Supp. Figs. 2.1, 2.2, 2.3, 2.4, 2.5). These results indicate that PcG genes are necessary for proper development of limbs along the proximal-distal axis, potentially interfacing with the leg gap gene system described in (Bruce and Patel, 2020; Clark-Hachtel and Tomoyasu, 2020), and that loss of PcG function can induce general morphogenic irregularities.

We also observed defects in the development of tergal, coxal, and basal plates in a number of PcG knockout hatchlings. For example, for all genes other than *Psc-1* and *Psc-2*, we observed reduced coxal, tergal, and basal plates (Fig. 3B, B', D, D'). These phenotypes were reminiscent of previous phenotypes induced by knockout of wing developmental genes such as *vestigial* and *scalloped* (Clark-Hachtel and Tomoyasu, 2020). For many PcG genes, we also observed total loss of plate morphogenesis, such as loss of the entire T2 tergal and coxal plates, which usually present with a

stereotypical triangular shape (Fig. 3C, C'). In other cases, we observed "floating" tergal plates, where the left and right halves of the hatchling failed to properly close on the dorsal side of the embryo (Fig. 3E). Some of these "floating" plates appeared to be caused by improper pairing of the left and right halves of the embryo (Fig. 3F, F', F''). For example, Figure 3F' shows a hatchling where the left T2 plate has joined with the right T3 plate. These results suggest that PcG genes may also play a role in regulating the process of dorsal closure.

The phenotypes we observed appear to affect body segments in more restricted domains than those governed by *Hox* regionalization alone. For example, the T2 segment, in which we frequently observed total loss of the tergal and coxal plates, normally develops into a clawed appendage. The T3 segment also develops a clawed appendage, but we did not observe total loss of T3 tergal and coxal plates in our mutants. These results reveal potential segmental identities unique to individual numerical body segments, regulated by a more specific mechanism than that of *Hox* regional identity, in which PcG genes play some important role.

2.1.4 – *Parhyale Psc-2*, but not *Psc-1*, is essential for proper embryonic development

Parhyale Psc-1 and *Psc-2* were the only ambiguously identified orthologs of PcG or TrxG genes found in our examination of the *Parhyale* genome. CRISPR-Cas9 mutagenesis of *Psc-2* caused major developmental defects, including apparent fusion of body segments and dramatic truncation of limbs (Supp. Fig. 2.1). Most *Psc-2* hatchlings did not hatch on their own, and required dissection from the egg at late developmental stages. CRISPR treatment with guides targeting *Psc-1* produced no observable phenotypes. Thus, *Psc-2* is indispensable for embryonic development, and likely serves a role in PcG complex function, while *Psc-1* does not appear to be essential for development.

2.1.5 – TrxG KO induces embryonic lethality, major developmental defects, or no phenotype; but not homeosis

In contrast to the consistent phenotypes observed for PcG knockouts, TrxG knockout hatchlings displayed a wider variety of phenotypes including embryonic lethality, major developmental defects, or no visible phenotype. We also did not observe any obvious homeotic transformations in TrxG knockout hatchlings.

Knockouts of *Brm* and *Fs1h* in *Parhyale* appeared to result in embryonic lethality (Supp. Fig. 2.6D-D', E). None of the *Fs1h* knockout embryos survived long enough to hatch, and only a small number of *Brm* embryos survived to late developmental stages. In the few *Brm* hatchlings we recovered, it appeared that approximately half of the embryo was deleted (Supp. Fig. 2.6D). This phenotype is consistent with the loss of one-half of the embryo starting at the 2-cell stage, based on previous lineage tracing experiments (Gerberding et al., 2002; Price et al., 2010), and suggests that *Brm* is essential for cell viability during *Parhyale* development.

Knockouts of *kto* and *skd*, members of the Mediator complex, showed major morphological defects such as compaction along the anterior-posterior axis, including fusion between adjacent body segments and defects in dorsal closure (Supp. Fig. 2.6A-A'', B-B''). Careful analysis of *kto* and *skd* mutant limbs did not reveal any obvious homeotic transformations. *Ash2* knockout hatchlings showed occasional defects in plate development, with reduced plates reminiscent of PcG mutants, as well as

twisted limbs, but did not appear to show any obvious homeotic transformations (Supp. Fig. 2.6C-C'').

Trx and *Kis* knockout embryos did not show any obvious phenotypic differences from WT embryos with the first pair of guide RNAs we used for each gene (Supp. Fig. 2.7). We designed additional pairs of guide RNAs targeted towards the 5' end of conserved domains of these genes that have previously been characterized as being important for gene function. In the case of *Trx*, careful analysis of gene models from several different transcriptomes revealed that the *Trx* gene may contain two separate, non-contiguous ORFs. The 5' ORF appeared to contain the conserved PHD domain and FYRN domains found in the 5' end of *Drosophila Trx*, while the 3' ORF appeared to contain FYRC and SET domains found in the 3' end of *Drosophila Trx* (Supp. Fig. 2.7C). When we identified *Trx* in the Mikado transcriptome, the 5' and 3' ORFs were annotated as two separate genes, which we referred to as *Trx (DBD)* and *Trx (SET)*. Quantification of the expression of each of these two transcripts across a developmental RNA-Seq timecourse (from Sun et al., 2021) revealed that the two ORFs appeared to have differential expression over time (Supp. Fig. 2.7B). To account for the possibility of the *Trx (SET)* ORF performing a function separate from the *Trx (DBD)* ORF, we also designed a pair of guides specifically targeting the *Trx (SET)* ORF. In total, we designed 8 guides targeting *Trx* and 4 guides targeting *Kis*. We did not observe any obvious homeotic phenotypes for any of the guides we tested for these genes. Hatchlings overall showed WT morphology or mild morphological defects (such as twisted body axis) which have previously been observed in sham injection experiments (Schmid, 2011).

Our experiments were unable to reveal any obvious role of TrxG in *Hox* regulation in *Parhyale*. It is possible that the homeotic effects of TrxG knockout are more subtle than the limb transformations observed in PcG knockouts; deeper analysis will be necessary to define the morphological landmarks of such phenotypes, such as the exact pattern of setae (sensory hairs) and other fine structures which may be specified via *Hox* mechanisms (such as was performed in Almazán et al., 2021). Additional experiments, such as using RNAi to induce gene knockdown for essential genes, or CRISPR mutagenesis of multiple genes simultaneously, could also help more clearly reveal whether TrxG genes play any role in *Hox* regulation in this organism.

2.1.6 – PcG gene knockout induces *Hox* misexpression during limb developmental stages

Based on the homeotic phenotypes we observed, we hypothesized that *Ubx* would be misexpressed in anterior segments in PcG mutants. To test this hypothesis, we performed *in situ* hybridization chain reaction (Bruce et al., 2021; Choi et al., 2018) on CRISPR-treated embryos for *Parhyale Ubx*, along with other *Hox* genes, including *abd-A* and *Abd-B*. We performed HCR on dissected embryos treated with CRISPR guides targeting *Sce* and *Pcl* at developmental stage 23, during which limb morphogenesis is underway. For both genes, we observed strong misexpression of *Ubx* in anterior body segments (Fig. 4A, B, C; *Ubx* panel), as predicted by the observed homeotic transformations. Thus, PcG knockout induces ectopic expression of *Ubx*, resulting in homeotic transformations of anterior body segments to T4/T5 identity.

We also examined the expression of *abd-A* and *Abd-B* in mutant embryos (Fig. 4A, B; *abd-A* and *Abd-B* panels). At late developmental stages, the tissue location and degree of misexpression of

abd-A and *Abd-B* differed when compared to *Ubx*. For example, in *Sce* mutant embryos, whereas *Ubx* misexpression appear to expand anteriorly broadly across all tissues, *Abd-B* expression expanded anteriorly, but appeared to be restricted to the medial region of the body, where the neurogenic ectoderm is located. *abd-A* in *Sce* mutant embryos appeared to be only misexpressed in the nervous system in a few cells per body segment. These results suggest that *Sce* is necessary for repression of *Abd-B* in the nervous system, but not in the limbs, and that *Sce* weakly represses *abd-A*.

We also examined the expression of two other *Hox* genes, *Scr* and *Antp*, in mutant embryos at S23 (Fig. 5A, B). *Scr* in WT embryos is expressed in the Mx1, Mx2, and T1 segments. In mutant embryos, *Scr* appeared to be expressed within the same regions, but with a notable reduction in expression in the T1 appendage. In the Mn segment, we observed patches of ectopic *Scr* expression. *Scr* expression also appeared to exhibit trace misexpression in further anterior structures, with a few transcriptional spots in the brain and antennae (Fig. 5B, *Scr* channel, arrows). When we compared the domain of *Scr* expression to *Ubx* expression, we observed that *Ubx* expression appeared to expand into the T1 appendage, suggesting a mechanism for the reduction of *Scr* expression in that limb. In a previous study (Pavlopoulos et al., 2009), heat shock overexpression of *Ubx* was associated with decreased or nearly abolished *Scr* expression within its normal domain, providing an explanation for the reduction in *Scr* expression in T1 as *Ubx* expression expands into that segment, as well as a lack of strong expression in more anterior segments, where *Ubx* is strongly misexpressed.

However, in PcG knockout embryos, *Scr* expression is retained in Mx1 and Mx2 (Fig. 5B). Closer examination of the patterns of *Scr* and *Ubx* expression within the same embryo revealed that *Scr* expression in PcG mutants is maintained in the posterior half of the Mx1 and Mx2, and missing in the anterior half of the limbs. In these appendages, *Ubx* is expressed in a nearly mutually-exclusive domain when compared to *Scr*. This phenotype can be explained if *Scr* loss in the anterior half of the Mx1 and Mx2 limbs is also mediated by *Ubx* repression of *Scr*.

In addition, these results suggest that the posterior half of the Mx1 and Mx2 appendages may lack some activating factor necessary for PcG derepression-mediated *Ubx* misexpression. These findings of mosaic *Hox* expression in mouthparts may also explain the relatively infrequent occurrence of mouthpart -> T4/5 transformations we observed in our knockout experiments, as well as the stubby or malformed morphology of mouthparts in most PcG mutants, which did not clearly resemble any known limb type.

To examine *Antp* expression, we used the cross-reactive mouse anti-Antp antibody (8C11) rather than *in situ* HCR. A challenge of *in situ* hybridization for detecting *Antp* in crustaceans is that crustaceans express a hybrid *Antp-Ubx* transcript in limbs across the broad region overlapping both *Antp* and *Ubx* expression, in addition to broad expression of *Antp* in the nervous system (Serano et al., 2016; Shiga et al., 2006). Antibody staining has previously revealed that functional Antp protein is only produced in the T2/3 appendages and in the nervous system – thus, examination of *Antp* expression requires direct detection of Antp protein. Using 8C11, we observed that the anterior boundary of *Antp* expression was perturbed in mutant embryos (Fig. 5C, D). In WT embryos, Antp protein in the nervous system is bounded in the anterior at the Mx2 segment. However, we observed strong Antp expression in the nervous system throughout the entire body axis in *Sce* and *Pcl* mutant embryos.

Together, these results indicate that PcG genes are essential for proper repression of *Hox* expression in *Parhyale*, as loss of PcG expression induces ectopic *Hox* expression for all of the five *Hox* genes we examined. In addition, these results reveal that each *Hox* gene in *Parhyale* appears to have a differential requirement for PcG function in different tissues and regions of the body.

2.1.7 – PcG knockout-induced misexpression begins after wild-type *Ubx* boundary establishment

To examine the role of PcG genes in the establishment of *Hox* expression patterns, we performed HCR on a developmental time course of *Sce* and *Pcl* mutant embryos at developmental stages S13 (early germband), S19 (posterior flexure/ late germband), S21 (early limb morphogenesis) and S23 (late limb morphogenesis) and compared the expression of *Ubx* and *Abd-B* to that of wild-type embryos.

We observed that the anterior boundary of *Ubx* expression appeared unchanged in CRISPR-treated versus WT embryos (Fig. 6A, Supp. Fig. 6.1A, B). In WT embryos, *Ubx* expression begins at the very earliest stages in parasegment 5, and then expands anteriorly into parasegment 4 by around S13 (Liubicich et al., 2009). At S13, for both *Sce* and *Pcl* knockout embryos, we observed *Ubx* expression with an anterior boundary of parasegment 4 (5/5 *Sce* embryos, 9/9 *Pcl* embryos; Fig. 6A, Supp Fig. 6.1B). In the case of *Sce* we observed several embryos where one side of the embryo appeared developmentally delayed relative to the other, suggesting that one half of the embryo had been affected by CRISPR-Cas9 mutagenesis, while the other half had not (Fig. 6A). In these embryos, we also observed that the anterior boundary of *Ubx* expression was unchanged.

When we examined the expression of *Ubx* at S19 in knockout embryos, we began to see misexpression anterior to the WT expression boundary. Among the embryos we dissected at the S19 developmental stage, some embryos appeared to have developmental delay such that they had not yet formed the ventral furrow characteristic of S19 embryos. We labeled these embryos "S17-like" based on their morphology, and embryos with a more typical morphology "S19-like". Among S19-like embryos, we observed spots of *Ubx* misexpression anterior to the WT boundary (*Sce*: 1/3 embryos; *Pcl*: 3/3 embryos; Fig. 6C, D, Supp. Fig. 6D-D''). The misexpression seemed to be localized primarily to the neurogenic ectoderm. These results suggest that PcG genes are dispensable for wildtype *Ubx* anterior boundary establishment, but are essential for anterior boundary maintenance.

We observed a difference in the timing of misexpression relative to anterior boundary establishment when we examined *Abd-B* expression in knockout embryos at S19, in both S17-like and S19-like embryos. In S17-like embryos, we observed that *Abd-B* misexpression appeared to begin nearly concurrently with the early stages of *Abd-B* expression at the wildtype boundary (*Sce*: 3/5 embryos; *Pcl* 3/3 embryos; Fig. 6B, B'; Supp. Fig. 6.1C, C'). Moreover, S17-like embryos displayed *Abd-B* misexpression, but not *Ubx* misexpression. All S19-like embryos expressing *Ubx* also misexpressed *Abd-B* (Fig. 6E, Supp. Fig. 6.1E). These results suggest that the timing of *Abd-B* dependence on PcG function may differ from that of *Ubx*; namely, that *Abd-B* dependence on PcG function precedes *Ubx* dependence, and that *Abd-B* PcG-dependent boundary maintenance occurs simultaneously with wildtype boundary establishment.

2.1.8 – Tissue- and PcG knockout gene-specific effects on *Hox* expression

As development progresses to S21, we observed misexpression of *Ubx, abd-A,* and *Abd-B* in mutant embryos (Supp. Fig. 6.2). For *Sce* mutant embryos, we observed broad misexpression of *Ubx* in both limbs and the nervous system, weak misexpression of *Abd-B* in the nervous system, and very weak misexpression of *abd-A* in a few cells of the nervous system. The expression of these genes appeared to differ relative to that of *Pcl* mutant embryos, which displayed strong and broad misexpression of *Ubx* in both limbs and the nervous system, strong misexpression of *Abd-B* in the nervous system, and weak misexpression of *abd-A* in the nervous system. Thus, by the S21 stage, we began to observe differences in the effects of individual PcG gene knockouts on the expression of individual *Hox* genes.

At S23, the differences between *Sce* and *Pcl* mutant embryos became more dramatic (Fig. 4). While both knockouts displayed strong and broad *Ubx* misexpression, as well as strong *Abd-B* expression in the nervous system, we observed striking differences in the degree of *abd-A* misexpression between mutants. In *Sce* mutants, a small number of cells, restricted to the neurogenic ectoderm, appeared to show strong *abd-A* misexpression. In contrast, in *Pcl* mutants, we observed strong misexpression of *abd-A* all across the neurogenic ectoderm, as well as patchy misexpression in the limbs. The difference in these phenotypes suggests that *Pcl* is indispensable for proper *abd-A* repression, while *Sce* more weakly represses *abd-A*, indicating that individual PcG genes may have different roles in regulating individual *Hox* genes.

To assess the tissue localization of the apparent *Hox* misexpression, we performed *in situ* HCR on *abd-A* and *Abd-B* along with the pan-neural marker *elav* in *Sce* and *Pcl* knockout embryos (Fig. 7A, B; Supp. Fig. 7.1). We observed restriction of the *Abd-B* misexpression region to the neurogenic ectoderm, as we had imputed based on morphological evidence. Moreover, we confirmed that *abd-A* misexpression in *Sce* mutants is primarily restricted to small patches of cells in the nervous system, while *abd-A* misexpression in *Pcl* mutants expands into both neural and limb tissue.

Based on these data, we further quantified the degree of *Hox* misexpression for *Ubx*, *abd-A*, and *Abd-B* across our HCR-stained wildtype and knockout embryos. For each segment of the body, we evaluated the expression of each of the three posterior *Hox* genes in both the medial portion of the embryo containing the neurogenic ectoderm, and the lateral regions containing limb tissue. We assigned a value of 1 to a wildtype-like high level of *Hox* expression, a value of 0.5 to expression below the maximum wildtype expression or for patchy or partial expression, and a value of 0 to absence of expression. The results of this quantification are summarized in Fig. 7C for S21 embryos and Fig. 7D for S23 embryos. Notably, we observed that *abd-A* expression appeared to be lower in the nervous system in *Sce* knockout embryos at both stages S21 and S23 when compared to *Pcl* knockout embryos. We also observed that *abd-A* appeared largely absent from the limbs in *Sce* knockout embryos, but showed strong expression in *Pcl* knockout embryos, further confirming our observation that *Sce* and *Pcl* appear to have different roles in regulating *abd-A*.

These results together reveal that individual *Hox* genes in *Parhyale* appear to be regulated by distinct genetic mechanisms. For *Ubx*, both *Sce* and *Pcl* mutants showed similarly strong misexpression in the nervous system and limbs. *Abd-B* misexpression appeared to be concentrated in the nervous system in *Sce* knockout embryos, with slight misexpression in the limbs – for *Pcl* knockout embryos, we also observed strong misexpression in the nervous system and slightly higher *Abd-B* misexpression

in the limbs. Among the three posterior *Hox* genes, *abd-A* showed the strongest difference in effect between *Pcl* and *Sce* knockout embryos, with weak, neurally-restricted misexpression induced by *Sce* knockout and broad, strong misexpression induced by *Pcl* knockout. These results suggest that each of the three posterior *Hox* genes have differential dependence on the function of individual PcG genes, and that the role of PcG regulation in the limbs and nervous system also differs between *Hox* genes.

2.2 – Discussion

In this paper, we reveal the essential roles of PcG genes in maintaining proper *Hox* expression in the amphipod crustacean *Parhyale hawaiensis*. Loss of PcG function induces *Hox* misexpression and homeotic transformations in a manner similar to that observed in other systems. However, the effects of PcG loss of function appeared to manifest differently for individual *Hox* genes, in different regions of the embryo, and for individual PcG genes. The striking differences in the effects of PcG knockout suggest idiosyncratic regulation of *Hox* genes by PcG group genes. Moreover, the phenotypes we observed across tissues and timepoints also provided additional insights into the mechanisms of *Hox* regulation in this organism.

2.2.1 – Proper *Hox* regulation in *Parhyale* requires at least three distinct mechanisms

The phenotypes we observed, when considered in relation to previous experiments, suggest that there are at least three distinct regulatory mechanisms involved in proper *Hox* regulation in *Parhyale*: boundary establishment, boundary maintenance, and *Hox* cross-regulation. These results are consistent with previous work on *Hox* regulation in *Drosophila* (Akam, 1987; Mallo and Alonso, 2013).

1. A PcG-independent mechanism governs early *Hox* anterior boundary establishment

First, our data suggest that an early boundary establishment mechanism precedes the PcG-dependent phase of *Hox* regulation in *Parhyale*. In PcG knockout embryos at the early germband stage (S13), we consistently observed proper establishment of the anterior expression boundary of *Ubx*. This suggests that PcG genes are dispensable for the proper establishment of the anterior expression boundary, and that the early regionalization of the *Parhyale* embryo is governed by other mechanisms.

In insects, gap and pair rule transcription factors are the upstream regulators of *Hox* genes, and are responsible for anterior boundary establishment (Maeda and Karch, 2006, 2015). While the precise determinants of anterior boundary establishment in *Parhyale* remain unknown, a similar phase of regulation prior to PcG dependence must also proceed during early development. We hypothesize that a family of transcription factors may act in a similar function to the gap and pair-rule genes in insects to establish the early anterior expression boundaries of *Hox* genes in *Parhyale*. Future experiments using techniques such as single-cell RNA sequencing may provide opportunities to identify such factors.

2. A PcG-dependent mechanism governs *Hox* anterior expression boundary maintenance

Second, our data clearly demonstrate that, following the boundary establishment phase of *Hox* expression, PcG genes become essential for proper anterior boundary maintenance. For the five *Hox* genes we examined (*Scr*, *Antp*, *Ubx*, *abd-A*, and *Abd-B*), we observed evidence of *Hox* misexpression

anterior to the wildtype anterior expression boundary. This result demonstrates a second phase of *Hox* regulation, wherein PcG function is required to maintain proper *Hox* repression. The degree of PcG dependence appeared to vary at the level of tissue location and PcG gene for each of the genes we examined.

The observation of two phases of early *Hox* regulation in *Parhyale* is notable when compared to previous experiments in insects. In the long-germ insect *Drosophila*, the boundary establishment and boundary maintenance phases of *Hox* regulation appear to occur in an ordered sequence, wherein gap and pair rule genes serve as upstream inputs to *Hox* regulation, and PcG genes later become essential for *Hox* repression (Jaeger, 2011; Simon et al., 1992; Struhl and Akam, 1985). In contrast, in the short-germ insect *Gryllus bimaculatus*, RNAi knockdown of PRC2 complex components *E(z)* and *Esc* appeared to suggest that the *Hox* genes *abd-A* and *Abd-B* lacked the gap gene-based establishment phase of *Hox* expression (Matsuoka et al., 2015). The authors made this assertion based on data indicating that anterior *Hox* genes (*Scr*, *Antp*, *Ubx*) in PRC2-RNAi embryos displayed wild-type anterior expression boundaries early in development, followed by expansion of expression later in development. In contrast, for the posterior *Hox* genes *abd-A* and *Abd-B*, PRC2-RNAi embryos exhibited much earlier misexpression, which appeared at the same time that the wildtype expression boundary would be observed in WT embryos.

We observed a similar contrast in the potential timing of boundary establishment and boundary maintenance between *Ubx* and *Abd-B* in *Parhyale*. Namely, the *Ubx* anterior expression boundary in PcG knockouts appeared normal at early stages (~S13), while misexpression began later in development (~S19). In contrast, for *Abd-B*, misexpression appeared to begin concurrently with the wildtype expression boundary (~S17). These data in *Parhyale* could indicate a similar mechanism to that observed in *Gryllus*: that anterior and posterior *Hox* genes have different boundary maintenance and boundary establishment requirements. However, in *Parhyale*, while *Abd-B* misexpression appeared to occur simultaneously with normal boundary establishment, the tissue location and intensity of expression differed. *Abd-B* misexpression appeared primarily in the neurogenic ectoderm at that early stage, and the broad normal anterior boundary was also clearly visible. Thus, it is possible that, despite the simultaneous appearance of the normal anterior boundary and ectopic expression, both mechanisms could still be required for *Parhyale Abd-B*.

Our results suggest that the previous data in *Gryllus* could benefit from reinvestigation. For *Gryllus abd-A* and *Abd-B*, the colorimetric *in situ* hybridization results presented show strong misexpression at developmental stages concurrent with strong wildtype expression; however, it is possible that at slightly earlier stages, one might still observe a stronger and clearer normal expression boundary, as we observed in *Parhyale*. Using a more sensitive technique such as *in situ* HCR at earlier stages could reveal the potential for a simultaneous biphasic regulatory mechanism, as our results in *Parhyale* suggest.

3. PcG knockouts reveal conditional cross-regulation of *Hox* genes

Third, our results provide further insight into the complexities of *Hox* cross-regulatory mechanisms in *Parhyale*. Previous studies have revealed that *Hox* genes in *Parhyale* exhibit cross-regulation. For example, CRISPR-Cas9 knockout of *Abd-B* results in posterior expansion of *Ubx*

expression, and homeotic transformations of abdominal segments to walking and jumping legs (Martin 2016, Jarvis 2018). This indicates that *Abd-B* represses *Ubx*. In a different set of studies, heat shock-induced misexpression of *Ubx* resulted in decreased *Scr* expression, indicating that *Ubx* represses *Scr* (Pavlopoulos et al., 2009). In our study, we observed contrasting effects of PcG knockout on the *AbdB* repression of *Ubx* and the *Ubx* repression of *Scr*.

In PcG knockout embryos, *Abd-B* appeared to properly repress *Ubx* in the abdomen of the embryo, while in regions where *Abd-B* was ectopically expressed, *Ubx* did not appear to be ectopically repressed. In wildtype embryos, *Ubx* expression is strongly observed in cells that begin expressing *Abd-B* at S17, and gradually grows weaker after S19, such that *Ubx* expression in the abdominal limbs and nervous system is weak but still observable by S21. By S23, *Ubx* expression in wildtype embryos is not observable in abdominal limbs, but weakly persists in some cells of the nervous system.

In PcG knockouts, the same trend is observed: strong *Ubx* expression in *Abd-B*-expressing cells of the abdomen, which gradually tapers into weak expression restricted to the nervous system by S23. In the same PcG knockouts, *Abd-B* misexpression begins with a few cells of the neurogenic ectoderm at early stages (S17), expanding to encompass much of the nervous system of all anterior segments by S23. Despite this strong expression of *Abd-B*, *Ubx* expression does not appear to be repressed in the nervous system. Cells expressing both *Ubx* and *Abd-B* can be found even at S23 in the nervous system of PcG knockout embryos. Thus, despite the fact that both anterior neural cells and posterior abdominal tissue contain cells that simultaneously express *Ubx* and *Abd-B* beginning at S17, *Abd-B* only represses *Ubx* in the abdomen, and not in the anterior neural tissue where it is ectopically expressed.

One explanation for this observation may be that *Abd-B* repression of *Ubx* depends on some third factor expressed in abdominal tissue, and that *Abd-B* repression by *Ubx* in the abdomen is not dependent on PcG function. Alternatively, anteriorly expressed *Ubx* could undergo auto-regulation, thus escaping later ectopic *Abd-B* expression. Heat shock misexpression of *Abd-B* could more precisely reveal whether *Abd-B* repression of *Ubx* in anterior tissue is dependent on PcG function.

PcG knockout embryos also revealed further complexity in the repression of *Scr* by *Ubx*. In PcG mutants, *Scr* expression in T1 is substantially reduced, likely due to strong misexpression of *Ubx*. Moreover, *Scr* misexpression in the antennae is observed in only a few cells, potentially also due to strong *Ubx* misexpression in those tissues. These results indicate that, in the antennae and T1 appendages, *Ubx* repression of *Scr* is not dependent on PcG function.

A more complicated mechanism appears to take place in the Mn, Mx1, and Mx2 appendages, where mutually exclusive *Scr* and *Ubx* expression domains are observed. In anterior regions of these limbs that ectopically express *Ubx*, *Scr* expression is almost completely eliminated. However, the posterior regions of these limbs appear to retain *Scr* expression. In these tissues, it is possible that either *Ubx* requires PcG function to repress *Scr*, or that some other mechanism prevents ectopic *Ubx* expression, such as the presence of a *Ubx* repressor or absence of *Ubx* activators. Thus, these results indicate that *Hox* cross-regulation may depend on factors found only in particular regions of an individual body segment.

Together, these data indicate that *Hox* cross-regulation occurs in *Parhyale* via layered mechanisms, potentially dependent on tissue type, expression timing, and body region. Future

experiments using knockouts of *Hox* genes and PcG genes, or heat shock induced misexpression combined with PcG knockout, could clarify the role of PcG function in *Hox* cross-regulation.

2.2.2 – Each posterior *Hox* gene in *Parhyale* exhibits idiosyncratic regulation

In addition to revealing insights into broader *Hox* regulatory mechanisms, our results also suggest that individual *Hox* genes in *Parhyale* follow idiosyncratic regulation. The three posterior *Hox* genes showed differences in the timing of their dependence on PcG function, the tissues in which ectopic expression was observed, and their dependence on specific PcG complex genes.

Differences in relative *Hox* misexpression timing

We were able to characterize the timing of misexpression for *Ubx* and *Abd-B*, and observed differences in the relative timing of *Hox* misexpression in relation to the normal anterior boundary establishment. *Ubx* showed no misexpression at the time of *Hox* initiation (S13), and did not exhibit observable misexpression in segments more anterior than the normal anterior expression boundary until around S19. In contrast, *Abd-B* showed simultaneous misexpression and normal anterior boundary establishment at S17. These results suggest that *Ubx* and *Abd-B* become dependent on PcG function at different times during development.

Differences in tissues competent for ectopic expression

While *Ubx* became strongly misexpressed upon knockout of PcG genes, *Abd-B* misexpression appeared to exhibit tissue dependence. For both *Pcl* and *Sce* knockout, *Abd-B* became strongly misexpressed in the nervous system, but did not appear to exhibit strong misexpression in the limbs. This result suggests that PcG may not be the primary method of repression for *Abd-B* in the limbs, and that additional mechanisms may be needed to explain how *Abd-B* is expressed within the limbs.

The presence of differential expression of *Hox* genes along the proximal-distal axis (in this case, the medial neurogenic ectoderm and the more distal limbs) is not without precedent in *Parhyale*. Indeed, more anterior *Hox* genes such as *Scr* appear to have differential regulation along the proximal-distal axis. *Scr* expression in the T1/Mxp appendage is restricted to the limb, and is absent from the nervous system. This *Scr* expression is critical to T1/Mxp identity, as revealed in knockout experiments. Our results indicate that a similar partitioning of limb and nervous system expression potential also exists for *Abd-B*. The precise mechanisms of this proximal-distal regulation remain to be investigated.

Together, these data suggest that proximal-distal regulation of *Hox* expression may be a general feature of the crustacean *Hox* regulatory architecture. Moreover, the existence of such mechanisms provide another regulatory parameter for the fine tuning of *Hox* expression, which may have played a role in the evolution of different expression patterns across crustaceans.

Differences in PcG complex gene dependence

Among the three genes we examined, *abd-A* appeared to show a striking difference in its dependence on different PcG genes for proper regulation. The phenotypes for PcG knockout-induced *Ubx* and *Abd-B* misexpression appeared similar in *Sce* and *Pcl* mutants; however, they differed

substantially in the case of *abd-A*. *Sce* knockout induced misexpression in a few cells of the nervous system, whereas *Pcl* knockout induced strong and broad *abd-A* misexpression in the nervous system and strong but patchy misexpression in the limbs.

These results indicate that individual PcG genes may have gene-specific regulatory functions for individual *Hox* genes. In the case of *Sce* and *Pcl*, members of PRC1 and PRC2.1 respectively, complex-specific functions may explain the differences in phenotypes. For example, PRC1 is responsible for H2AK188 ubiquitylation, whereas PRC2 is responsible for H3K4 trimethylation. Thus, it is possible that H2AK188ub is largely dispensable for *abd-A* regulation, whereas H3K4me3 is essential. Immunofluorescence experiments on PcG knockout embryos could reveal such differences, while ChIP-Seq or CUT&RUN experiments could reveal the precise cis-regulatory targets of these genes.

The presence of differential regulation of *Hox* genes by PcG genes reveals that the upstream factors governing *Hox* regulation in *Parhyale* may have become distinct. Alterations to the function of *Sce*, for example, could preferentially result in changes to *Ubx* and *Abd-B* expression, without altering *abd-A* expression. While we only examined two PcG genes in detail in our study, it is possible that other PcG genes may also have regulatory functions that preferentially affect or exclude individual *Hox* genes. Such regulatory mechanisms would provide a direct method to alter the expression of a single *Hox* gene while leaving the expression of other *Hox* genes intact, providing a potential mechanism for the evolution of new *Hox* expression patterns.

2.2.3 – PcG genes are crucial for diverse developmental processes in *Parhyale*

In addition to homeotic phenotypes and effects on *Hox* regulation, we characterized a number of additional roles for PcG genes during *Parhyale* development. In particular, we observed defects in the development of body and limb plates, as well as defects in proximal-distal patterning, suggesting that PcG genes may also play important roles in these developmental processes.

PcG knockout embryos displayed reduced plates reminiscent of those observed in knockouts of genes crucial for the development of insect wings, such as *vestigial*, *nubbin*, and *apterous*, which resulted in coxal, tergal, and basal plates of reduced size. This suggests that PcG genes may regulate such plate developmental genes. For one gene, *pho*, we observed potential homeotic transformations of plates, suggesting that PcG genes may also play a role in plate identity. We also observed phenotypes indicating that PcG genes play a role in the closing of the dorsal side of the embryo, as we observed tergal plates that failed to fuse dorsally, fused asymmetrically between left and right halves of the embryo, or fused adjacent segments along the anterior-posterior axis. The genes involved in dorsal closure in *Parhyale* remain uninvestigated, but we expect such genes are likely also regulated by PcG function.

Moreover, the consistent loss of the T2 coxal and tergal plates across PcG genes reveals subdivisions of regions of *Hox*-specified identity in the *Parhyale* embryo. Both T2 and T3 limbs in *Parhyale* develop into a claw, or cheliped. However, knockout of PcG frequently ablates the entire T2 tergal and coxal plates, in addition to inducing T2 -> T4/5 transformations. In the T3 segment, which also consistently had homeotic transformations of T3 -> T4/5 identity, we sometimes observed failure of fusion of the tergal plate, but we did not observe total loss of the tergal and coxal plates. Thus, PcG

knockout reveals that distinct genetic mechanisms may separate body segments with similar homeotic identity, revealing additional layers of distinction that subdivide body segments with similar *Hox* expression.

Finally, we observed defects in proximal-distal development for most of the PcG knockouts we examined, including truncation of limbs and fusion of limb segments. For some of these phenotypes, it was possible to infer the specific limb segments that were lost or fused. For example, *Sce* knockout appears to induce variable fusion of the propodus, carpus, and merus. Such fusions are not identical to any of the phenotypes previously described for the leg gap gene or *Hox* gene knockouts in *Parhyale*. These results indicate that PcG genes may work in concert with leg gap genes, and other unknown genes, to specify proximal-distal identity along the legs in *Parhyale*.

2.2.4 – TrxG proteins show no obvious role in *Parhyale Hox* regulation

We were unable to recover any homeotic phenotypes for the TrxG genes we targeted using CRISPR-Cas9 mutagenesis. Some of the genes we analyzed displayed their roles in *Hox* regulation in suppressor screens of PcG genes in *Drosophila* – performing combined knockouts of these genes and PcG genes in *Parhyale* could clarify their roles in *Hox* regulation.

The absence of any obvious phenotypes for *Trx* is somewhat surprising, given the phenotypes observed in *Drosophila* knockouts, which show homeotic transformations of haltere to wing structures (Breen and Harte, 1991, 1993). It is possible that the function of *Trx* could be replaced in *Parhyale* by *Trithorax-related* (*Trr*), which we also found in the *Parhyale* genome, which would be an interesting question for a future study. More careful analysis of the *Trx* locus, including long-read RNA-sequencing, RACE, or other experiments that examine the exact sequence content of *Trx* transcripts, could also clarify whether *Trx* in *Parhyale* is indeed expressed as two separate transcripts, and lead to clearer understanding of the function and structure of this gene.

2.3 – Ideas and Speculation

When considered alongside previous data in this organism, our results provide a number of interesting testable hypotheses and avenues for future investigation. In particular, we propose that currently undiscovered mechanisms, both ancestral to arthropods and evolved in crustaceans, may play a role in crustacean *Hox* regulation. We also examine the implications of our data in relationship to previous studies characterizing the functions of PcG genes in other organisms.

2.3.1 – PcG knockout phenotypes suggest a limb-specific repression mechanism for *Abd-B*

While we observed strong and broad misexpression of *Ubx* in both limbs and the nervous system in response to PcG knockout at late stages (S21-S23), *Abd-B* misexpression was mostly restricted to the nervous system for both *Sce* and *Pcl* knockouts. This result contrasts with PcG knockout phenotypes observed in *Drosophila*, where broad *Abd-B* misexpression and homeotic transformations of body segments towards the *Abd-B*-specified A8 identity have been observed. The absence of strong *Abd-B* misexpression in the limbs of PcG knockout *Parhyale* suggests the presence of

a secondary limb-specific repression mechanism. Such a mechanism could be either the absence of an essential *Abd-B* activator, or the presence of an additional *Abd-B* repressor in limb tissue.

This observation reveals potential insights into crustacean *Hox* regulatory evolution when compared to the results found in *Gryllus bimaculatus* PRC2 knockdown experiments (Matsuoka et al., 2015). In those studies, *Abd-B* misexpression also appeared to be restricted to the neurogenic ectoderm, although the authors do not make any special note of this phenotype. Given that *Abd-B* is not normally expressed in any limb tissue in insects, the absence of *Abd-B* expression in limbs may not be surprising for this organism. However, the phenotypes observed in *Parhyale* provide additional context, namely suggesting that an ancestral secondary mechanism for *Abd-B* repression in thoracic limbs may be shared between crustaceans and short-germ insects. This secondary repression mechanism may then have been lost in long-germ insects, such as *Drosophila*, enabling broader derepression of *Abd-B* in response to loss of PcG function (Supp. Fig. 8.1A).

2.3.2 – PcG knockouts reveal gene-specific roles in *abd-A* regulation and an unknown posterior boundary maintenance mechanism

In addition to the differences in effect for individual PcG gene knockouts on *abd-A*, it is also notable that the posterior boundary of *abd-A* appears to remain intact in PcG mutants. In *Drosophila*, while the anterior boundaries of *Hox* expression are maintained by PcG function, the posterior boundaries of *Hox* expression are thought to be maintained by *Hox* cross-regulation. In *Parhyale*, this posterior boundary maintenance appears to be true for *Abd-B* repression of *Ubx* and *Ubx* repression of *Scr*. However, previous knockout experiments have shown that neither *Ubx* nor *Abd-B* is necessary for the proper posterior boundary maintenance of *abd-A*, suggesting no role of *Hox* cross-regulation in this process (Jarvis et al., 2022). Our data indicate that PcG knockout also does not disrupt the posterior boundary of *abd-A* expression. Thus, an additional, unknown mechanism governs the position of the posterior boundary of *abd-A* expression which is independent of known *Hox* maintenance mechanisms in arthropods.

The existence of such a mechanism could provide an explanation for the divergent expression patterns of *abd-A* in different groups of crustaceans. For example, in the decapod crustacean *Procambarus fallax* and the mysid *Mysidium columbiae*, *abd-A* is strongly expressed in segments A4 and A5, whereas its expression is weaker in *Parhyale*. The evolution of a novel, gradated *abd-A* repression mechanism could explain the shift in *abd-A* expression in *Parhyale* and other amphipods, providing a mechanistic explanation for crustacean body plan evolution. We propose that a Posterior Boundary Factor may govern the regulation of the posterior boundary of *abd-A* in crustaceans, and that this factor operates independently of *Hox* cross-regulatory mechanisms (Supp. Fig. 8.1B). This factor may be one of the many additional PcG genes for which phenotypes have not been thoroughly investigated, or it may be another mechanism entirely. Such a Posterior Boundary Factor, if identified, could represent a novel regulatory mechanisms evolved within crustaceans.

2.3.3 – Implications for the cis-regulatory architecture of the *Parhyale Hox* cluster

Our experiments demonstrate a clear role for PcG genes in *Parhyale Hox* regulation. Given the current understanding of PcG function in *Drosophila* and its relationship to cis-regulatory elements,

we can make certain hypotheses about the cis-regulatory architecture of the *Parhyale Hox* cluster. In particular, our data imply the existence of Polycomb Response Elements (PREs) in the *Parhyale Hox* cluster. PREs are specialized cis-regulatory elements that recruit *pho* binding and later PcG complex activity to a locus in *Drosophila*.

Future studies using ChIP-Seq or CUT&RUN, combined with newly available functional genomic data for *Parhyale*, such as ATAC-Seq (Sun et al., 2021), could reveal the precise location of PREs and other important cis-regulatory elements in the *Parhyale Hox* cluster. Such approaches could then enable the precise dissection of cis- and trans-regulatory mechanisms of *Hox* regulation, and enable direct comparisons of the *Parhyale Hox* cluster to the more well-studied *Drosophila Hox* cluster.

2.3.4 – Implications for the ancestral roles of PcG in *Hox* regulation

Previous work in flies and mice has revealed two different potential roles for *PcG* regulation of *Hox* genes. In *Drosophila*, *Hox* boundary establishment occurs independently of PcG function, and PcG genes are required for proper boundary maintenance (Maeda and Karch, 2006, 2015). In vertebrates, boundary establishment and maintenance both appear to depend on PcG/TrxG-mediated chromatin states (Chambeyron and Bickmore, 2004; Deschamps and van Nes, 2005; Kmita and Duboule, 2003; Rastegar Mojgan et al., 2004; Soshnikova and Duboule, 2009; Tschopp and Duboule, 2011).

The previous study in the cricket *Gryllus bimaculatus* suggested that *Gryllus* may represent an intermediate between the mouse and fly literature, wherein anterior *Hox* genes depend on PcG for boundary maintenance, but not boundary establishment, whereas *abd-A* and *Abd-B* have a vertebrate-like dependence on PcG for both boundary establishment and maintenance (Matsuoka et al., 2015).

Our data suggest that more careful examination of the timing and intensity of misexpression could reveal that *Hox* genes require PcG only for boundary maintenance, rather than both maintenance and establishment, in short-germ arthropods such as *Gryllus* and *Parhyale*. Notably, short-germ arthropods have progressive segmentation of the body, where anterior regions segments develop prior to more posterior segments. Thus, within an embryo, anterior segments are "older" than posterior segments, and often further along in developmental processes such as limb bud formation and *Hox* expression. In PcG loss of function experiments in both *Gryllus* and *Parhyale*, ectopic *Hox* expression in anterior tissues occurs concurrently with appearance of the wildtype boundary. The previous study in *Gryllus* argued that this result suggested that *Gryllus abd-A* and *Abd-B* did not utilize a PcG-independent boundary establishment phase, and instead used a vertebrate-like PcG-dependent boundary establishment mechanism. We favor the interpretation that the PcG-dependent boundary maintenance phase begins in anterior tissue at the same time that more posterior tissue enters the PcG-independent boundary establishment phase in short germ insects such as *Gryllus* and *Parhyale*, and that the apparent concurrent timing is a result of the different "ages" of the anterior and posterior regions.

As our data do not directly identify upstream PcG-independent boundary establishment factors for *Parhyale Abd-B*, we provide two potential interpretations of our findings in the context of broader bilaterian *Hox* regulation by PcG (Supp. Fig. 8.1C). In the first hypothesis, *Parhyale Abd-B*

behaves analogously to the previous interpretation of *Gryllus abd-A* and *Abd-B* regulation, wherein the more posterior *Hox* genes utilize a PcG-independent mechanism of boundary establishment. The second hypothesis, which we favor based on our interpretation of the early *Abd-B* misexpression pattern in *Parhyale*, suggests that all arthropods may deploy two phases of boundary regulation: PcG-independent boundary establishment, and PcG-dependent boundary maintenance.

In the case of the second hypothesis, a striking asymmetry is observed between arthropod and vertebrate *Hox* regulation by PcG genes, wherein arthropod PcG function is primarily necessary for *Hox* expression maintenance, rather than initial boundary establishment, while vertebrate PcG function is important for both establishing and maintaining *Hox* patterns. These results refute the assertion that a transition in *Hox* regulation occurred during arthropod evolution from an ancestral PcG-based *Hox* boundary establishment mechanism towards a gap-gene mechanism during arthropod evolution. Instead, our results suggest that understanding the ancestral role of PcG in regulating the urbilaterian *Hox* cluster will require further sampling at more early-branching protostome and deuterostome lineages.

2.4 – Conclusion

Distinct regulatory strategies provide possible mechanisms for *Hox* regulatory evolution

Together, our PcG knockout and *Hox* expression data indicate that *Parhyale Hox* genes have meaningful differences in their upstream regulatory mechanisms. These differences can occur with respect to the timing of PcG dependence, tissue-specific regulatory mechanisms, and PcG gene-specific effects. The presence of distinct regulatory mechanisms for each *Hox* gene in *Parhyale* suggest that crustacean *Hox* genes more broadly could have evolved distinct *Hox* regulatory mechanisms, wherein perturbations to canonical regulation could have effects on individual *Hox* genes, at specific times in development, and in individual tissues, rather than affecting all *Hox* genes uniformly across the body.

In addition, the phenotypes we have observed, when considered in conjunction with previous work examining *Hox* cross-regulatory mechanisms, suggest that additional layers and mechanisms of regulation remain to be discovered within crustacean *Hox* clusters. Such unknown mechanisms may include: the affectors responsible for *Hox* anterior boundary establishment prior to PcG dependence; the activators and repressors governing *Ubx* expression potential in mouthparts; the secondary repression mechanism for *Abd-B* in the limbs; and the affectors maintaining the posterior boundary of *abd-A*. Identifying such factors and examining their function across crustaceans could lead to a deeper mechanistic understanding of the ancient processes underlying crustacean body plan evolution.

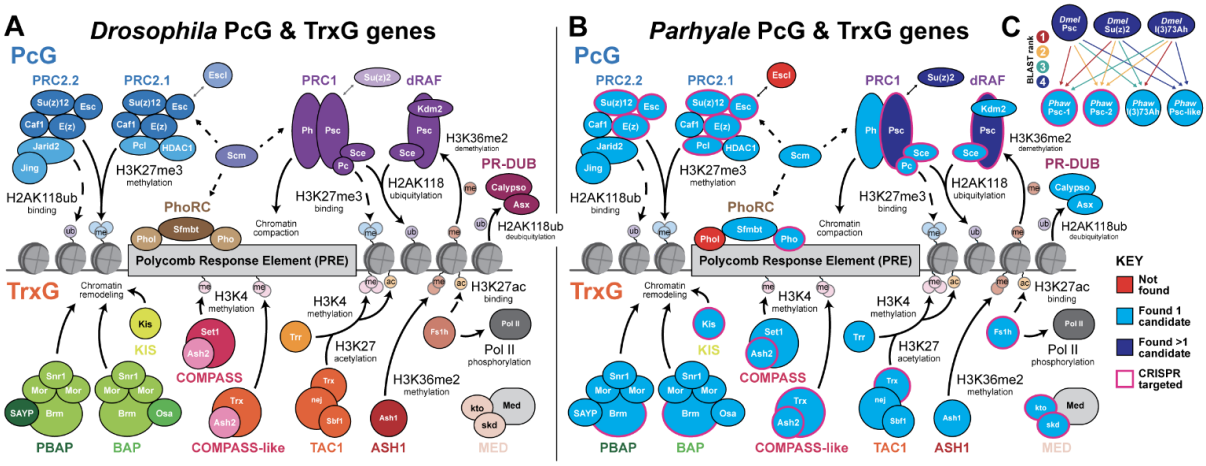


Figure 2.1: The Parhyale hawaiensis genome contains all core PcG and TrxG genes.

A) Schematic of the core PcG and TrxG complexes found in Drosophila, in addition to other genes classified as PcG/TrxG genes, alongside known gene regulatory functions. Adapted from (Erokhin et al., 2018; Geisler and Paro, 2015; Kassis et al., 2017; Schuettengruber et al., 2017).

B) Summary of PcG/TrxG genes found in Parhyale using reciprocal BLAST. All genes except esc-l and pho-l were found in the Parhyale genome.

C) Drosophila Psc and Su(z)2 in Parhyale map ambiguously to two transcripts, which we have named Psc-1 and Psc-2. A third potential Psc paralog was also identified. We named this gene Psc-like.

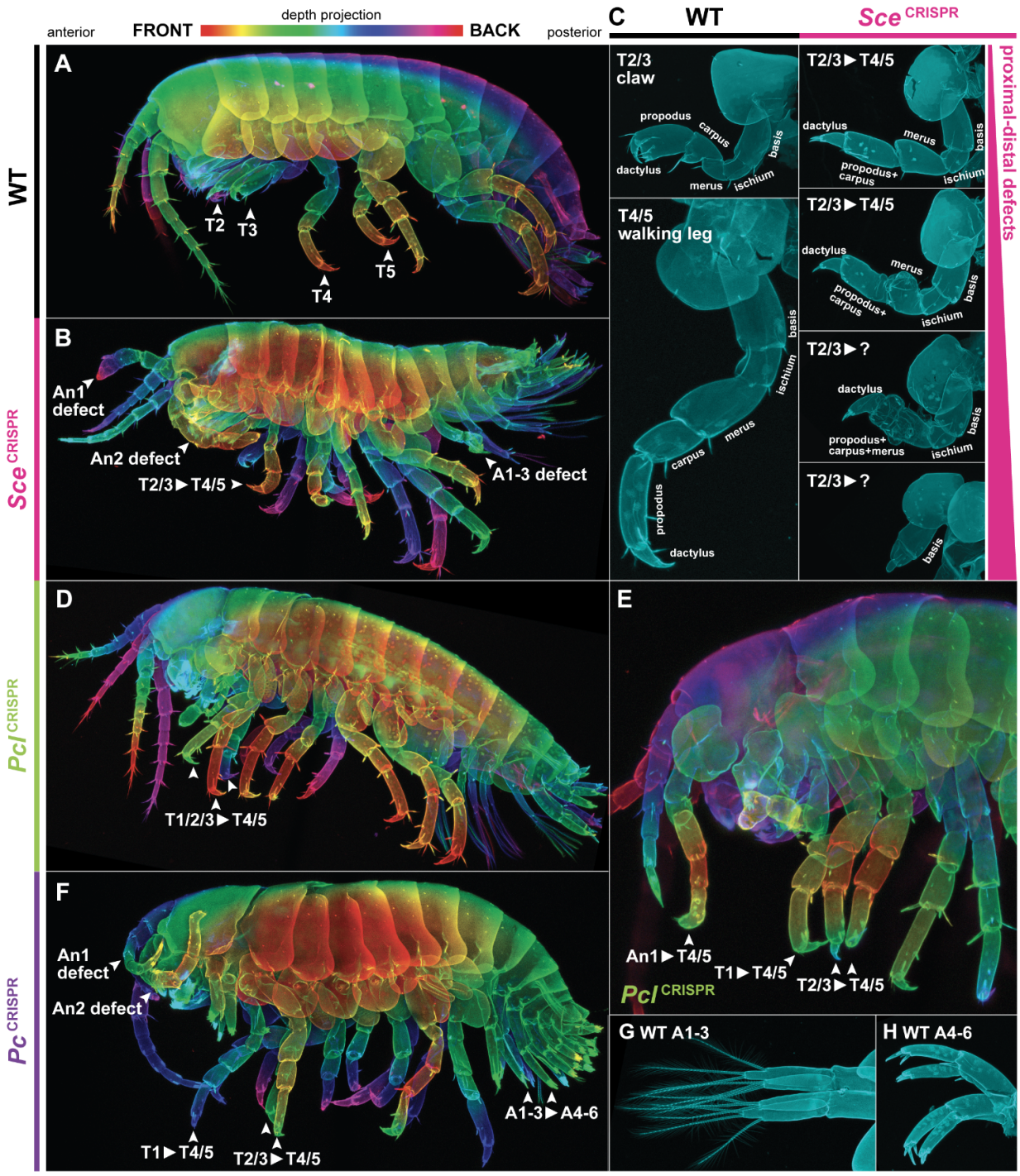


Fig. 2.2. PcG knockout hatchlings exhibit homeotic transformations

A) Representative wildtype *Parhyale* hatchling. Of note, the T2 and T3 appendages are claws (chelipeds) and the T4 and T5 appendages are forward walking legs (forward pereopods).

B) Representative *Sce* mutant hatchling. *Sce* mutants show various limb defects, in addition to homeotic transformation of anterior thoracic appendages (T1-3) toward T4/5 identity.

C) Proximal-distal phenotypes concurrent with homeotic phenotypes in *Sce* mutants. Left column shows dissected WT T2/3 and T4/5 appendages. Right column shows a range of different proximal-distal defects in *Sce* T2/3

appendages. Despite these proximal-distal defects, comparison of the merus of *Sce* mutant T2/3 appendages to the merus of WT T2/3 and WT T4/5 shows a clear homeosis.

D) Representative *Pcl* mutant hatchling, showing prominent T1/2/3 -> T4/5 homeotic transformations. *Pcl* mutants did not often appear to have proximal-distal defects.

E) Representative *Pcl* mutant hatchling showing T1/2/3 -> T4/5 homeotic transformations, as well as An2 -> T4/5 transformation. Notably, the formation of an ectopic coxal plate suggests T4/5 transformation and phenocopies previous *Ubx* heat shock overexpression experiments (Pavlopoulos et al., 2009)**.**

F) Representative *Pc* mutant hatchling showing T1/2/3 -> T4/5 homeosis, as well as A1-3 -> A4-6 homeosis and antennal defects.

G) Dissected WT A1-3 swimming leg (pleopod).

H) Dissected WT A4-6 anchoring leg (uropod).

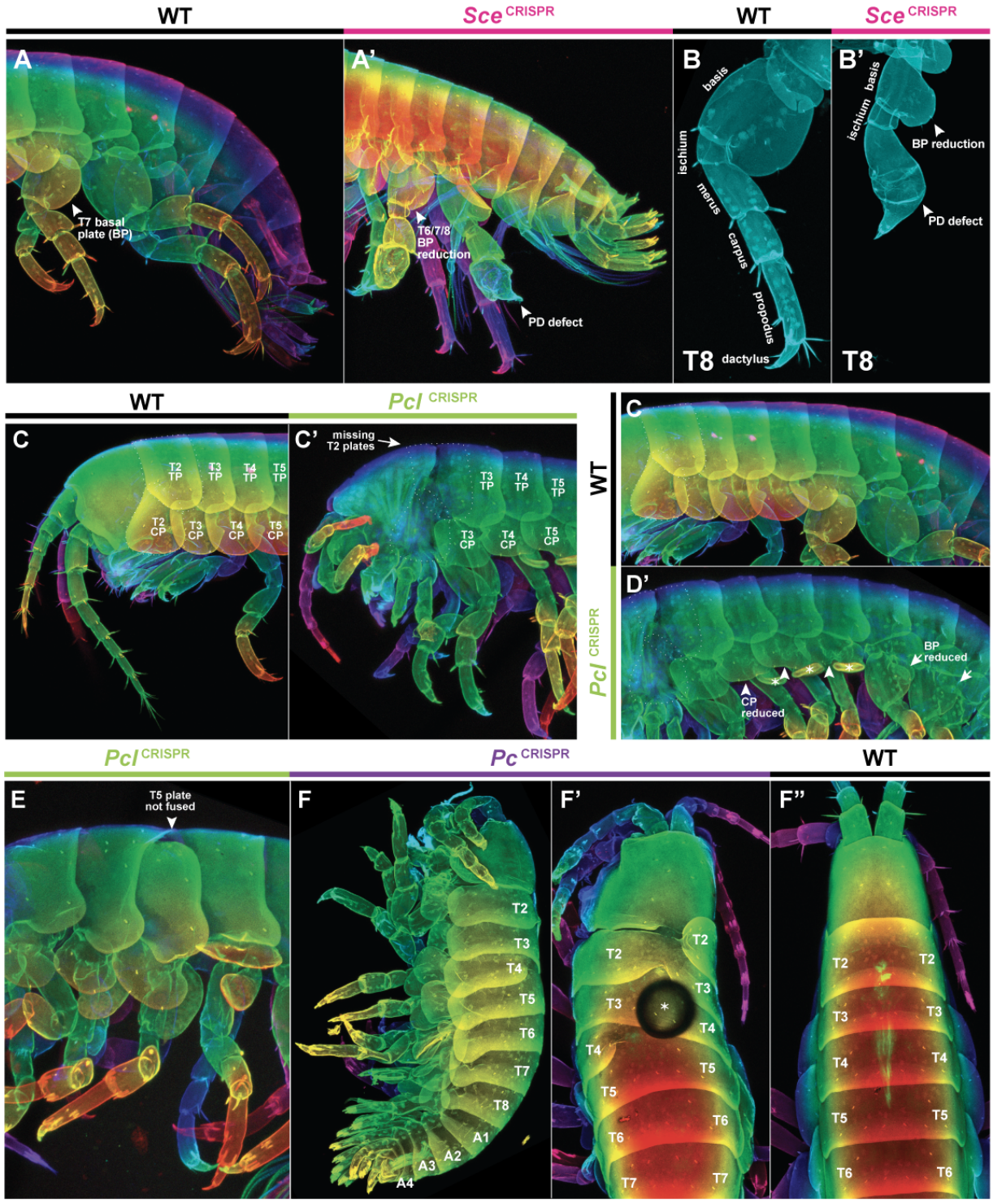


Fig. 2.3: Proximal-distal and plate defects of PcG mutants

A) Wildtype *Parhyale* posterior showing T6-8 jumping legs and abdomen.
A') *Sce* mutant hatchling showing proximal-distal defects in T6/8 appendages and deformed basal plates.
B) WT T8 appendage.
B') *Sce* mutant T8 appendage showing proximal-distal leg segment fusion and reduced basal plate.
C) WT anterior, with tergal and coxal plates labeled.

C') *Pcl* mutant showing loss, with dotted line indicating position of missing T2 tergal and coxal plates, as evidenced by the lack of the triangular T2 coxal plate.
D) View of tergal and coxal plates along WT thorax.
D') *Pcl* mutant showing reduced coxal and basal plates. Marked with asterisks are the gills, normally not visible from a lateral view in a WT hatchling, but revealed by coxal plate reduction.
E) *Pcl* mutant showing unfused tergal plate (labeled). This sample also shows reduced coxal plates and visible gills.
F-F') Lateral and dorsal views of the same *Pc* mutant hatchling, illustrating asymmetrical fusion of plates between left and right halves of the body. Asterisk marks a bubble in the mounting medium.
F'') Dorsal view of a WT hatchling.

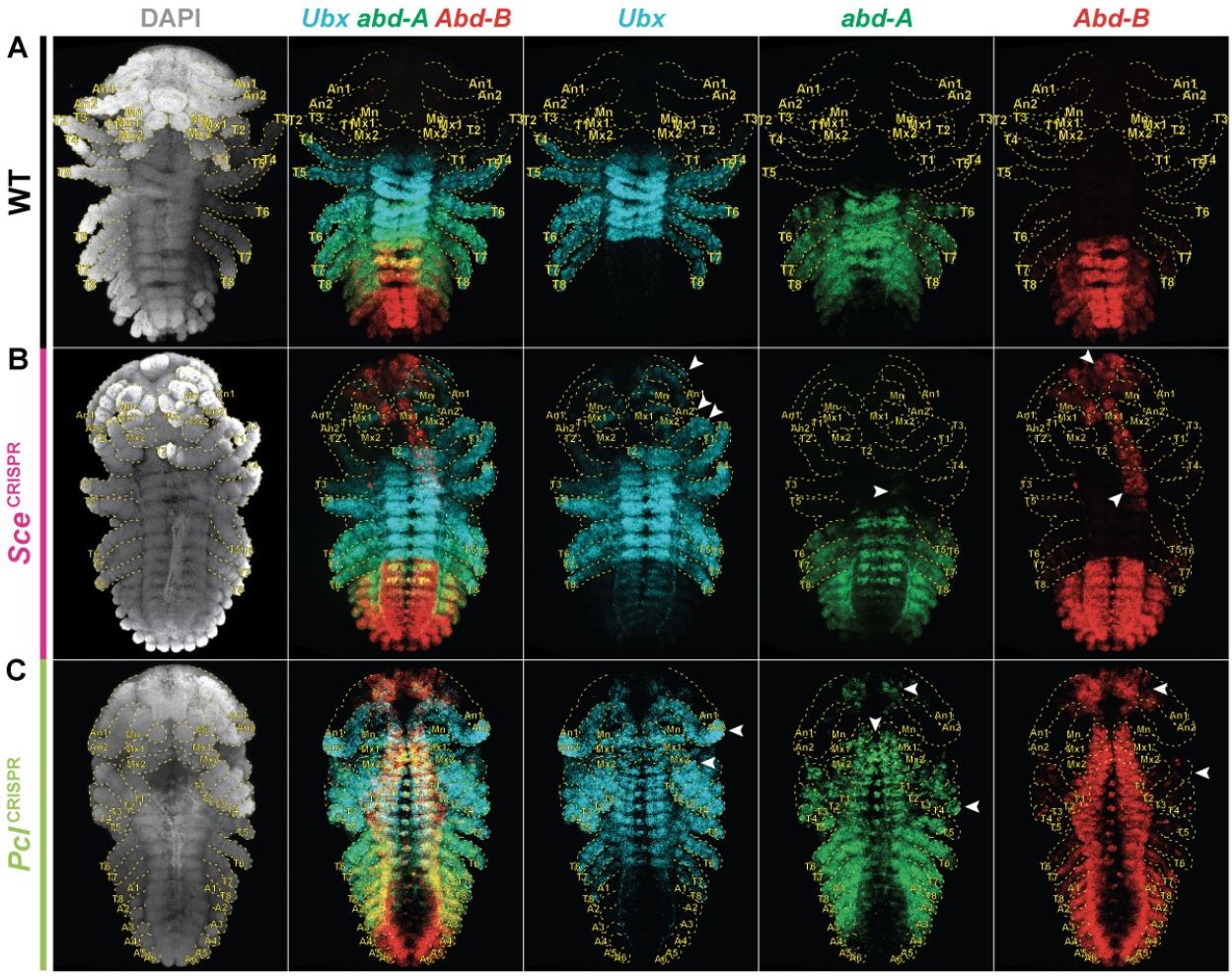


Fig. 2.4: CRISPR mutagenesis of PcG induces *Hox* misexpression

A) *In situ* HCR of *Ubx, abd-A,* and *Abd-B* in a representative WT embryo at stage 23. The anterior boundary of *Ubx* expression is in the T2 appendage, with weak *Ubx* expression in T2-3 and stronger *Ubx* expression in T4-8. The anterior boundary of *abd-A* expression is the T6 appendage. Strong *abd-A* expression is found in T6-A3, with weaker, gradated expression found in A4 and A5. The anterior boundary of *Abd-B* expression is A1, and *Abd-B* expression is found throughout the abdomen (A1-6).

B) *In situ* HCR of *Ubx, abd-A,* and *Abd-B* in a representative *Sce* mutant embryo at stage 23. The embryo appears to have mosaic loss of function, with the head and anterior-left half of the embryo showing *Ubx* misexpression beyond the normal WT anterior boundary (the embryo is mounted ventral side up; in the image, the left half of the embryo is on the right side). *abd-A* misexpression is slight, restricted to a few cells of the nervous system. *Abd-B* is misexpressed strongly in the nervous system and not in the limbs.

C) *In situ* HCR of *Ubx, abd-A,* and *Abd-B* in a representative *Pcl* mutant embryo at stage 23. The embryo appears to have bilateral loss of function. *Ubx* misexpression expands throughout the anterior of the embryo. *abdA* misexpression appears strongly throughout the nervous system and in strong patches in the limbs, in stark contrast to weak misexpression in *Sce* mutant embryos. *Abd-B* appears strongly misexpressed in the nervous system, including long filaments of *Abd-B* misexpression in the limbs, which resembles the pattern of limb enervation.

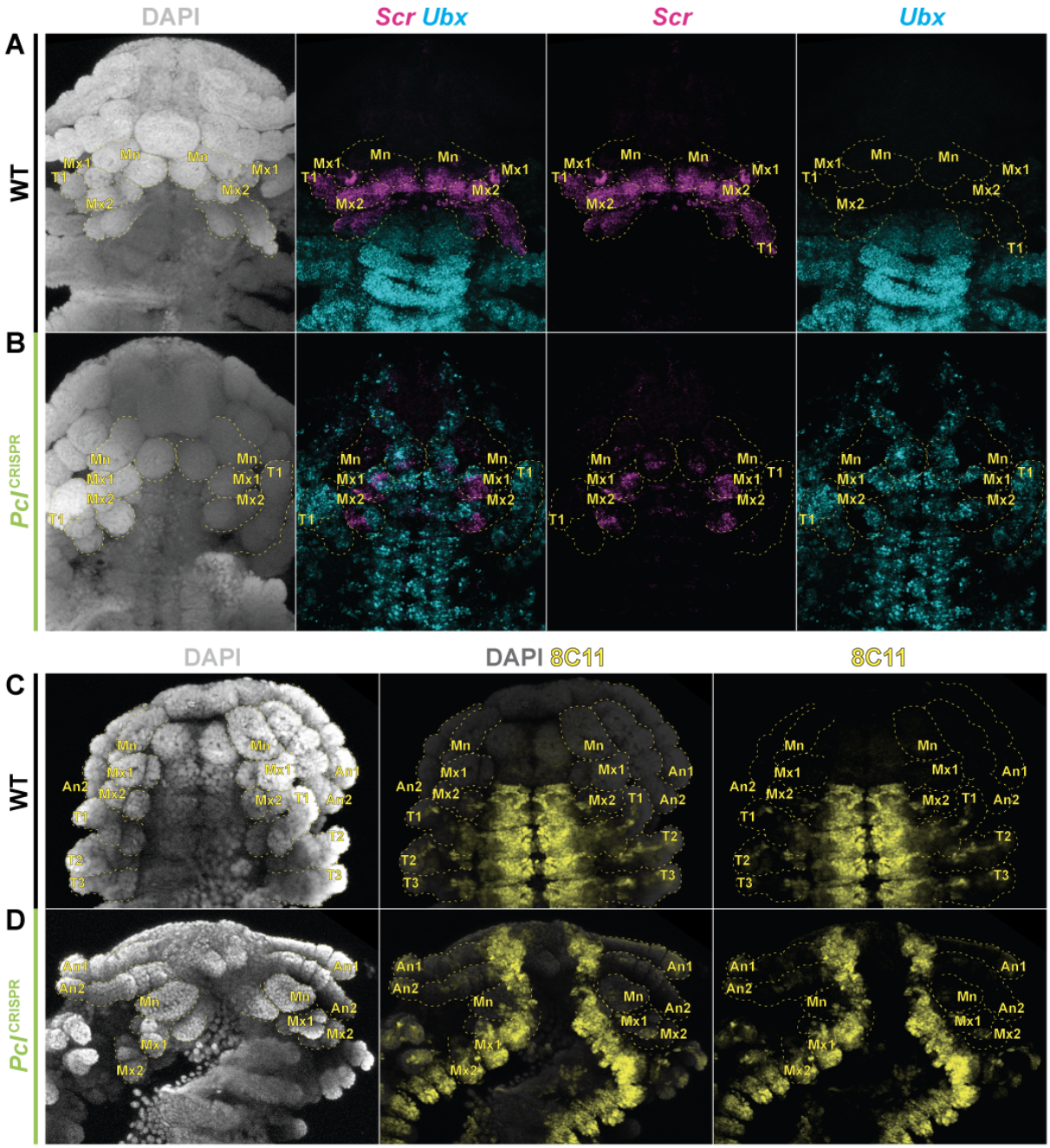


Fig. 2.5. Misexpression of *Scr* and *Antp* in PcG mutants

A) *In situ* HCR of *Scr* and *Ubx* in the head of a WT embryo at stage 23. *Scr* is expressed in Mx1, Mx2, and T1. *Ubx* is expressed in T2 and T3.

B) *In situ* HCR of *Scr* and *Ubx* in the head of a *Pcl* mutant embryo at stage 23. *Scr* expression expands anteriorly into Mn, as well as a few transcriptional spots (arrowheads) in the antennal segments. *Scr* expression is largely abolished from T1, and is retained in patches of Mx1 and Mx2. *Ubx* expression expands anteriorly into all anterior segments. *Ubx* and *Scr* appear to be expressed in mutually-exclusive domains in Mx1 and Mx2.

C) Antibody stain using ms anti-Antp (8C11) in WT embryo at stage 21. The anterior boundary of Antp protein is the Mx2 segment.

D) Antibody stain using 8C11 in a *Pcl* mutant embryo at stage 21. Antp protein is found more anterior than the normal boundary, expanding into the nervous system of the An1, An2, Mn, and Mx1 segments.

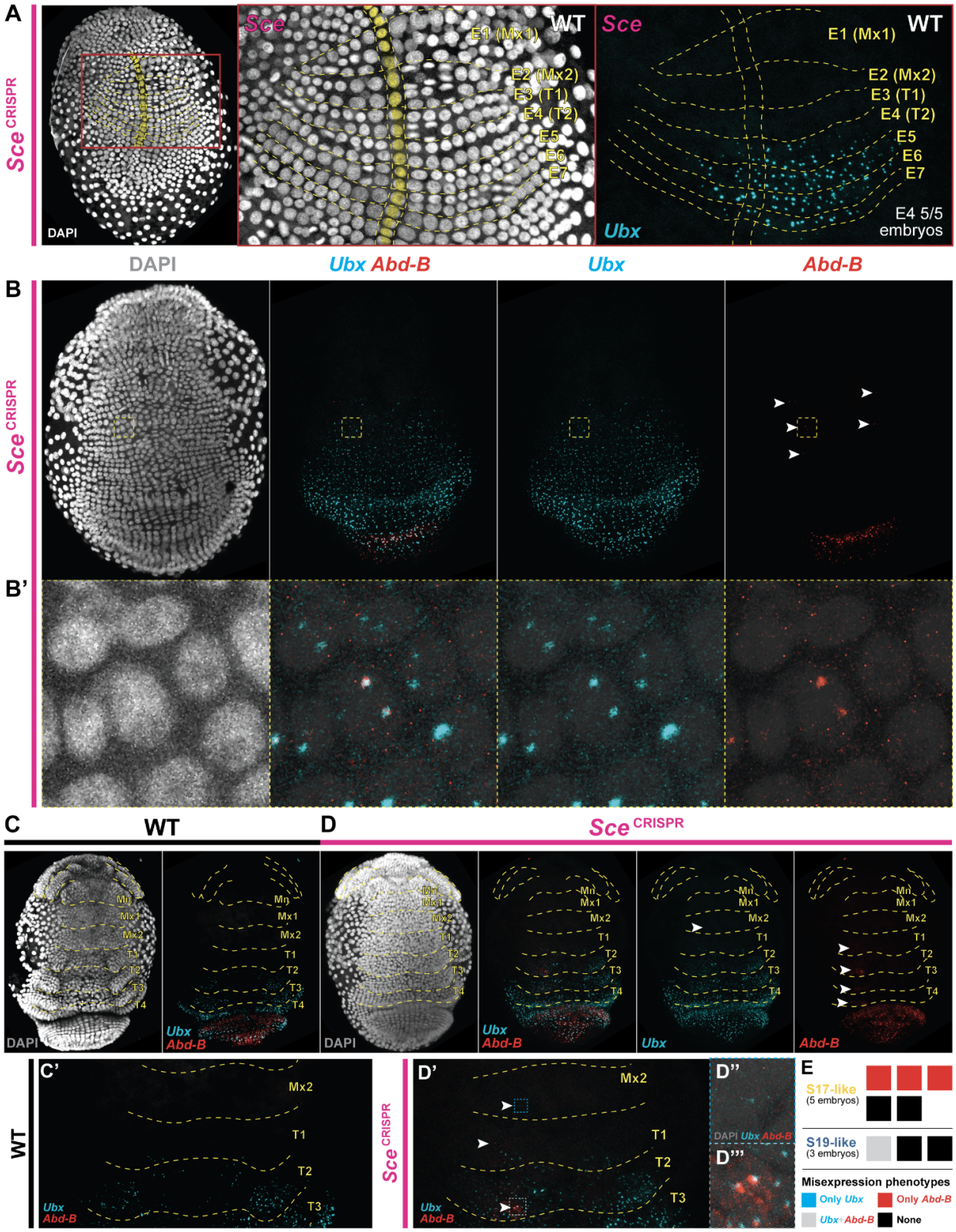

A
Sce CRISPR
DAPI
Sce
WT
E1 (Mx1)
E2 (Mx2)
E3 (T1)
E4 (T2)
E5
E6
E7
Sce
WT
E1 (Mx1)
E2 (Mx2)
E3 (T1)
E4 (T2)
E5
E6
E7
Ubx
E4 5/5 embryos
DAPI
Ubx Abd-B
Ubx
Abd-B
B
Sce CRISPR
B'
C
WT
D
Sce CRISPR
Mn
Mx1
Mx2
T1
T2
T3
T4
DAPI
Ubx
Abd-B
C'
WT
D'
Sce CRISPR
D''
DAPI Ubx Abd-B
D'''
E
S17-like
(5 embryos)
S19-like
(3 embryos)
Misexpression phenotypes
Only Ubx
Only Abd-B
Ubx+Abd-B
None

Fig. 2.6: Differential timing of PcG knockout-mediated misexpression for *Ubx* and *Abd-B*

A) *In situ* HCR of *Ubx* in a unilateral *Sce* mutant embryo. Based on a delayed pattern of cell divisions found in the right half of the embryo (embryo mounted ventrally; right half of embryo is on the left of the image), we infer that this half has lost *Sce* expression. The anterior boundary of *Ubx* expression between the *Sce* mutant and wildtype halves is the same, at parasegment E4, which primarily contributes to the T2 appendage. We observed a parasegment E4 anterior boundary of *Ubx* expression in 5/5 embryos examined at this stage.

B) *In situ* HCR of *Ubx* and *Abd-B* in an S17-like embryo. The very early stages of *Abd-B* expression have begun in the posterior of the embryo. Concurrently, *Abd-B* misexpression is found in the neurogenic ectoderm of anterior segments.

B') Inset region of interest showing clear *Ubx* and *Abd-B* transcription spots in the nucleus of an *Sce* S17-like mutant embryo. Previous work has shown that the *Parhyale Hox* cluster is a single cluster. *Ubx* and *Abd-B* transcription spots should be nearly co-localized within the nucleus.

C) *In situ* HCR of *Ubx* and *Abd-B* in a WT embryo at S19.

C') Inset region of Mx2, T1, T2, and T3 segments in embryo from panel C.

D) *In situ* HCR of *Ubx* and *Abd-B* in a *Sce* S19-like mutant embryo. Arrows in *Ubx* and *Abd-B* panels mark regions of misexpression.

D') Inset region of Mx2, T1, T2, and T3 segments in embryo from panel D.

D'') Inset region from panel D' showing *Ubx* misexpression spots in a cell found in the Mx2 segment.

D''') Inset region from panel D' showing *Ubx* and *Abd-B* misexpression in cells found in the T2 segment.

E) Schematic representation of *Ubx* and *Abd-B* misexpression patterns in S17-like and S19-like *Sce* mutant embryos. All S17-like mutants showing misexpression appeared to show only *Abd-B* misexpression. The single S19-like mutant embryo showed both *Ubx* and *Abd-B* misexpression.

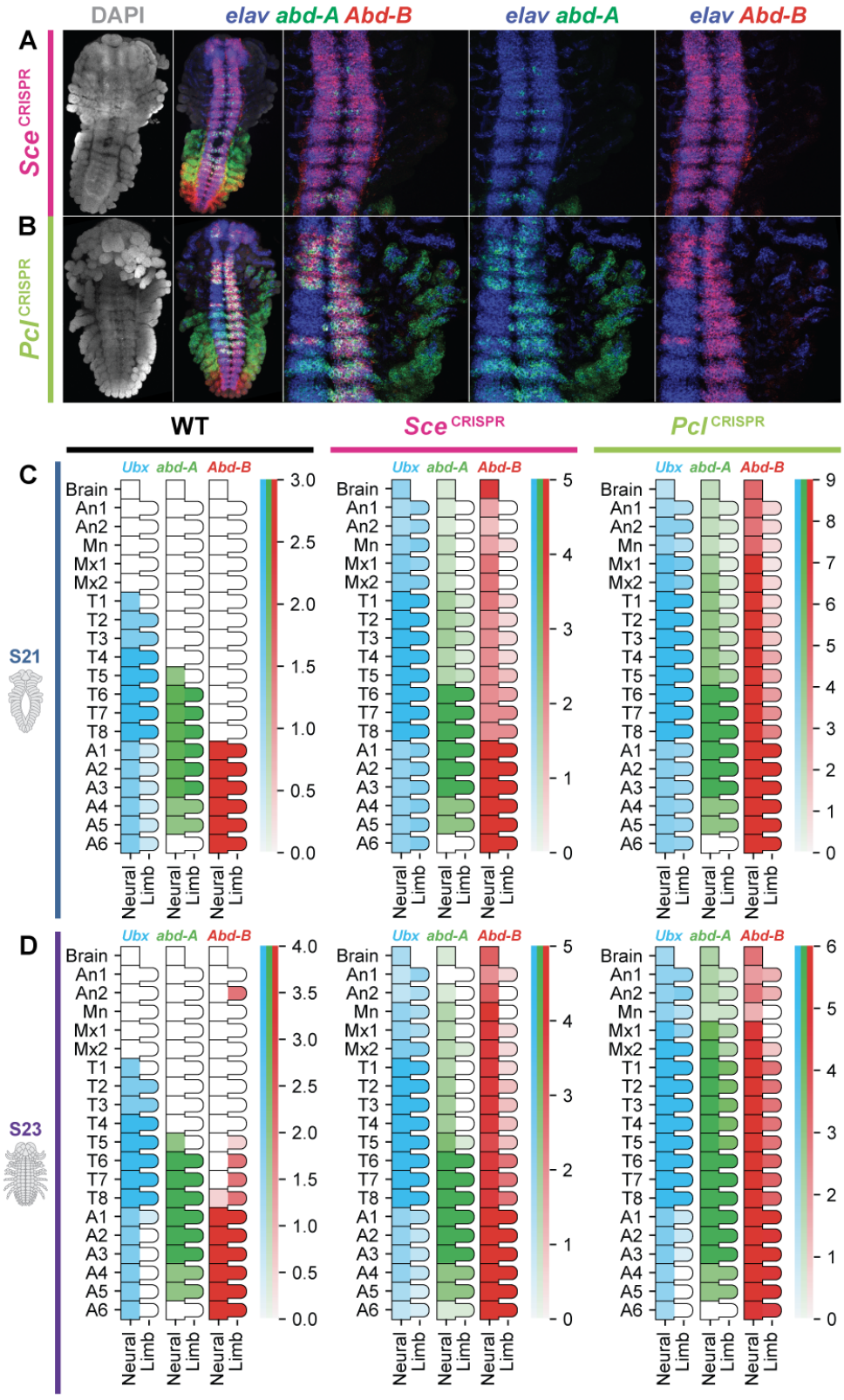

DAPI
elav abd-A Abd-B
elav abd-A
elav Abd-B
A
Sce CRISPR
B
Pcl CRISPR
WT
Sce CRISPR
Pcl CRISPR
C
S21
D
S23
Ubx abd-A Abd-B
Brain
An1
An2
Mn
Mx1
Mx2
T1
T2
T3
T4
T5
T6
T7
T8
A1
A2
A3
A4
A5
A6
Neural
Limb
3.0
2.5
2.0
1.5
1.0
0.5
0.0
5
4
3
2
1
0
9
8
7
6
4.0
3.5

Fig. 2.7: Differential tissue-specific misexpression of *Hox* genes between PcG mutants

A) *In situ* HCR of *abd-A* and *Abd-B* as well as the neural marker *elav* in a *Sce* mutant embryo at stage 23. *Abd-B* misexpression appears to be restricted to the nervous system. *abd-A* misexpression is found in a small number of nervous system cells.

B) *In situ* HCR of *abd-A* and *Abd-B* as well as the neural marker *elav* in a *Pcl* mutant embryo at stage 23. *Abd-B* misexpression appears to be primarily found in the nervous system. *abd-A* misexpression, in contrast to *Sce* mutants, is stronger in the nervous system and also occurs in patches of the limbs.

C) Quantification of *Ubx*, *abd-A*, and *Abd-B* expression in the limbs and nervous system of each segment of the body in WT, *Sce*, and *Pcl* embryos at stage 21. A region containing strong expression equivalent to the strongest WT expression was assigned a value of 1; a region containing patchy expression or expression less than the strongest WT expression was assigned a value of 0.5. 3 WT embryos, 5 *Sce* mutants, and 9 *Pcl* mutants were examined.

D) Quantification of *Ubx*, *abd-A*, and *Abd-B* expression in the limbs and nervous system of each segment of the body in WT, *Sce*, and *Pcl* embryos at stage 21. Analysis performed identically to panel C. 4 WT embryos, 5 *Sce* mutants, and 6 *Pcl* mutants were examined. Differences were observed between *abd-A* and *Abd-B* expression in *Sce* and *Pcl* mutants. *abd-A* expression in *Sce* mutants is weaker in the nervous system and largely absent from the limbs, as compared to in *Pcl* mutants. *Abd-B* misexpression was stronger in the limbs of *Sce* mutants, but was largely found in strips reminiscent of limb enervation.

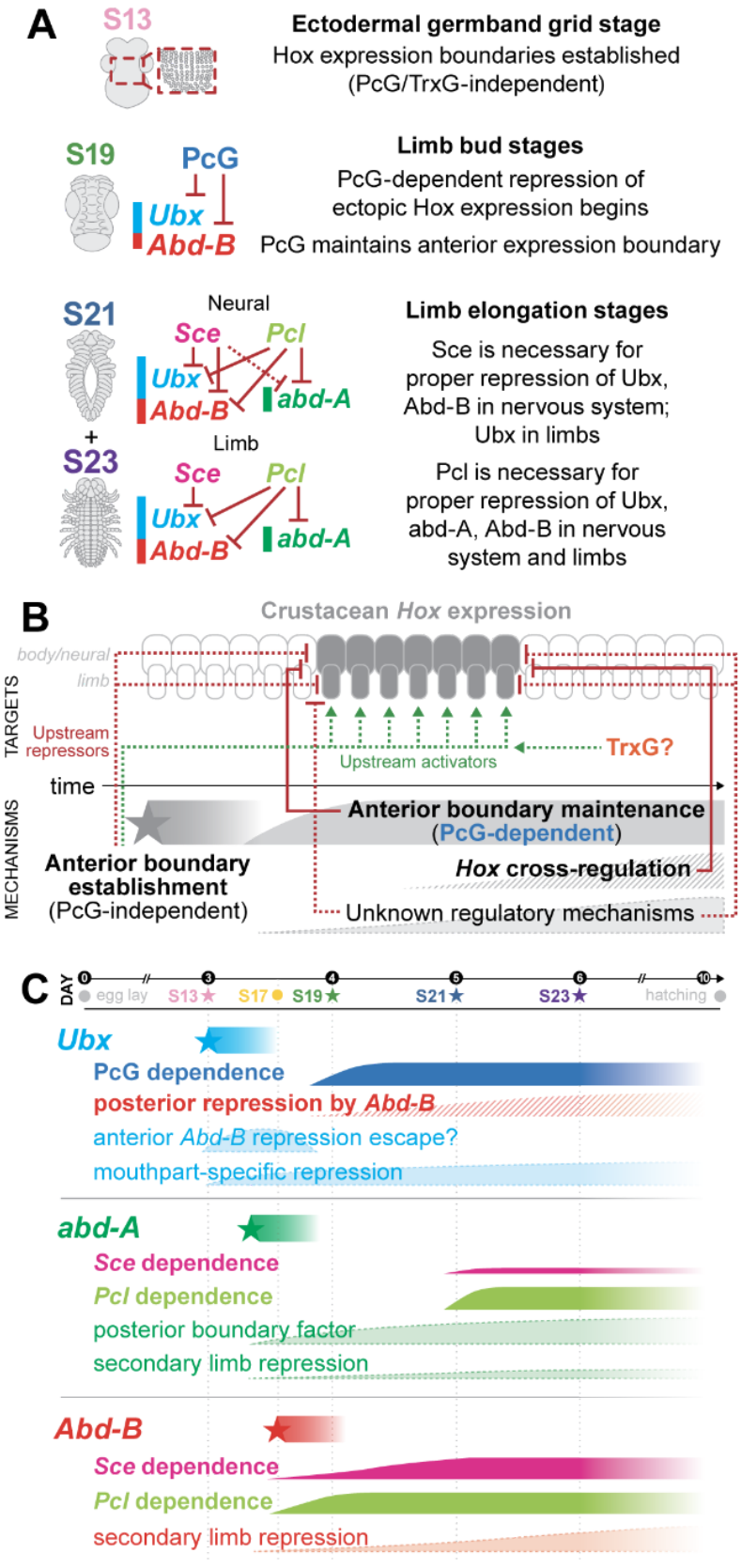

A
S13
Ectodermal germband grid stage
Hox expression boundaries established
(PcG/TrxG-independent)
S19
PcG
Ubx
Abd-B
Limb bud stages
PcG-dependent repression of
ectopic Hox expression begins
PcG maintains anterior expression boundary
S21
Neural
Sce
Pcl
Ubx
Abd-B
abd-A
+
S23
Limb
Sce
Pcl
Ubx
Abd-B
abd-A
Limb elongation stages
Sce is necessary for
proper repression of Ubx,
Abd-B in nervous system;
Ubx in limbs
Pcl is necessary for
proper repression of Ubx,
abd-A, Abd-B in nervous
system and limbs
B
Crustacean Hox expression
body/neural
limb
TARGETS
Upstream
repressors
Upstream activators
TrxG?
time
MECHANISMS
Anterior boundary maintenance
(PcG-dependent)
Anterior boundary
establishment
(PcG-independent)
Hox cross-regulation
Unknown regulatory mechanisms
C
DAY
0
egg lay
3
S13
S17
4
S19
5
S21
6
S23
10
hatching
Ubx
PcG dependence
posterior repression by Abd-B
anterior Abd-B repression escape?
mouthpart-specific repression
abd-A
Sce dependence
Pcl dependence
posterior boundary factor
secondary limb repression
Abd-B
Sce dependence
Pcl dependence
secondary limb repression

Fig. 2.8: Summary of findings and implications for *Hox* regulatory evolution

A) Summary of the phenotypes and inferred genetic relationships of *Hox* genes and PcG genes at S13, S19, S21, and S23.

B) Model of *Hox* regulation in crustaceans consisting of at least three mechanisms: anterior boundary establishment, anterior boundary maintenance (PcG-dependent), and *Hox* cross-regulation. Comparison of the results of PcG knockout to *Hox* gene knockout suggests that additional unknown regulatory mechanisms may also be involved in crustacean *Hox* regulation.

C) Summary of the *Hox* regulatory mechanisms for each of the three posterior *Hox* genes in *Parhyale*. *Ubx* anterior boundary establishment occurs at around S13, and the PcG-dependent phase of *Ubx* expression begins around S19. The inability of ectopic *Abd-B* to repress *Ubx* in anterior tissues suggests that anterior *Ubx* expression may achieve an autoregulatory state, preventing repression by *Abd-B*. The mutually exclusive expression domains of *Ubx* and *Scr* in the mouthparts suggest differential competence of mouthpart tissue to express *Ubx*. *abd-A* establishes its *Hox* expression between S13 and S17, but does not appear to become dependent on PcG repression until around S21. *abd-A* is differentially dependent on *Sce* and *Pcl* function for proper repression. An unknown factor may be responsible for the posterior boundary of *abd-A* expression, which is not perturbed by either *Abd-B* knockout or PcG loss of function. *Abd-B* establishes its anterior boundary and becomes dependent on PcG function at S19. *Abd-B* misexpression induced by PcG knockout appears to be primarily restricted to neural tissue. An unknown mechanism may repress *Abd-B* expression in the anterior limbs.

Table 1: Summary of PcG and TrxG CRISPR phenotypes

Gene	Family	Complex	Homeosis	Plate Defects	Limb Defects	Other Phenotypes
Polycomb (Pc)	PcG	PRC1	T1 -> T4/5 T2/3 -> T4/5 A1-3 -> A4-6	Missing T2 coxal and tergal plates, unfused plates, asymmetric plate fusion	Deformed antenna, deformed mouthparts, deformed limbs, proximal-distal fused limbs	–
Sex combs extra (Sce)	PcG	PRC1, dRAF	T1 -> T4/5 T2/3 -> T4/5	Missing T2 coxal and tergal plates, reduced basal plates, unfused tergal plates	Deformed antenna, deformed mouthparts, deformed limbs, proximal-distal fused limbs	–
Posterior sex combs-1 (Psc-1)	PcG	?	–	–	–	–
Posterior sex combs-2 (Psc-2)	PcG	PRC1, dRAF	Not observable (no limbs)	Loss of plates	Loss or major reduction of limbs	High embryonic lethality
Suppressor of zeste 12 (Su(z)12)	PcG	PRC2	An -> T4/5 Mp -> T4/5 T1 -> T4/5 T2/3 -> T4/5	Missing T2 coxal and tergal plates, reduced basal plates, unfused tergal plates	Deformed antenna, deformed mouthparts, deformed limbs, proximal-distal fused limbs, twisted limbs	–
Enhancer of zeste (E(z))	PcG	PRC2	T2/3 -> T4/5	Reduced basal and coxal plates	Deformed antenna, deformed mouthparts, deformed limbs, proximal-distal fused limbs, twisted limbs	Major malformations, possible body segment fusion
Extra sex combs (Esc)	PcG	PRC2	T2/3 -> T4/5	Reduced basal and coxal plates	Deformed antenna, deformed mouthparts, deformed limbs, proximal-distal fused limbs, twisted limbs	Major malformations
Polycomb-like (Pcl)	PcG	PRC2.1	An -> T4/5 Mp -> T4/5 T1 -> T4/5 T2/3 -> T4/5	Missing T2 coxal and tergal plates, reduced basal and coxal plates, unfused tergal plates	Deformed antenna, deformed mouthparts, deformed limbs, proximal-distal fused limbs, twisted limbs	–
Pleiohometic (pho)	PcG	PhoRC	T1 -> T4/5 T2/3 -> T4/5 Coxal plate T2-5 -> posterior plates	Reduced basal and coxal plates	Deformed antenna, deformed mouthparts, deformed limbs, proximal-distal fused limbs, twisted limbs	–
Trithorax (Trx)	TrxG	TAC1	–	–	–	–
Brahma (Brm)	TrxG	PBAP, BAP	–	Missing plates	Missing limbs	Embryonic lethality, major malformations
Kismet (Kis)	TrxG	–	–	–	–	–
Absent, small, or homeotic discs 2 (Ash2)	TrxG	COMPASS, COMPASS-like	–	Reduced coxal, tergal, basal plates; unfused tergal plates	Rounded limbs	Major malformations
Female sterile (1) homeotic (Fs1h)	TrxG	–	–	–	–	Embryonic lethality
Kohtalo (kto) aka Med12	TrxG (?)	Mediator	–	Unfused tergal plates	Deformed limbs	Embryonic lethality, major malformations
Skuld (skd) aka Med13	TrxG (?)	Mediator	–	Unfused tergal plates	Deformed limbs	Embryonic lethality, major malformations
CTCF	Other	–	–	–	–	–

2.5 – Methods

BLAST identification of PcG/TrxG genes

General strategy outlined in Supp Fig. 1.2b. We identified core PcG/TrxG genes from the *Drosophila* and *Mus* literature, with special attention to those genes that had previously been shown to induce homeotic phenotypes (Supp Tables 1 and 2). We downloaded peptide sequences for these genes from UNIPROT and used tblastn to identify potential best hits to the *Parhyale* genome. For each peptide, we used BLAST against 5 different *Parhyale* transcript sources: the *Parhyale* MAKER genome annotation (acquired from Leo Blondel), the Kao *et al.* Mikado transcriptome, the Sun *et al.* Mikado transcriptome, the Sun *et al.* StringTie2 short + long read embryonic transcriptome, and the Trinity-Limb transcriptome from the Sun *et al.* manuscript (generated by Heather Bruce and Jessen Bredeson). Several groups of genes appeared to have shared BLAST hits, including *Trx*, *Setd1*, and *Trr*. We disambiguated any inconclusive results using reciprocal best BLAST searches using the online BLAST portal. The full list of the *Parhyale* PcG/TrxG genes we identified is found in Supp. Tables 3 and 4.

RNA-Seq expression quantification

We used RNA-Seq data from the Sun et al. 2022 manuscript for developmental stages S13, S19, S21, and S23 (three replicates for each stage). We quantified the expression of all transcripts using Kallisto alignment-free transcript abundance estimation to the Sun-Mikado transcriptome. For each PcG/TrxG gene, we used the best Mikado transcript BLAST hit.

CRISPR-Cas9 guide design

We used the CRISPR-Cas9 targeting tool in Benchling to choose guide RNA sequences. We designed guides targeting either a 5' portion of the ORF of a gene of interest, determined based on BLAST to all organisms to assess the true START codon, or to a specifically identifiable BLAST homology domain with suggested importance in gene function.

CRISPR-Cas9 mutagenesis

Cas9 ribonucleoprotein complexes were assembled using Cas9-NLS enzyme purchased from the QB3 MacroLab at Berkeley and guide RNAs purchased from Synthego. Cas9 enzyme was used at a final concentration of 333 ng/μL and guide RNAs were used at a final concentration of 100ng/μL for each guide RNA in multi-guide knockout experiments, and 200ng/μL for single-guide knockout experiments. Final guide injection mix also included 0.05% phenol red; all injection components were mixed in nuclease-free water. For most experiments, we used two guide RNAs targeting each gene of interest. To demonstrate the reproducibility of single guide RNAs to generate the same phenotypes, thus controlling for potential off-target effects, we performed single-guide knockouts of *Sce*, *Pcl*, and *Su(Z)12* (see Supp. Figs. 2.2, 2.3, and 2.4) and quantified the proportion of phenotypes we observed between single and double-guide knockout experiments. We recovered the same classes of phenotypes between the single-guide and double-guide knockout experiments, although the relative proportion of each phenotype and total mutagenic efficacy varied.

Embryo culture and dissection

Embryos were raised as previously described. After injection, embryos were transferred to filter-sterilized artificial sea water (FASW) made using Tropic Marin sea salt to a final salinity of ~31-35%. Embryos were raised at 27°C in a humidified chamber in an incubator. We performed regular water changes on embryos every 1-2 days, removing any dead embryos. Wildtype embryos usually hatch by 10-11dpf. For our experiments, if any embryos remained unhatched but alive at 14dpf, we sacrificed and dissected those embryos using sharpened tungsten needles. For *in situ* hybridization and immunofluorescence experiments, we dissected embryos in 9 parts FASW, 1 part PEM buffer, and 1 part 32% paraformaldehyde and fixed for 25-35 minutes at room temperature (~21°C) using previously described techniques.

Microscopy and phenotype analyses

Hatchlings were fixed in 9 parts FASW, 1 part PEM buffer, and 1 part 32% paraformaldehyde. After removing FASW and applying the 9:1:1 fix, we used a tungsten needle to pierce the gut tube, usually through the ventral side of the hatchling around the juncture between the head and thorax. This "throat punch" technique largely prevented gross morphological distortions often observed in fixed hatchlings. Hatchlings were fixed for 1hr at room temperature in 9:1:1 fix, or overnight at 4°C. After fixation, hatchlings were washed 3X with PT and transferred to 50% glycerol/ 50% 1XPBS, then 70% glycerol/ 30% 1X PBS for mounting. For limb dissection panels, limbs were dissected in the 70% glycerol/ 30% 1X PBS mounting medium. Hatchlings were imaged on a Zeiss LSM 780 confocal microscope using the 10X objective and 405 laser; hatchling cuticle is highly autofluorescent in the DAPI channel. Images were processed using FIJI. First, the Gamma value was adjusted to create a more visually uniform cuticle, and levels were adjusted to bring the background fluorescence to 0. A depth projection was then generated using the Stacks > Hyperstacks > Temporal Color Code command and the "Spectrum" LUT. Images were adjusted for additional brightness using the Levels command in Adobe Photoshop Creative Cloud.

***In situ* HCR probes**

We ordered *in situ* hybridization chain reaction probes for *Ph-Ubx* isoform 2, *Ph-abd-A* isoform 2, *Ph-Abd-B* isoform 2, *Ph-Scr*, and *Ph-elav* through Molecular Instruments (Lot #PRG221-225).

***In situ* hybridization chain reaction**

We performed *in situ* hybridization chain reaction on dissected embryos using the Bruce et al. modifications to the Choi et al. *in situ* hybridization chain reaction 3.0 method. Embryos were mounted as previously described and imaged using a Zeiss LSM 780 confocal microscope using a 10X or 20X objective. We substantially narrowed the detection range for the fluorophores for each probe to avoid bleedthrough between channels, using wildtype single-probe stains as a benchmark for expected expression boundaries. We utilized the same detection ranges for imaging our mutant and wildtype embryos. Z-stack maximum intensity projection images were generated using FIJI, and background (non-embryo tissue) fluorescence was adjusted to 0 using the Window/Level tool.

Hox expression analysis

For each embryo used for analysis in Fig. 7, we examined the expression of each *Hox* gene compared to the DAPI-stained morphology of each embryo. For each embryo, we evaluated whether any expression was observed in the medial region or the limbs. We assigned a 1 to expression equivalent to the highest wildtype expression (evaluated by eye), and a 0.5 to lower levels of expression or patchy expression. As embryos are symmetrical, but we observed mosaicism in our knockouts, if either half of the embryo showed any expression, we used that value as our measure for that region of the embryo. This allowed us to gauge the maximum possible extent of misexpression in mutants.

Acknowledgments

We would like to thank members of the Nipam Patel, Hernan Garcia, David Weisblat, and Iswar Hariharan labs for feedback on the experiments and analyses performed for this project. RJC and YT were supervised by Hernan Garcia during their undergraduate thesis work that contributed to this project.

Competing Interests

The authors declare no competing interests.

Contributions

DAS, YT, and RJC conceived of the project. YT, RJC, and DAS performed literature research to identify candidate PcG/TrxG genes. YT compiled literature summary tables. YT, DAS, and RJC performed bioinformatic analyses to identify PcG/TrxG genes, and DAS performed bioinformatic analyses to evaluate gene expression levels. DAS and RJC performed CRISPR injections. RJC, YT, and DAS raised and phenotyped mutant hatchlings. DAS and YT performed confocal microscopy on hatchlings. YT wrote scripts to visualize phenotype data with input from DAS. DAS performed embryo dissections and *in situ* hybridization and antibody stains. DAS wrote and revised the manuscript with input from YT, RJC, and NHP. NHP supervised project.

2.6 – References, Chapter 2

Akam, M. (1987). The molecular basis for metameric pattern in the Drosophila embryo. Development 101, 1–22.

Akam, M. (1987). The molecular basis for metameric pattern in the Drosophila embryo. Development *101*, 1–22. .

Akasaka, T., van Lohuizen, M., van der Lugt, N., Mizutani-Koseki, Y., Kanno, M., Taniguchi, M., Vidal, M., Alkema, M., Berns, A., and Koseki, H. (2001). Mice doubly deficient for the Polycomb Group genes Mel18 and Bmi1 reveal synergy and requirement for maintenance but not initiation of Hox gene expression. Development *128*, 1587–1597. https://doi.org/10.1242/dev.128.9.1587.

Almazán, A., Çevrim, Ç., Musser, J.M., Averof, M., and Paris, M. (2021). Regenerated crustacean limbs are precise replicas. BioRxiv 2021.12.13.472338. https://doi.org/10.1101/2021.12.13.472338.

Angelini, D.R., and Kaufman, T.C. (2005). Comparative Developmental Genetics and the Evolution of Arthropod Body Plans. Annu. Rev. Genet. *39*, 95–119. https://doi.org/10.1146/annurev.genet.39.073003.112310.

Averof, M., and Akam, M. (1995). Hox genes and the diversification of insect and crustacean body plans. Nature *376*, 420–423. https://doi.org/10.1038/376420a0.

Averof, M., and Patel, N.H. (1997). Crustacean appendage evolution associated with changes in Hox gene expression. Nature *388*, 682. .

Breen, T.R., and Duncan, I.M. (1986). Maternal expression of genes that regulate the bithorax complex of Drosophila melanogaster. Dev. Biol. *118*, 442–456. https://doi.org/10.1016/0012-1606(86)90015-1.

Breen, T.R., and Harte, P.J. (1991). Molecular characterization of the trithorax gene, a positive regulator of homeotic gene expression in Drosophila. Mech. Dev. *35*, 113–127. https://doi.org/10.1016/0925-4773(91)90062-B.

Breen, T.R., and Harte, P.J. (1993). Trithorax regulates multiple homeotic genes in the bithorax and Antennapedia complexes and exerts different tissue-specific, parasegment-specific and promoter-specific effects on each. Development *117*, 119–134. https://doi.org/10.1242/dev.117.1.119.

Brinkmeier, M.L., Geister, K.A., Jones, M., Waqas, M., Maillard, I., and Camper, S.A. (2015). The Histone Methyltransferase Gene Absent, Small, or Homeotic Discs-1 Like Is Required for Normal Hox Gene Expression and Fertility in Mice1. Biol. Reprod. *93*, 121, 1–12. https://doi.org/10.1095/biolreprod.115.131516.

Brown, J.L., Fritsch, C., Mueller, J., and Kassis, J.A. (2003). The Drosophila pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing. Development *130*, 285–294. https://doi.org/10.1242/dev.00204.

Bruce, H.S., and Patel, N.H. (2020). Knockout of crustacean leg patterning genes suggests that insect wings and body walls evolved from ancient leg segments. Nat. Ecol. Evol. *4*, 1703–1712. https://doi.org/10.1038/s41559-020-01349-0.

Bruce, H., Jerz, G., Kelly, S., McCarthy, J., Pomerantz, A., Senevirathne, G., Sherrard, A., Sun, D., Wolff, C., and Patel, N. (2021). Hybridization Chain Reaction (HCR) In Situ Protocol. Protocols.Io.

Brunk, B.P., Martin, E.C., and Adler, P.N. (1991). Molecular genetics of the Posterior sex combs/Suppressor 2 of zeste region of Drosophila: aberrant expression of the Suppressor 2 of zeste gene results in abnormal bristle development. Genetics *128*, 119–132. https://doi.org/10.1093/genetics/128.1.119.

Chambeyron, S., and Bickmore, W.A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. Genes Dev. *18*, 1119–1130. https://doi.org/10.1101/gad.292104.

Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., and Pierce, N.A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development *145*, dev165753. https://doi.org/10.1242/dev.165753.

Clark-Hachtel, C.M., and Tomoyasu, Y. (2020). Two sets of candidate crustacean wing homologues and their implication for the origin of insect wings. Nat. Ecol. Evol. *4*, 1694–1702. https://doi.org/10.1038/s41559-020-1257-8.

de Graaff Wim, Tomotsune Daihachiro, Oosterveen Tony, Takihara Yoshihiro, Koseki Haruhiko, and Deschamps Jacqueline (2003). Randomly inserted and targeted Hox/reporter fusions transcriptionally silenced in Polycomb mutants. Proc. Natl. Acad. Sci. *100*, 13362–13367. https://doi.org/10.1073/pnas.2237046100.

Denell, R.E. (1978). Homoeosis in Drosophila. II. A genetic analysis of Polycomb. Genetics *90*, 277. .

Deschamps, J., and van Nes, J. (2005). Developmental regulation of the Hox genes during axial morphogenesis in the mouse. Development *132*, 2931–2942. https://doi.org/10.1242/dev.01897.

Deschamps, J., and Wijgerde, M. (1993). Two Phases in the Establishment of HOX Expression Domains. Dev. Biol. *156*, 473–480. https://doi.org/10.1006/dbio.1993.1093.

Deutsch, J.S., and Mouchel-Vielh, E. (2003). Hox genes and the crustacean body plan. BioEssays *25*, 878–887. https://doi.org/10.1002/bies.10319.

Erokhin, M., Georgiev, P., and Chetverina, D. (2018). Drosophila DNA-Binding Proteins in Polycomb Repression. Epigenomes *2*. https://doi.org/10.3390/epigenomes2010001.

Fauvarque, M.-O., Laurenti, P., Boivin, A., Bloyer, S., Griffin-Shea, R., Bourbon, H.-M., and Dura, J.-M. (2001). Dominant modifiers of the polyhomeotic extra-sex-combs phenotype induced by marked P element insertional mutagenesis in Drosophila. Genet. Res. *78*, 137–148. .

Geisler, S.J., and Paro, R. (2015). Trithorax and Polycomb group-dependent regulation: a tale of opposing activities. Development *142*, 2876–2887. https://doi.org/10.1242/dev.120030.

Gerberding, M., Browne, W., and Patel, N. (2002). Cell lineage analysis of the amphipod crustacean Parhyale hawaiensis reveals an early restriction of cell fates. Development *129*, 5789–5801. https://doi.org/10.1242/dev.00155.

Guenther, M., Jenner, R., and Chevalier, B. (2005). Global and Hox-specific roles for the MLL1 methyltransferase. Proc. ... https://doi.org/10.1073/pnas.0503072102.

Hrycaj, S.M., and Wellik, D.M. (2016). Hox genes and evolution. F1000Res *5*, 859. https://doi.org/10.12688/f1000research.7663.1.

Hughes, C.L., and Kaufman, T.C. (2002). Hox genes and the evolution of the arthropod body plan. Evol. Dev. *4*, 459–499. https://doi.org/10.1046/j.1525-142X.2002.02034.x.

Irish, V.F., Martinez-Arias, A., and Akam, M. (1989). Spatial regulation of the Antennapedia and Ultrabithorax homeotic genes during Drosophila early development. EMBO J. *8*, 1527–1537. .

Isono Kyo-ichi, Fujimura Yu-ichi, Shinga Jun, Yamaki Makoto, O-Wang Jiyang, Takihara Yoshihiro, Murahashi Yasuaki, Takada Yuki, Mizutani-Koseki Yoko, and Koseki Haruhiko (2005). Mammalian Polyhomeotic Homologues Phc2 and Phc1 Act in Synergy To Mediate Polycomb Repression of Hox Genes. Mol. Cell. Biol. *25*, 6694–6706. https://doi.org/10.1128/MCB.25.15.6694-6706.2005.

Jaeger, J. (2011). The gap gene network. Cell. Mol. Life Sci. *68*, 243–274. https://doi.org/10.1007/s00018-010-0536-y.

Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in Drosophila. Genetics *206*, 1699–1725. https://doi.org/10.1534/genetics.115.185116.

Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. Proc. Natl. Acad. Sci. *85*, 8136–8140. .

Kmita, M., and Duboule, D. (2003). Organizing axes in time and space; 25 years of colinear tinkering. Science *301*, 331–333. .

Krumlauf, R. (1994). Hox genes in vertebrate development. Cell *78*, 191–201. https://doi.org/10.1016/0092-8674(94)90290-9.

Kurzhals, R.L., Tie, F., Stratton, C.A., and Harte, P.J. (2008). Drosophila ESC-like can substitute for ESC and becomes required for Polycomb silencing if ESC is absent. Dev. Biol. *313*, 293–306. https://doi.org/10.1016/j.ydbio.2007.10.025.

Liubicich, D.M., Serano, J.M., Pavlopoulos, A., Kontarakis, Z., Protas, M.E., Kwan, E., Chatterjee, S., Tran, K.D., Averof, M., and Patel, N.H. (2009). Knockdown of Parhyale Ultrabithorax recapitulates evolutionary changes in crustacean appendage morphology. Proc Natl Acad Sci USA *106*, 13892–13896. https://doi.org/10.1073/pnas.0903105106.

Lo Stanley M., Ahuja Nitin K., and Francis Nicole J. (2009). Polycomb Group Protein Suppressor 2 of Zeste Is a Functional Homolog of Posterior Sex Combs. Mol. Cell. Biol. *29*, 515–525. https://doi.org/10.1128/MCB.01044-08.

Maeda, R.K., and Karch, F. (2006). The ABC of the BX-C: the bithorax complex explained. Development *133*, 1413–1422. https://doi.org/10.1242/dev.02323.

Maeda, R.K., and Karch, F. (2015). The open for business model of the bithorax complex in Drosophila. Chromosoma *124*, 293–307. https://doi.org/10.1007/s00412-015-0522-0.

Mallo, M., and Alonso, C.R. (2013). The regulation of Hox gene expression during animal development. Development *140*, 3951–3963. https://doi.org/10.1242/dev.068346.

del Mar Lorente, M., Marcos-Gutierrez, C., Perez, C., Schoorlemmer, J., Ramirez, A., Magin, T., and Vidal, M. (2000). Loss- and gain-of-function mutations show a polycomb group function for Ring1A in mice. Development *127*, 5093–5100. https://doi.org/10.1242/dev.127.23.5093.

Martin, E.C., and Adler, P.N. (1993). The Polycomb group gene Posterior Sex Combs encodes a chromosomal protein. Development *117*, 641–655. https://doi.org/10.1242/dev.117.2.641.

Martin, A., Serano, J.M., Jarvis, E., Bruce, H.S., Wang, J., Ray, S., Barker, C.A., C, O., Liam, and Patel, N.H. (2015). CRISPR/Cas9 Mutagenesis Reveals Versatile Roles of Hox Genes in Crustacean Limb Specification and Evolution. Curr Biol *26*, 14–26. https://doi.org/10.1016/j.cub.2015.11.021.

Matsuoka, Y., Bando, T., Watanabe, T., Ishimaru, Y., Noji, S., Popadić, A., and Mito, T. (2015). Short germ insects utilize both the ancestral and derived mode of Polycomb group-mediated epigenetic silencing of Hox genes. Biol. Open *4*, 702–709. https://doi.org/10.1242/bio.201411064.

Ohno, K., McCabe, D., Czermin, B., Imhof, A., and Pirrotta, V. (2008). ESC, ESCL and their roles in Polycomb Group mechanisms. Mech. Dev. *125*, 527–541. https://doi.org/10.1016/j.mod.2008.01.002.

Panganiban, G., Sebring, A., Nagy, L., and Carroll, S. (1995). The Development of Crustacean Limbs and the Evolution of Arthropods. Science *270*, 1363–1366. https://doi.org/10.1126/science.270.5240.1363.

Pavlopoulos, A., Kontarakis, Z., Liubicich, D.M., Serano, J.M., Akam, M., Patel, N.H., and Averof, M.

(2009). Probing the evolution of appendage specialization by Hox gene misexpression in an emerging model crustacean. Proc Natl Acad Sci USA *106*, 13897–13902. https://doi.org/10.1073/pnas.0902804106.

Peifer, M., Karch, F., and Bender, W. (1987). The bithorax complex: control of segmental identity. Genes Dev. *1*, 891–898. https://doi.org/10.1101/gad.1.9.891.

Price, A.L., Modrell, M.S., Hannibal, R.L., and Patel, N.H. (2010). Mesoderm and ectoderm lineages in the crustacean Parhyale hawaiensis display intra-germ layer compensation. Dev Biol *341*, 256–266. https://doi.org/10.1016/j.ydbio.2009.12.006.

Puro, J., and Nygrén, T. (1975). Mode of action of a homoeotic gene in Drosophila melanogaster: Localization and dosage effects of Polycomb. Hereditas *81*, 237–247. .

Rastegar Mojgan, Kobrossy Laila, Kovacs Erzsebet Nagy, Rambaldi Isabel, and Featherstone Mark (2004). Sequential Histone Modifications at Hoxd4 Regulatory Regions Distinguish Anterior from Posterior Embryonic Compartments. Mol. Cell. Biol. *24*, 8090–8103. https://doi.org/10.1128/MCB.24.18.8090-8103.2004.

Schmid, B. (2011). Molecular Studies on Head Development of the Amphipod Crustacean Parhyale hawaiensis (Georg-August-University Göttingen).

Schuettengruber, B., Bourbon, H.-M., Di Croce, L., and Cavalli, G. (2017). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. Cell *171*, 34–57. https://doi.org/10.1016/j.cell.2017.08.002.

Schwentner, M., Combosch, D.J., Pakes Nelson, J., and Giribet, G. (2017). A Phylogenomic Solution to the Origin of Insects by Resolving Crustacean-Hexapod Relationships. Curr. Biol. *27*, 1818-1824.e5. https://doi.org/10.1016/j.cub.2017.05.040.

Schwentner, M., Richter, S., Rogers, D.C., and Giribet, G. (2018). Tetraconatan phylogeny with special focus on Malacostraca and Branchiopoda: highlighting the strength of taxon-specific matrices in phylogenomics. Proc. R. Soc. B Biol. Sci. *285*. https://doi.org/10.1098/rspb.2018.1524.

Serano, J.M., Martin, A., Liubicich, D.M., Jarvis, E., Bruce, H.S., La, K., Browne, W.E., Grimwood, J., and Patel, N.H. (2016). Comprehensive analysis of Hox gene expression in the amphipod crustacean Parhyale hawaiensis. Dev Biol *409*, 297–309. https://doi.org/10.1016/j.ydbio.2015.10.029.

Shiga, Y., Sagawa, K., Takai, R., Sakaguchi, H., Yamagata, H., and Hayashi, S. (2006). Transcriptional readthrough of Hox genes Ubx and Antp and their divergent post-transcriptional control during crustacean evolution. Evol. Dev. *8*, 407–414. https://doi.org/10.1111/j.1525-142X.2006.00114.x.

Simon, J., Chiang, A., and Bender, W. (1992). Ten different Polycomb group genes are required for spatial control of the abdA and AbdB homeotic products. Development *114*, 493–505. https://doi.org/10.1242/dev.114.2.493.

Soshnikova, N., and Duboule, D. (2009). Epigenetic temporal control of mouse Hox genes in vivo. Science *324*, 1320–1323. .

Struhl, G., and Akam, M. (1985). Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of Drosophila. EMBO J. *4*, 3259–3264. .

Struhl, G., and White, R.A.H. (1985). Regulation of the Ultrabithorax gene of drosophila by other

bithorax complex genes. Cell *43*, 507–519. https://doi.org/10.1016/0092-8674(85)90180-1.

Sun, D.A., Bredeson, J.V., Bruce, H.S., and Patel, N.H. (2021). Identification and classification of cis-regulatory elements in the amphipod crustacean Parhyale hawaiensis. BioRxiv 2021.09.16.460328. https://doi.org/10.1101/2021.09.16.460328.

Tschopp, P., and Duboule, D. (2011). A genetic approach to the transcriptional regulation of Hox gene clusters. Annu. Rev. Genet. *45*, 145–166. .

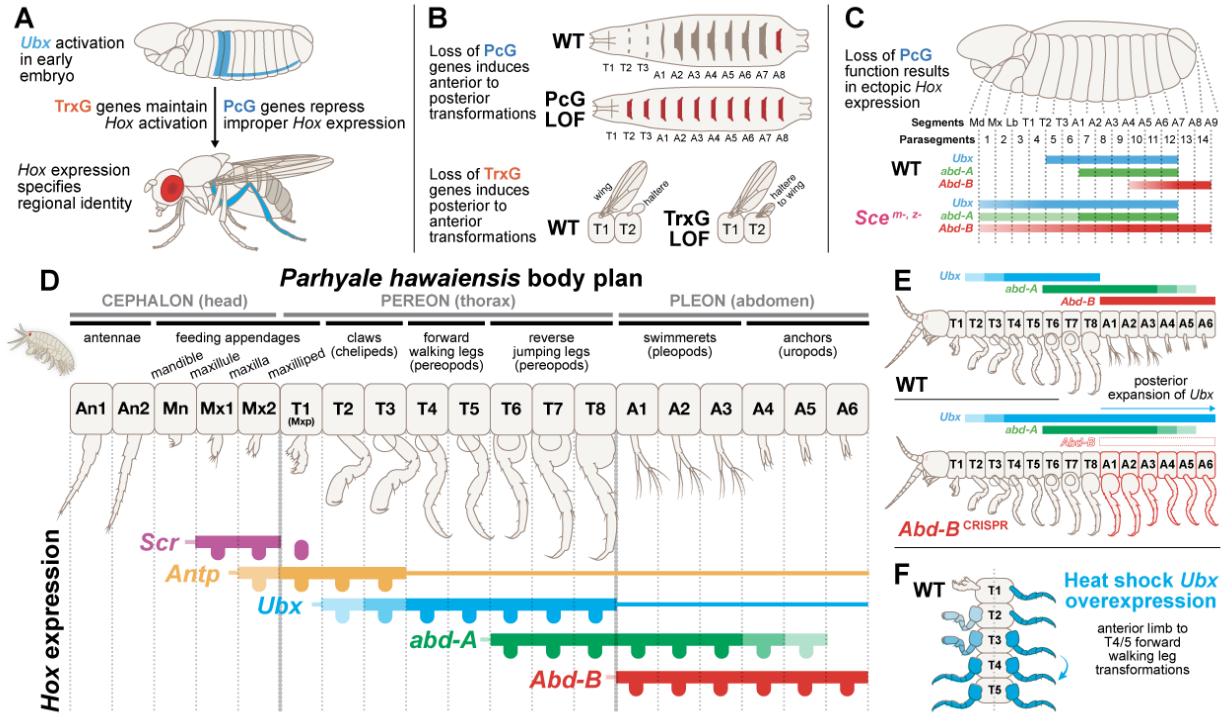
Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. Nature *431*, 873–878. .

Wang, S., He, F., Xiong, W., Gu, S., Liu, H., Zhang, T., Yu, X., and Chen, Y. (2007). Polycomblike-2-deficient mice exhibit normal left–right asymmetry. Dev. Dyn. *236*, 853–861. https://doi.org/10.1002/dvdy.21070.

Wedeen, C., Harding, K., and Levine, M. (1986). Spatial regulation of Antennapedia and bithorax gene expression by the Polycomb locus in Drosophila. Cell *44*, 739–748. .

Wolfe, J.M., Breinholt, J.W., Crandall, K.A., Lemmon, A.R., Lemmon, E.M., Timm, L.E., Siddall, M.E., and Bracken-Grissom, H.D. (2019). A phylogenomic framework, evolutionary timeline and genomic resources for comparative studies of decapod crustaceans. Proc. R. Soc. B Biol. Sci. *286*, 20190079. https://doi.org/10.1098/rspb.2019.0079.

Yu Benjamin D., Hanson Robin D., Hess Jay L., Horning Susan E., and Korsmeyer Stanley J. (1998). MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. Proc. Natl. Acad. Sci. *95*, 10632–10636. https://doi.org/10.1073/pnas.95.18.10632.



Supp. Fig. 2.1.1: Background

A) Summary model of phases of *Hox* regulation as described in *Drosophila*. *Hox* expression boundaries are specified in the early embryo via transcription factor-based mechanisms. As development progresses, PcG and TrxG genes are required to maintain proper *Hox* expression through repressive and activating mechanisms, respectively. The presence of *Hox* effectors in given body regions throughout development helps specify regional identity.

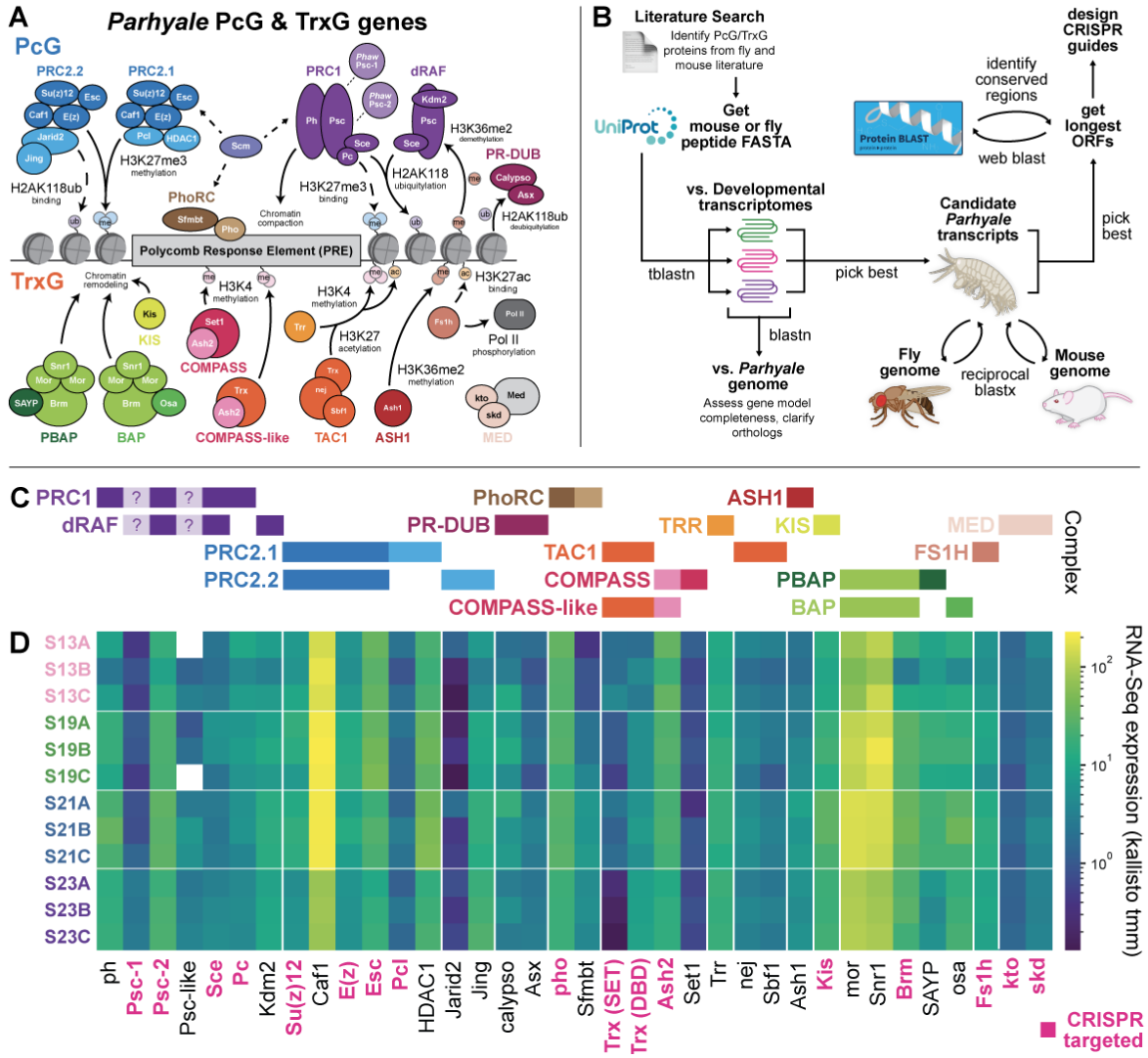
B) Summary of PcG and TrxG loss of function phenotypes. Loss of PcG genes frequently results in all body segments transformed to the most posterior body segment. Loss of TrxG genes can induce posterior to anterior transformations, such as haltere to wing transformation.

C) Example of *Hox* misexpression in PcG mutants. Loss of function of *Sce* results in broad misexpression of multiple *Hox* genes, with *Abd-B* overriding the other *Hox* identities to cause transformations of all thoracic segments to the A8 identity.

D) Limb identities and expression of five *Hox* genes (*Scr*, *Antp*, *Ubx*, *abd-A*, and *Abd-B*) in the wildtype *Parhyale* body plan.

E) Summary of *Abd-B* CRISPR knockout experiments from M(Martin et al., 2015) and Jarvis et al., 2022. Loss of *Abd-B* results in expansion of *Ubx* to the posterior of the embryo, and subsequent homeotic transformations. This result indicates that *Abd-B* represses *Ubx* to set the posterior boundary of *Ubx* expression.

F) Summary of *Ubx* heat shock overexpression results from Pavlopoulos et al., 2009. Heat shock results in transformation of anterior appendages to T4/5 forward walking leg identity.



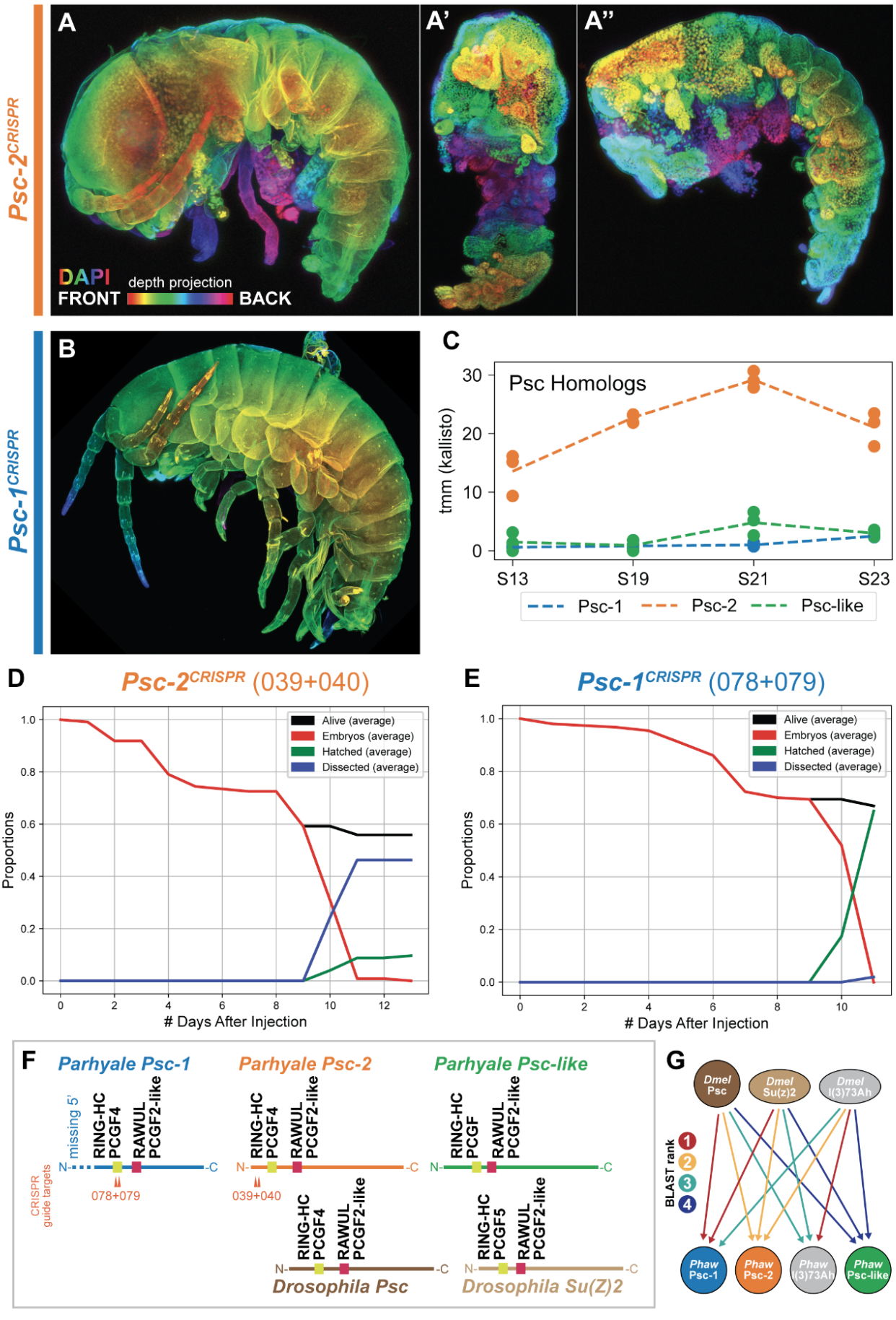
Supp. Fig. 2.1.2: Identification and expression of PcG/TrxG genes in Parhyale hawaiensis

A) Full diagram of PcG/TrxG genes found in the *Parhyale* genome.

B) Schematic of bioinformatic methods used to identify PcG/TrxG genes. A literature review was performed to identify the core components of PcG/TrxG genes in both *Mus musculus* and *Drosophila melanogaster*. Mouse and fly peptide sequences were obtained for each gene from UNIPROT and tblastn was used to identify the top BLAST hits in several transcript sources, including the phaw_5.0 MAKER genome annotation, the Kao et al. Mikado annotation, the Trinity-limb transcriptome from Sun et al., and the Sun et al. Mikado annotation. For each top candidate, the gene models were manually examined within the *Parhyale* genome to identify gene model completeness and clarify orthologs based on relative genomic location. We also identified the location of possible splice junctions at this step to avoid designing CRISPR guides that overlapped with slice junctions. After determining the best candidate locus and best gene model from across the transcript sources, we performed reciprocal BLAST to both the *Drosophila* and mouse genomes to confirm our top hits, and clarify potential paralog identities. Upon identifying the best candidate for each gene and paralog in the genome, we used online web BLAST to identify potential conserved protein domains and peptide sequences in order to design CRISPR guide RNAs.

C) Summary of PcG and TrxG complexes in the *Parhyale* genome.

D) Expression of each PcG/TrxG gene in the *Parhyale* genome across three replicates at four developmental stages (S13, S19, S21, S23). Transcriptome data from Sun et al. 2022 was used. All genes of interest showed expression at all timepoints.


Psc-2CRISPR
A
A'
A''
DAPI
depth projection
FRONT
BACK
Psc-1CRISPR
B
C
Psc Homologs
tmm (kallisto)
30
20
10
0
S13
S19
S21
S23
Psc-1
Psc-2
Psc-like
D
Psc-2CRISPR (039+040)
Alive (average)
Embryos (average)
Hatched (average)
Dissected (average)
Proportions
Days After Injection
E
Psc-1CRISPR (078+079)
F
Parhyale Psc-1
Parhyale Psc-2
Parhyale Psc-like
missing 5'
RING-HC
PCGF4
PCGF5
PCGF
RAWUL
PCGF2-like
CRISPR guide targets
078+079
039+040
Drosophila Psc
Drosophila Su(Z)2
G
Dmel Psc
Dmel Su(z)2
Dmel l(3)73Ah
BLAST rank
Phaw Psc-1
Phaw Psc-2
Phaw l(3)73Ah
Phaw Psc-like

Supp. Fig. 2.2.1: *Psc* family genes in *Parhyale*

A) *Psc-2* CRISPR knockout hatchlings. Hatchlings showed major morphological defects, including total loss of limbs and plates, with varying degrees of severity. A shows an apparently mosaic knockout, where some limbs and other body structures are visible. A' shows a severe phenotype from a ventral view, showing major reduction along the proximal-distal axis. A'' shows a severe phenotype from a lateral view, showing that body segments develop, but limbs do not.

B) *Psc-1* CRISPR knockout hatchling. We did not observe any noticeable phenotypes from dual-guide knockout of *Psc-1*.

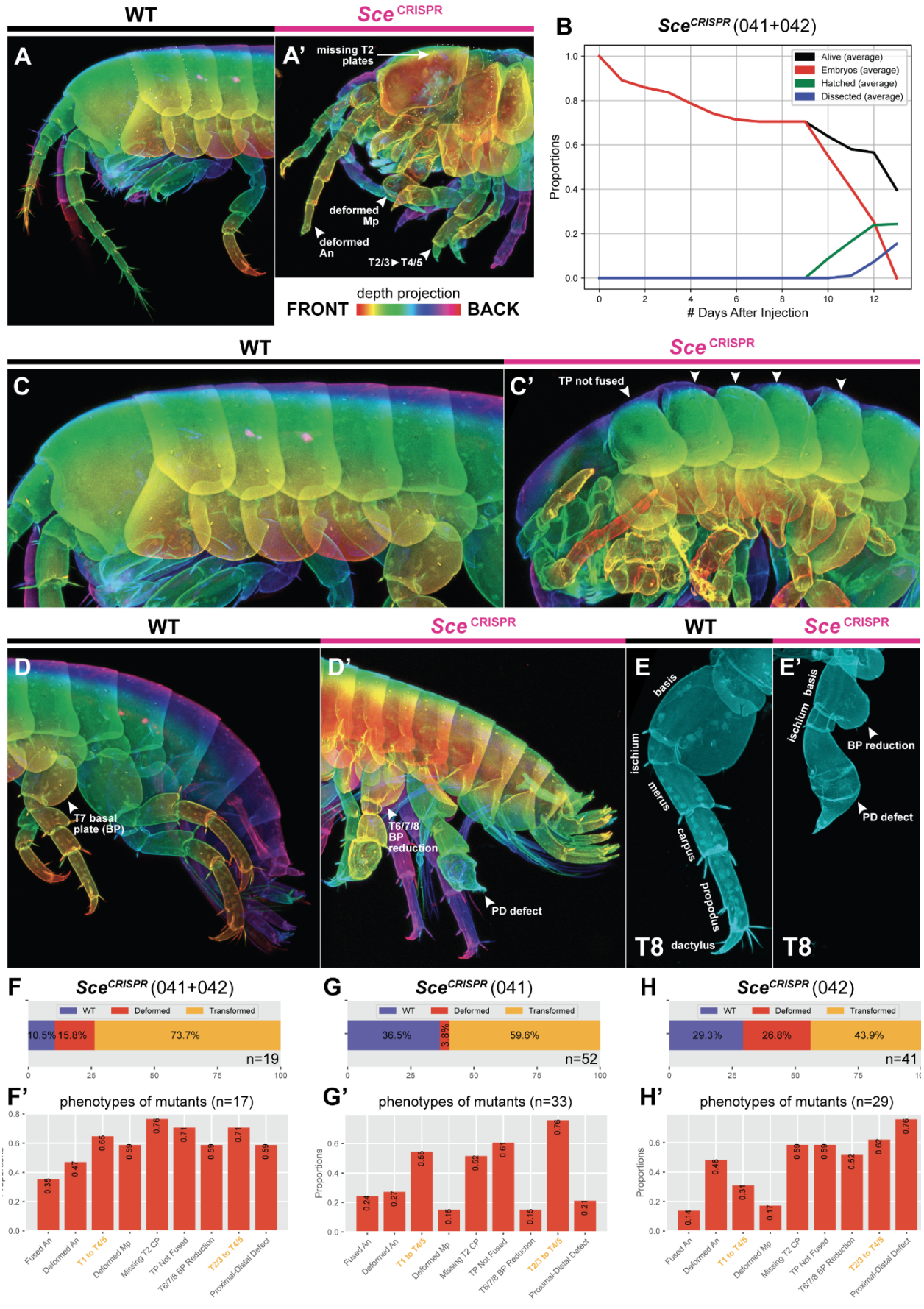
C) Expression of *Psc-1*, *Psc-2*, and *Psc-like* in the Sun et al. 2022 embryonic transcriptome dataset. *Psc-2* is expressed at a much higher level than *Psc-1* or *Psc-like*.

D) Average survivorship curve for *Psc-2* CRISPR experiments. The vast majority of surviving *Psc-2* hatchlings required dissection, and were not able to hatch on their own.

E) Average survivorship curve for *Psc-1* CRISPR experiments. Nearly all embryos surviving to late stages of development were able to hatch successfully.

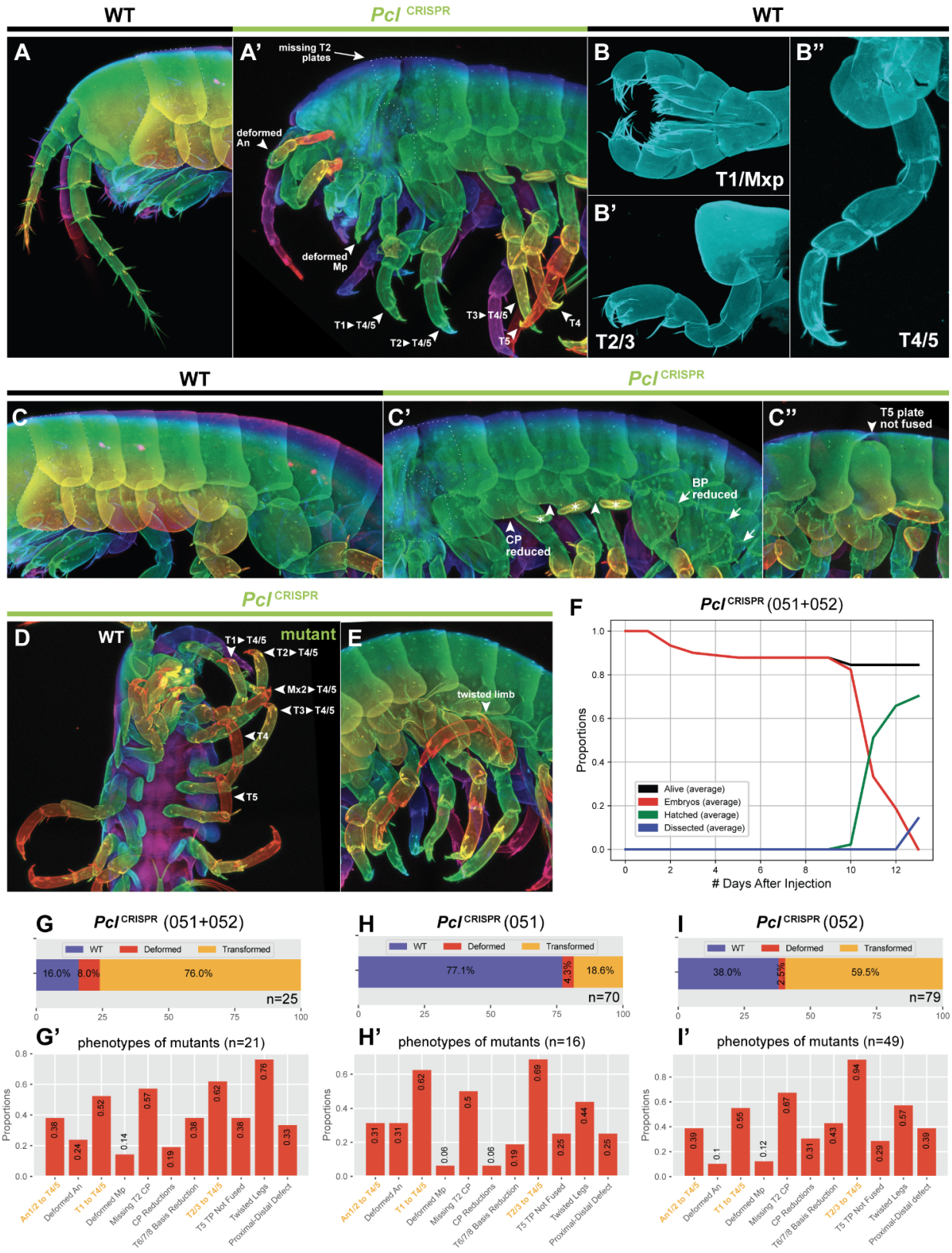
F) Diagram of *Parhyale Psc-1*, *Psc-2*, and *Psc-like* genes compared to *Drosophila Psc* and *Su(Z)2* genes. BLAST identified homology domains are placed in relative position along the peptide. All three *Parhyale Psc* homologs appear to have RING-HC PCGF domains and RAWUL PCGF2-like domains.

G) BLAST top hit mapping for *Drosophila* to *Parhyale* genes. *Parhyale Psc-1* is the top hit for both *Drosophila Psc* and *Su(Z)2*, while *Parhyale Psc-2* is the second best BLAST hit for *Drosophila Psc* and *Su(Z)2*. The third top BLAST hit for both *Drosophila Psc* and *Su(Z)2* is *Parhyale l(3)73Ah*, which was identified as being the top BLAST hit for *Drosophila l(3)73Ah*. A fourth gene, which we named *Parhyale Psc-like*, was a top blast hit for *Drosophila Psc*, *Su(Z)2*, and *l(3)73Ah*, but the reciprocal best BLAST hit for *Parhyale Psc-like* was *Drosophila Psc*. Given that *Parhyale Psc-1* and *Psc-2* are better reciprocal BLAST hits for *Drosophila Psc*, we termed this fourth gene *Parhyale Psc-like*.


WT
Sce CRISPR
A
A'
missing T2 plates
deformed Mp
deformed An
T2/3►T4/5
depth projection
FRONT
BACK
B
Sce CRISPR (041+042)
Alive (average)
Embryos (average)
Hatched (average)
Dissected (average)
Proportions
Days After Injection
C
C'
TP not fused
D
D'
T7 basal plate (BP)
T6/7/8 BP reduction
PD defect
E
E'
basis
ischium
merus
carpus
propodus
dactylus
T8
BP reduction
PD defect
F
Sce CRISPR (041+042)
WT
Deformed
Transformed
10.5%
15.8%
73.7%
n=19
G
Sce CRISPR (041)
36.5%
3.8%
59.6%
n=52
H
Sce CRISPR (042)
29.3%
26.8%
43.9%
n=41
F'
phenotypes of mutants (n=17)
G'
phenotypes of mutants (n=33)
H'
phenotypes of mutants (n=29)
Fused An
Deformed An
T1 to T4/5
Deformed Mp
Missing T2 CP
TP Not Fused
T6/7/8 BP Reduction
T2/3 to T4/5
Proximal-Distal Defect

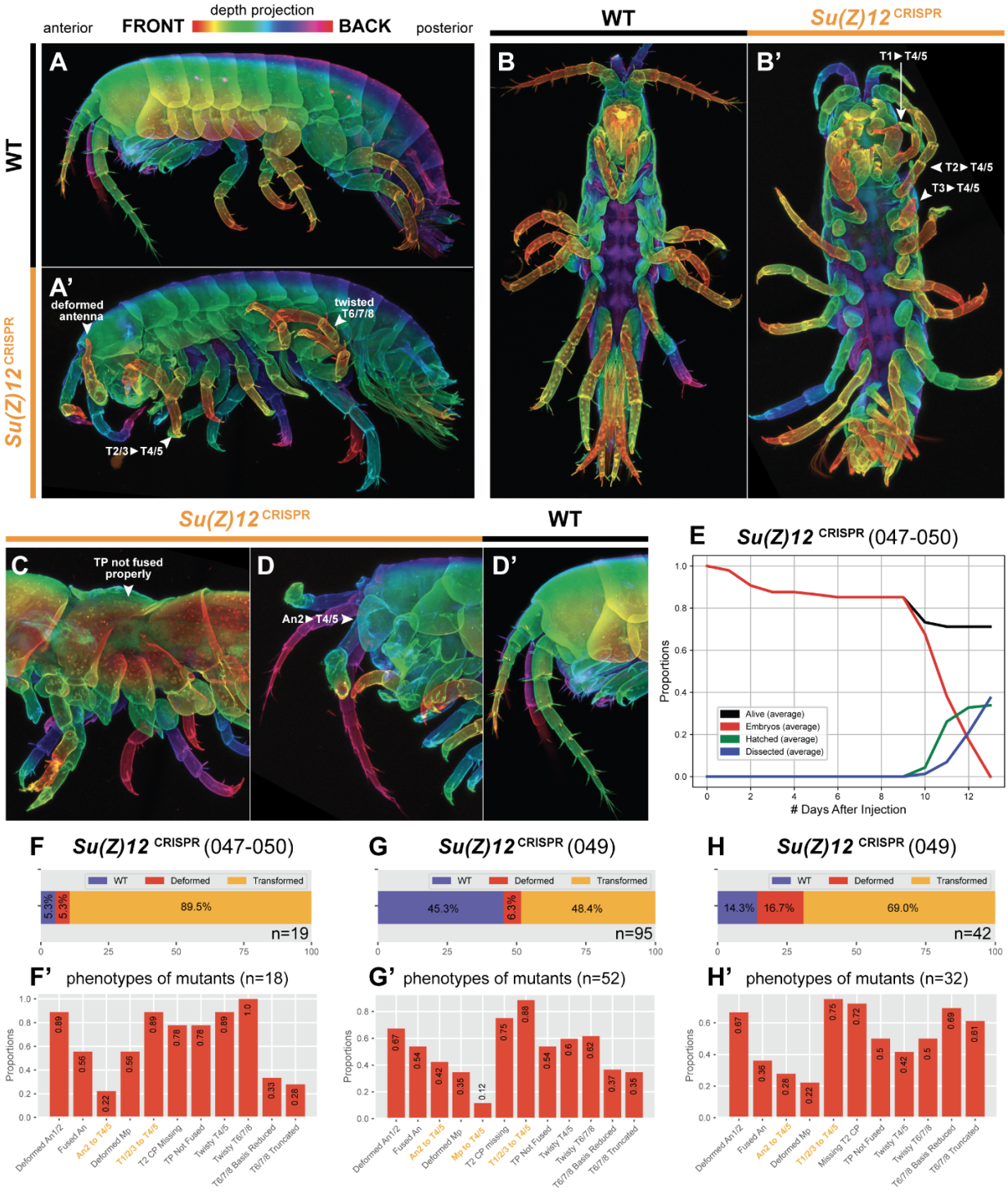
Supp. Fig. 2.2.2: *Sce* CRISPR phenotype data

A) Wildtype *Parhyale* hatchling head and anterior thorax.
A') *Sce* CRISPR hatchling head and anterior thorax showing various homeotic transformations and missing T2 tergal and coxal plates.
B) Average survivorship curve for dual-guide *Sce* CRISPR knockout experiments.
C) Wildtype *Parhyale* hatchling view focused on tergal and coxal plates.
C') *Sce* CRISPR hatchling showing defects in tergal plate fusion marked with arrowheads.
D) Wildtype *Parhyale* hatchling posterior, with T8 tergal plate marked with arrowhead.
D') *Sce* CRISPR hatchling showing tergal plate reduction and proximal-distal defects, marked with arrowheads.
E) Dissected wildtype T8 limb with leg segments labeled.
E') Dissected *Sce* CRISPR T8 limb with leg segments and limb defects labeled.
F) Proportion of WT, deformed, and transformed hatchlings for *Sce* dual-guide knockout (guide RNAs 041 and 042) from n=19 hatchlings.
F') Proportion of phenotypes from *Sce* (041+042) dual-guide knockout for n=17 hatchlings that we did not classify as wildtype. Yellow labels indicate phenotypes we classified as homeotic mutations.
G) Proportion of WT, deformed, and transformed hatchlings for *Sce* (041) single-guide knockout for n=52 hatchlings.
G') Proportion of phenotypes from *Sce* (041) single-guide knockout for n=33 mutant hatchlings.
H) Proportion of WT, deformed, and transformed hatchlings for *Sce* (042) single-guide knockout for n=41 hatchlings.
H') Proportion of phenotypes from *Sce* (042) single-guide knockout for n=29 mutant hatchlings.


WT
Pcl CRISPR
WT
A
A'
missing T2 plates
deformed An
deformed Mp
T1►T4/5
T2►T4/5
T3►T4/5
T4
T5
B
T1/Mxp
B'
T2/3
B''
T4/5
WT
Pcl CRISPR
C
C'
BP reduced
CP reduced
C''
T5 plate not fused
Pcl CRISPR
D
WT
mutant
T1►T4/5
T2►T4/5
Mx2►T4/5
T3►T4/5
T4
T5
E
twisted limb
F
Pcl CRISPR (051+052)
Proportions
Days After Injection
Alive (average)
Embryos (average)
Hatched (average)
Dissected (average)
G
Pcl CRISPR (051+052)
WT
Deformed
Transformed
16.0%
8.0%
76.0%
n=25
G'
phenotypes of mutants (n=21)
H
Pcl CRISPR (051)
77.1%
4.3%
18.6%
n=70
H'
phenotypes of mutants (n=16)
I
Pcl CRISPR (052)
38.0%
2.5%
59.5%
n=79
I'
phenotypes of mutants (n=49)
An1/2 to T4/5
Deformed An
T1 to T4/5
Deformed Mp
Missing T2 CP
CP Reductions
T6/7/8 Basis Reduction
T2/3 to T4/5
T5 TP Not Fused
Twisted Legs
Proximal-Distal Defect

Supp. Fig. 2.2.3: *Pcl* CRISPR phenotype data

A) Wildtype *Parhyale* hatchling head and anterior thorax.
A') *Pcl* CRISPR hatchling head and anterior thorax showing various homeotic transformations and missing T2 tergal and coxal plates.
B) Wildtype T1 appendages.
B') Wildtype T2/3 appendage.
B'') Wildtype T4/5 appendage.
C) Wildtype *Parhyale* hatchling view focused on tergal and coxal plates.
C') *Pcl* CRISPR hatchling showing reduction in basal plates (arrows) and exposed gills (asterisks) revealed by reduction in coxal plates.
C'') *Pcl* CRISPR hatchling showing tergal plate fusion defect.
D) *Pcl* CRISPR hatchling showing unilateral transformations from a ventral view. Mx2 - T3 appendages are all transformed towards T4/5 forward walking leg identity.
E) *Pcl* CRISPR hatchling showing twisted legs.
F) Average survivorship curve of *Pcl* dual-guide knockout experiments.
G) Proportion of WT, deformed, and transformed hatchlings for *Pcl* dual-guide knockout (guide RNAs 051 and 052) from n=25 hatchlings.
G') Proportion of phenotypes from *Pcl* (051+052) dual-guide knockout for n=21 hatchlings that we did not classify as wildtype. Yellow labels indicate phenotypes we classified as homeotic mutations.
H) Proportion of WT, deformed, and transformed hatchlings for *Pcl* (051) single-guide knockout for n=70 hatchlings.
H') Proportion of phenotypes from *Pcl* (051) single-guide knockout for n=16 mutant hatchlings.
I) Proportion of WT, deformed, and transformed hatchlings for *Pcl* (052) single-guide knockout for n=79 hatchlings.
I') Proportion of phenotypes from *Pcl* (052) single-guide knockout for n=49 mutant hatchlings.



Supp. Fig. 2.2.4: *Su(Z)12* CRISPR phenotype data

A) Wildtype *Parhyale* hatchling, lateral view.

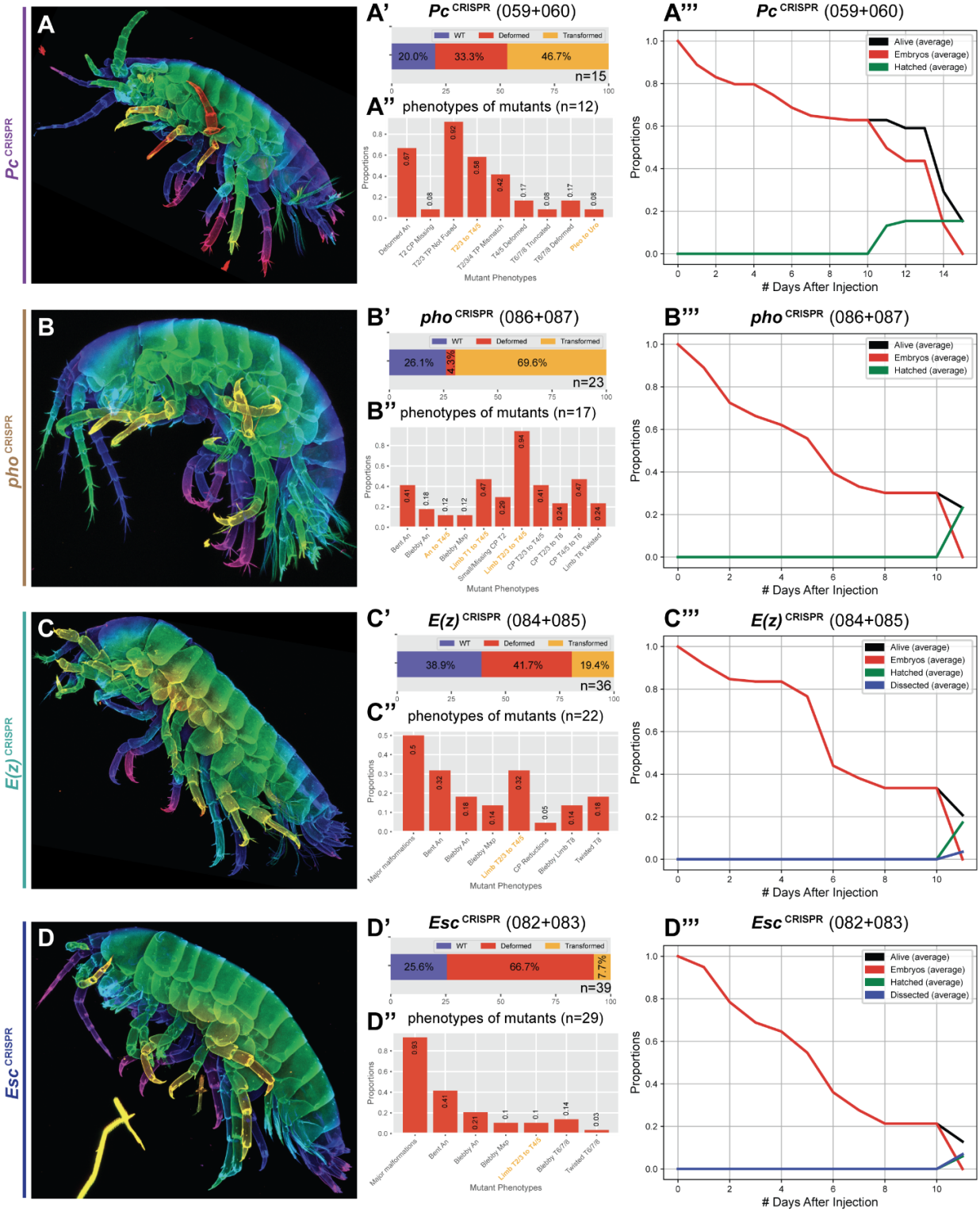
A') *Su(Z)12* CRISPR hatchling showing various homeotic transformations and limb defects.

B) Wildtype *Parhyale* hatchling, ventral view.

B') *Su(Z)12* CRISPR hatchling showing various unilateral homeotic transformations and limb defects, ventral view.

C) *Su(Z)12* CRISPR hatchling showing tergal plate fusion defect.

D) *Su(Z)12* CRISPR hatchling head showing antenna to T4/5 partial transformation. The hallmark of this transformation is the emergence of a coxal plate on the antennal segment, not normally observed.
D') Wildtype *Parhyale* hatchling head. Note the absence of the plate seen in the *Su(Z)12* knockout.
E) Average survivorship curve of *Su(Z)12* quadruple-guide knockout experiments.
F) Proportion of WT, deformed, and transformed hatchlings for *Pcl* quadruple-guide knockout (guide RNAs 047-050) from n=19 hatchlings. We also performed single-guide knockout experiments for guides 047 and 048, which we designed as a result of ambiguity about the true START codon for the *Su(Z)12* gene. These guides did not appear to show any obvious defects or homeotic transformations.
F') Proportion of phenotypes from *Su(Z)12* (047-050) quadruple-guide knockout for n=18 hatchlings that we did not classify as wildtype. Yellow labels indicate phenotypes we classified as homeotic mutations.
G) Proportion of WT, deformed, and transformed hatchlings for *Su(Z)12* (049) single-guide knockout for n=95 hatchlings.
G') Proportion of phenotypes from *Su(Z)12* (049) single-guide knockout for n=52 mutant hatchlings.
H) Proportion of WT, deformed, and transformed hatchlings for *Su(Z)12* (050) single-guide knockout for n=42 hatchlings.
H') Proportion of phenotypes from *Su(Z)12* (050) single-guide knockout for n=32 mutant hatchlings.


Pc CRISPR
A
A'
Pc CRISPR (059+060)
WT
Deformed
Transformed
20.0%
33.3%
46.7%
n=15
A''
phenotypes of mutants (n=12)
Proportions
Mutant Phenotypes
Deformed An
T2 CP Missing
T2/3 TP Not Fused
T2/3 to T4/5
T2/3/4 TP Mismatch
T4/5 Deformed
T6/7/8 Truncated
T6/7/8 Deformed
Pleo to Uro
A'''
Pc CRISPR (059+060)
Alive (average)
Embryos (average)
Hatched (average)
Proportions
Days After Injection
pho CRISPR
B
B'
pho CRISPR (086+087)
26.1%
4.3%
69.6%
n=23
B''
phenotypes of mutants (n=17)
Bent An
Blebby An
An to T4/5
Blebby Mxp
Limb T1 to T4/5
Small/Missing CP T2
Limb T2/3 to T4/5
CP T2/3 to T4/5
CP T2/3 to T6
CP T4/5 to T6
Limb T6 Twisted
B'''
pho CRISPR (086+087)
E(z) CRISPR
C
C'
E(z) CRISPR (084+085)
38.9%
41.7%
19.4%
n=36
C''
phenotypes of mutants (n=22)
Major malformations
Bent An
Blebby An
Blebby Mxp
Limb T2/3 to T4/5
CP Reductions
Blebby Limb T8
Twisted T8
C'''
E(z) CRISPR (084+085)
Dissected (average)
Esc CRISPR
D
D'
Esc CRISPR (082+083)
25.6%
66.7%
7.7%
n=39
D''
phenotypes of mutants (n=29)
Blebby T6/7/8
Twisted T6/7/8
D'''
Esc CRISPR (082+083)

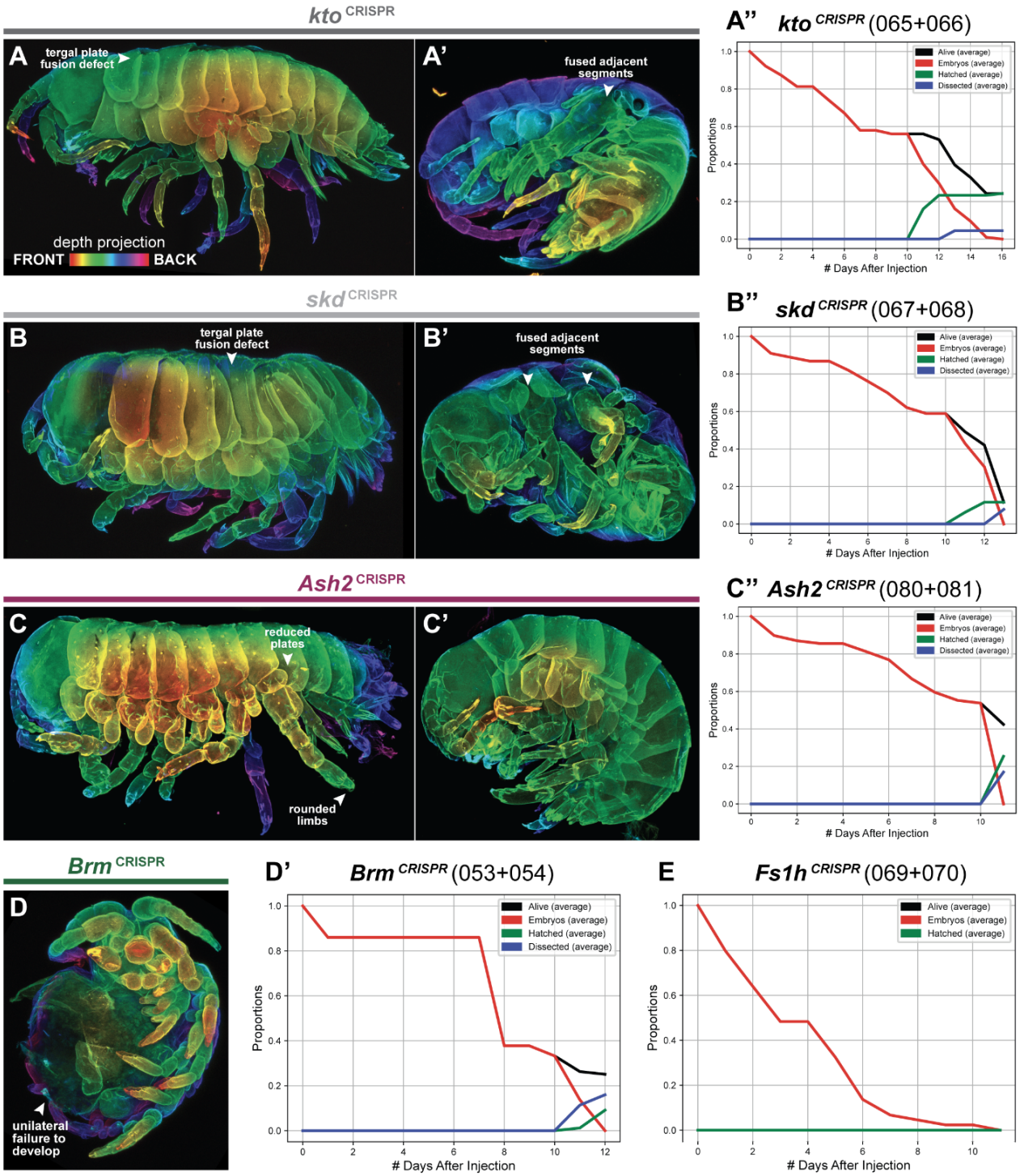
Supp. Fig. 2.2.5: Phenotype data for additional PcG genes

A-A’’’) Representative hatchling showing T2/3 to T4/5 transformations, mutant classes, phenotype proportions, and survivorship curve for *Pc* CRISPR knockout.

B-B’’’) Representative hatchling showing T2/3 to T4/5 transformations, mutant classes, phenotype proportions, and survivorship curve for *pho* CRISPR knockout.

C-C’’’) Representative hatchling showing T2/3 to T4/5 transformations, mutant classes, phenotype proportions, and survivorship curve for *E(z)* CRISPR knockout.

D-D’’’) Representative hatchling showing T2/3 to T4/5 transformations, mutant classes, phenotype proportions, and survivorship curve for *Esc* CRISPR knockout.



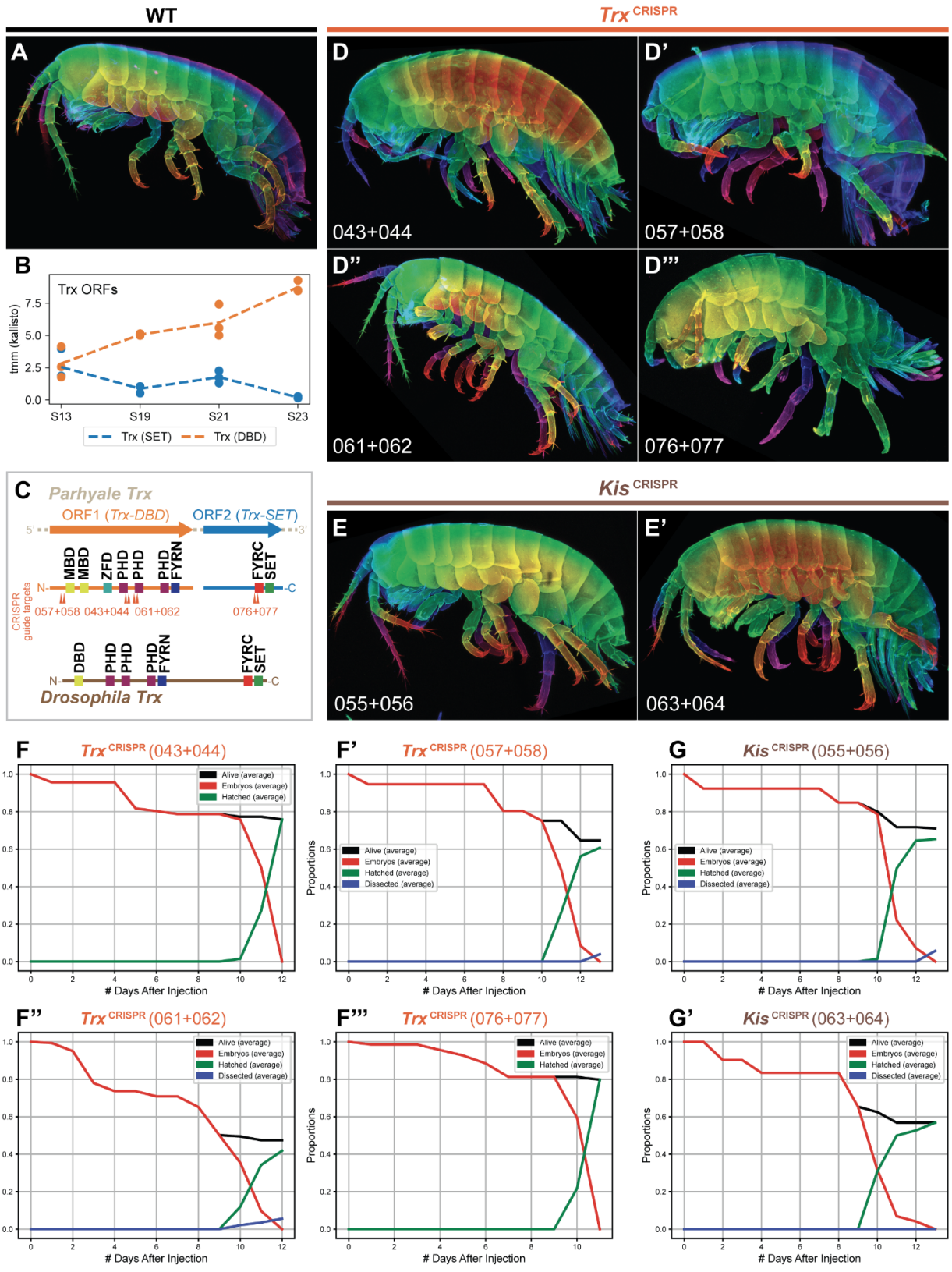
Supp. Fig. 2.2.6: Example phenotypes for TrxG CRISPR knockouts

A-A") Representative hatchlings and survivorship curve of *kto* CRISPR experiments. A) shows a milder *kto* hatchling, while A') shows a more severe phenotype, with fusion between adjacent body segments marked by arrowhead.

B-B") Representative hatchlings and survivorship curve of *skd* CRISPR experiments. B) shows a milder *skd* hatchling, while B') shows a more severe phenotype, with fusion between adjacent body segments marked by arrowhead.

C-C") Representative hatchlings and survivorship curve of *Ash2* CRISPR experiments. C) shows an *Ash2* hatchling with rounded legs and plates, while C') shows a hatchling with minor morphological defects.

D-D') Representative hatchlings and survivorship curve of *Brm* CRISPR experiments. D) shows an example of a *Brm* mutant hatchling where one half of the embryo appears not to have developed.
E) Survivorship curve of *Fs1h* CRISPR experiments. We recovered no surviving embryos from *Fs1h* knockout experiments, suggesting that *Fs1h* knockout is embryonic lethal.


WT
Trx CRISPR
A
B
Trx ORFs
tmm (kallisto)
7.5
5.0
2.5
0.0
S13
S19
S21
S23
Trx (SET)
Trx (DBD)
C
Parhyale Trx
ORF1 (Trx-DBD)
ORF2 (Trx-SET)
5'
3'
MBD
MBD
ZFD
PHD
PHD
PHD
FYRN
FYRC
SET
N-
-C
CRISPR guide targets
057+058
043+044
061+062
076+077
DBD
PHD
PHD
PHD
FYRN
FYRC
SET
N-
-C
Drosophila Trx
D
043+044
D'
057+058
D''
061+062
D'''
076+077
Kis CRISPR
E
055+056
E'
063+064
F
Trx CRISPR (043+044)
F'
Trx CRISPR (057+058)
G
Kis CRISPR (055+056)
F''
Trx CRISPR (061+062)
F'''
Trx CRISPR (076+077)
G'
Kis CRISPR (063+064)
Alive (average)
Embryos (average)
Hatched (average)
Dissected (average)
Proportions
Days After Injection

Supp. Fig. 2.2.7: Lack of obvious phenotypes for *Trx* and *Kis* knockout experiments

A) Wildtype *Parhyale* hatchling, lateral view.

B) Expression of two candidate *Parhyale Trx* ORFs, one containing DNA binding domains (Trx-DBD) and one containing the SET domain (Trx-SET). The two candidate ORFs appear to be expressed at different levels throughout development.

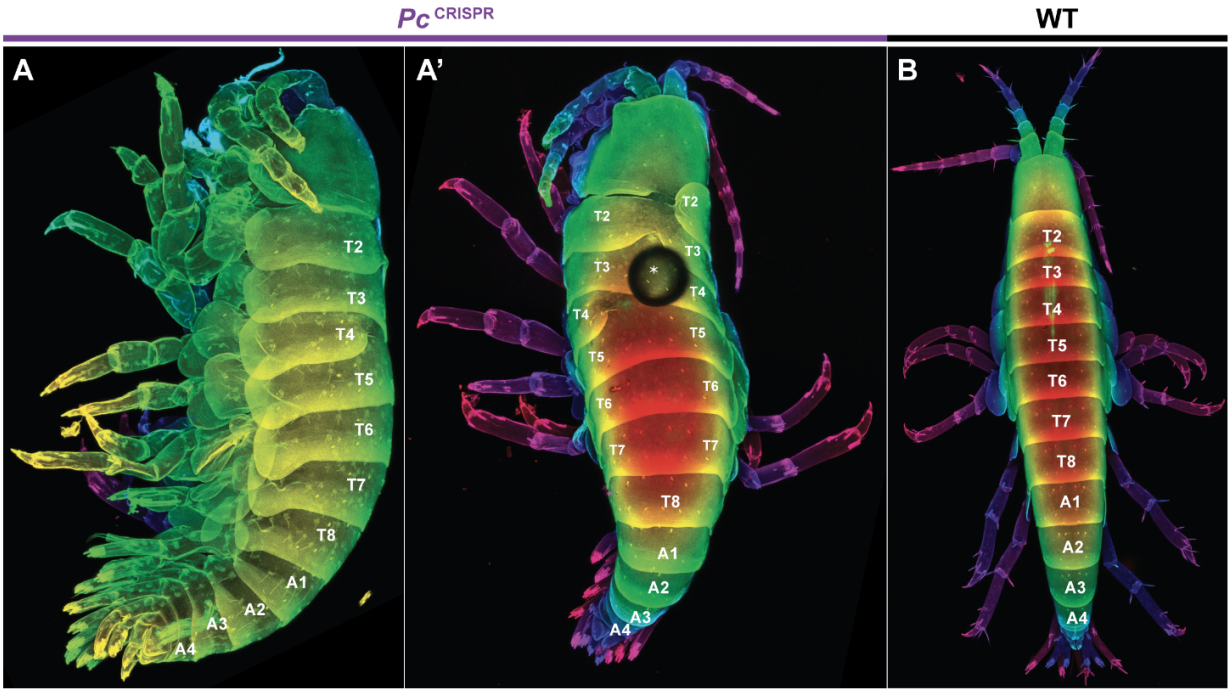
C) Representation of relative position of protein domains comparing *Parhyale Trx-DBD* and *Trx-SET* to *Drosophila Trx*. *Parhyale* may have a split in the two peptide domains. Relative position of CRISPR guide RNAs is illustrated for all 8 guides used to target *Trx*.

D-D''') Representative hatchlings from *Trx* CRISPR knockout experiments. Each experiment used a different pair of guide RNAs, shown in the bottom left corner. We did not observe obvious homeotic transformations for any of the guides.

E-E') Representative hatchlings from *Kis* CRISPR knockout experiments. Each experiment used a different pair of guide RNAs, shown in the bottom left corner. We did not observe obvious homeotic transformations for any of the guides.

F-F''') Survivorship curves for *Trx* CRISPR knockout experiments. We observed varying levels of survivorship for these experiments; however, during the course of the experiments, we did not observe hatchlings that showed obvious morphological defects, as was observed for PcG gene knockout experiments, or other TrxG knockouts, such as *Brm* or *Fs1h*.

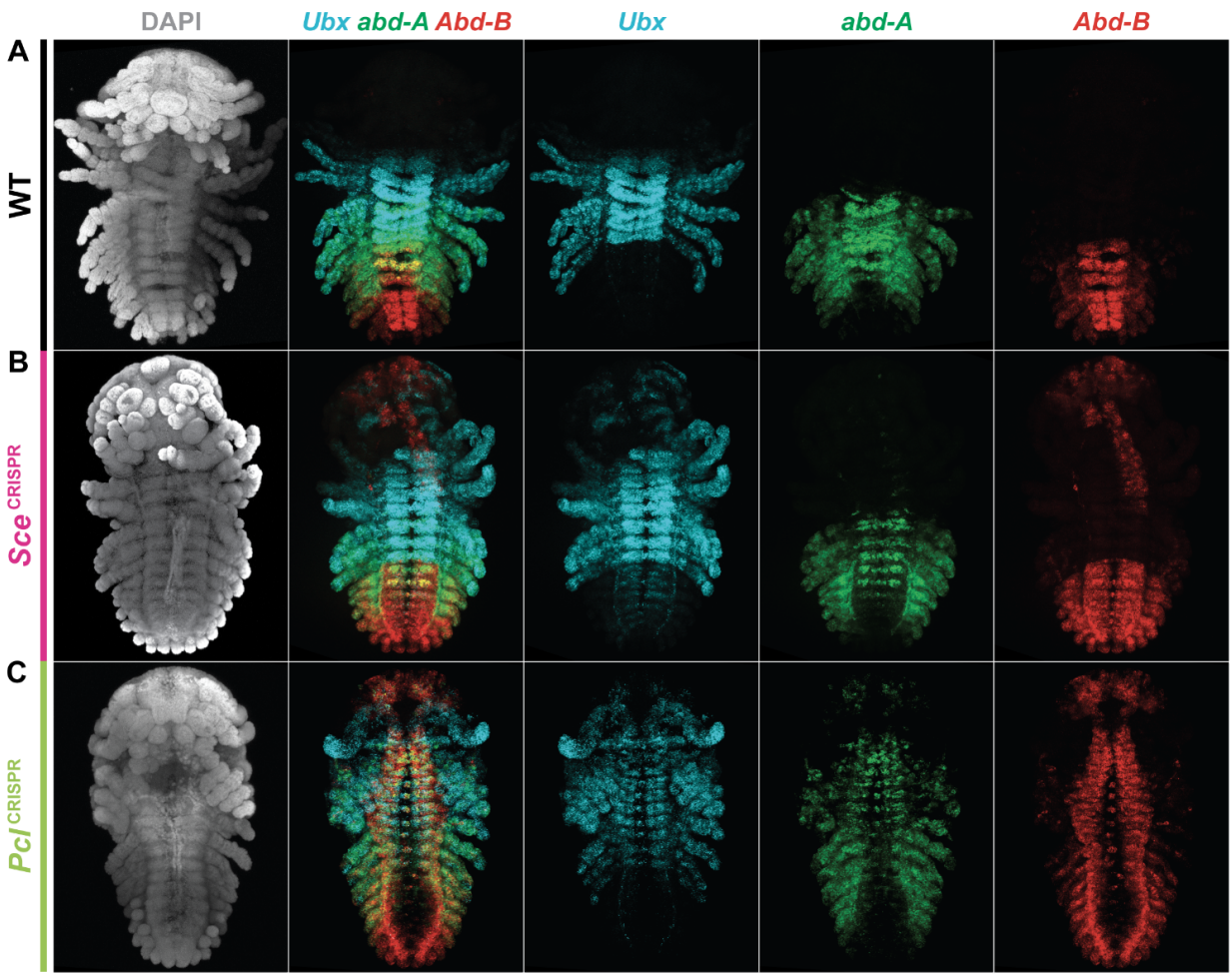
G-G') Survivorship curves for *Kis* CRISPR knockout experiments. We observed a similar lack of obvious defects in *Kis* knockout embryos during development.



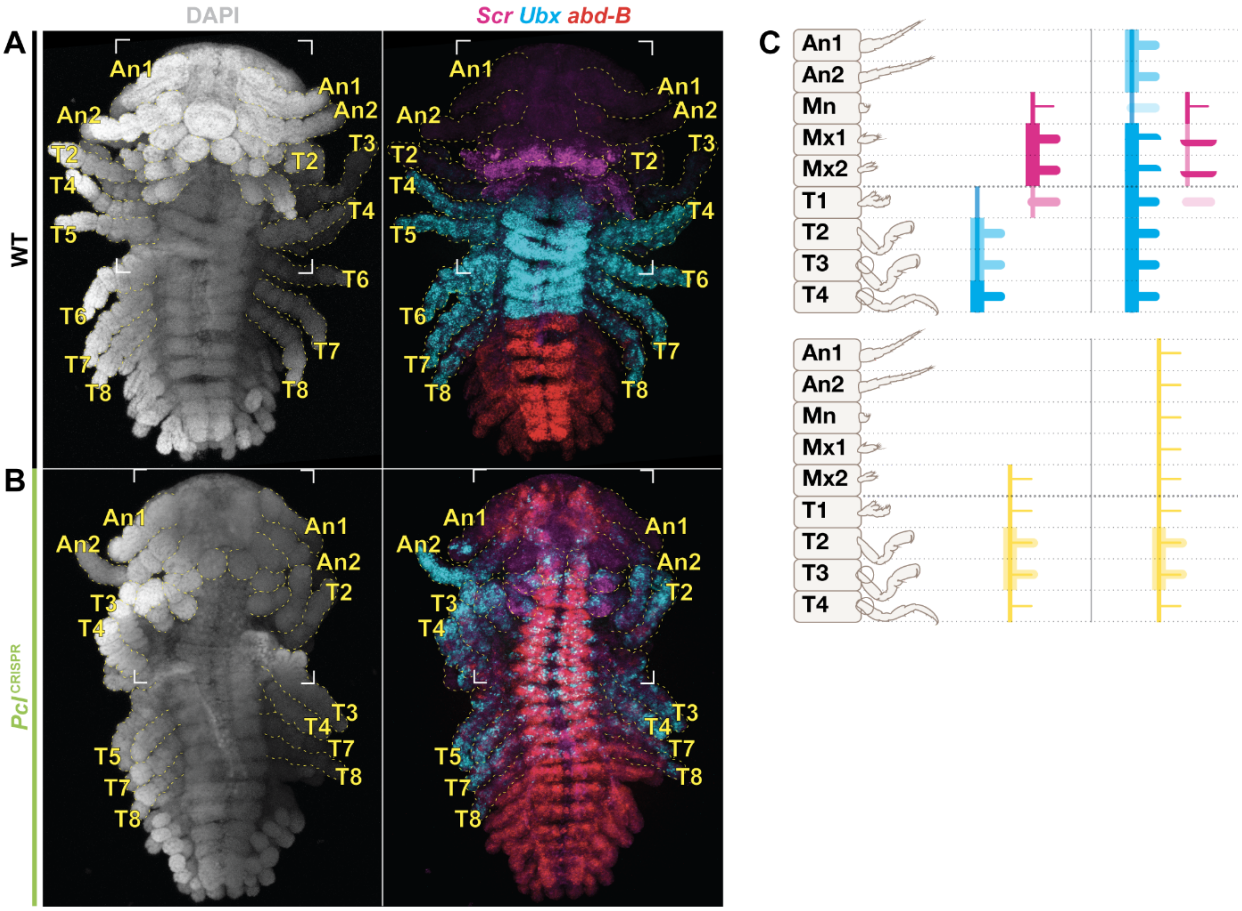
Supp. Fig. 2.3.1: Asymmetric fusion of plates in *Pc* mutant hatchlings

A-A') Lateral and dorsal views of the same *Pc* mutant hatchling, showing asymmetric fusion of plates. The same image is displayed as in Fig. 3; however, we display the entire hatchling here in A' to illustrate how we counted the number of tergal plates.

B) Dorsal view of a WT hatchling showing the counting of tergal plates.



Supp. Fig. 2.4.1: Expression of posterior *Hox* genes in PcG mutant embryos
This figure is identical to Fig. 4, with the exception that we have removed limb labels.

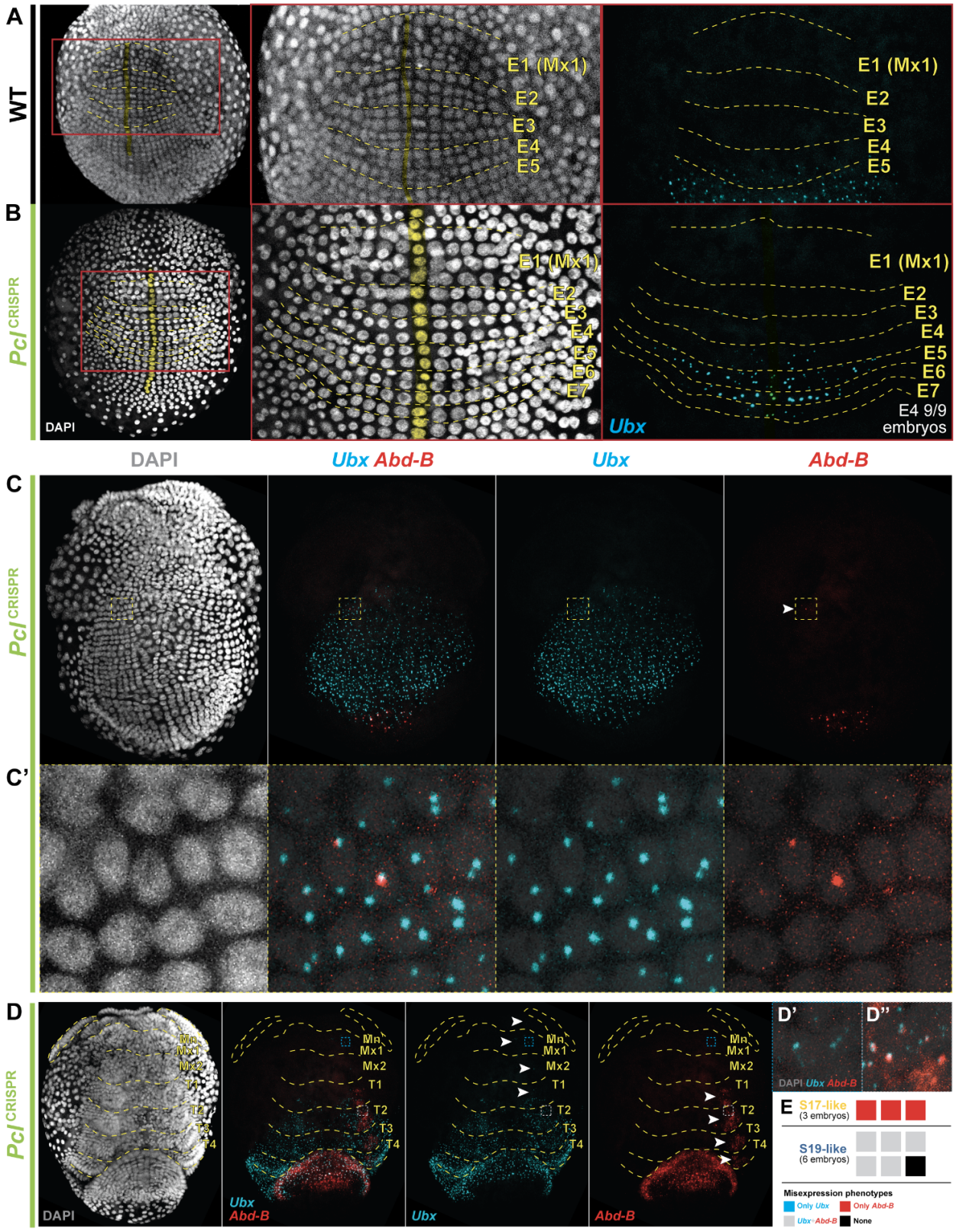


Supp. Fig. 2.5.1: Misexpression of *Scr*, *Ubx*, and *Abd-B* in *Pcl* mutant embryos

A) Wildtype *Parhyale* embryo at S23 showing expression domains of *Scr* (magenta), *Ubx* (cyan), and *Abd-B* (red).

B) *Pcl* CRISPR embryo at S23 showing expression domains of *Scr* (magenta), *Ubx* (cyan), and *Abd-B* (red). The embryo displayed appears to be bilaterally mutant throughout the body axis, as evidenced by strong *Abd-B* expression throughout the head and thoracic nervous system.

C) Schematic of patterns of *Sce*, *Ubx*, and *Antp* expression in wildtype versus PcG mutant embryos.


A
WT
E1 (Mx1)
E2
E3
E4
E5
B
Pcl CRISPR
DAPI
E6
E7
Ubx
E4 9/9
embryos
DAPI
Ubx Abd-B
Ubx
Abd-B
C
C'
D
Ma
Mx1
Mx2
T1
T2
T3
T4
D'
D''
E
S17-like
(3 embryos)
S19-like
(6 embryos)
Misexpression phenotypes
Only Ubx
Only Abd-B
Ubx+Abd-B
None

Supp. Fig. 2.6.1: Differences in PcG dependence timing in *Pcl* mutant embryos

A) Wildtype *Parhyale* embryo at S13 showing expression of *Ubx* (cyan) and DAPI. The anterior boundary of *Ubx* expression is the E4 parasegment.

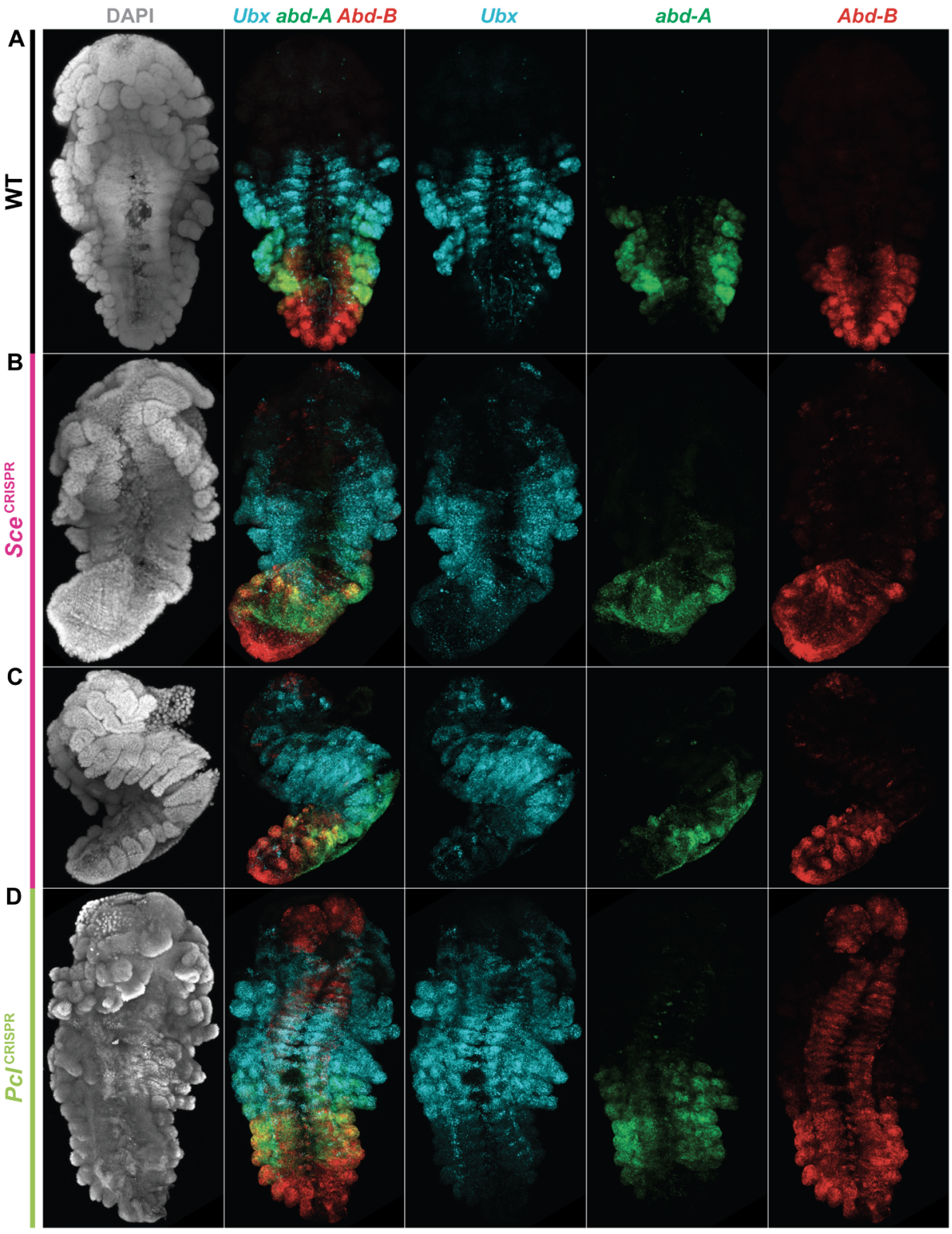
B) *Pcl* CRISPR embryo at S13 showing expression of *Ubx* (cyan). The anterior boundary of *Ubx* expression is the E4 parasegment. We observed the same anterior boundary in 9/9 embryos imaged at this stage.

C) Expression of *Ubx* and *Abd-B* in S17-like *Pcl* mutant embryos. At the same time as *Abd-B* expression is beginning in the posterior, we observed ectopic *Abd-B* expression initiation in the anterior.

C') Inset panel from C) showing a handful of cells with strong *Abd-B* transcriptional spots, and *Ubx* expression in the same cells.

D) Expression of *Ubx* and *Abd-B* in S19-like *Pcl* mutant embryos. We observed misexpression of both *Ubx* and *Abd-B* at this stage, as marked with arrowheads in gene-specific panels; all embryos that showed *Abd-B* misexpression also showed *Ubx* misexpression.

E) Number of embryos showing different patterns of misexpression in S17-like and S19-like embryos dissected concurrently when wildtype embryos were at stage S19. We observed that 3/3 S17-like embryos showed only misexpression of *Abd-B*, while 5/6 S19-like embryos showed simultaneous misexpression of *Ubx* and *Abd-B*.


DAPI
Ubx abd-A Abd-B
Ubx
abd-A
Abd-B
A
WT
B
Sce CRISPR
C
D
Pcl CRISPR

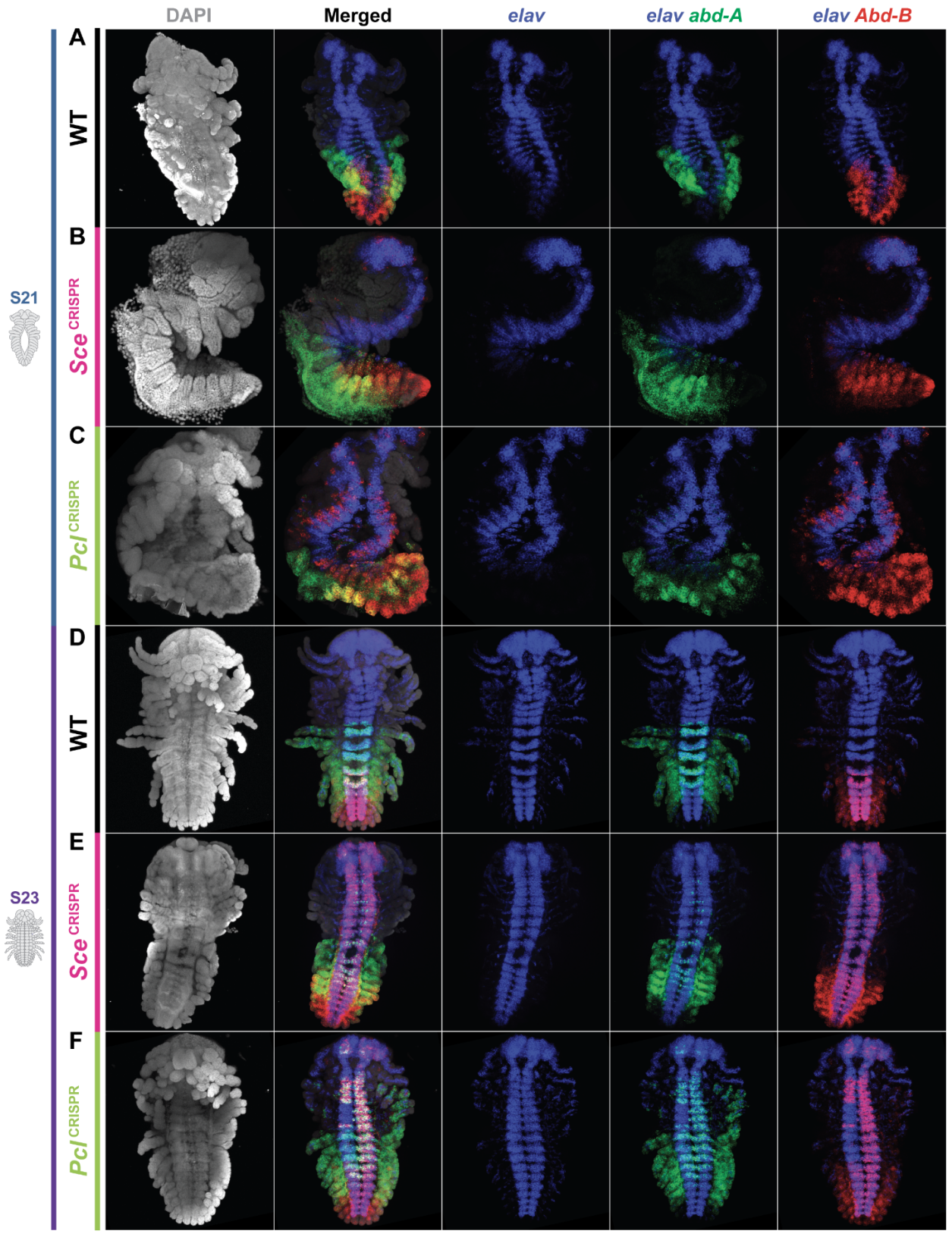
Supp. Fig. 2.6.2: Expression of posterior *Hox* genes in S21 embryos

A) Wildtype *Parhyale* embryo at S21 showing expression of *Ubx* (cyan), *abd-A* (green), *Abd-B* (red), and DAPI (gray).

B) *Sce* CRISPR embryo at S21. Weak *Abd-B* and *Ubx* misexpression anterior of the wildtype domain is observed. Very little *abd-A* misexpression is observed.

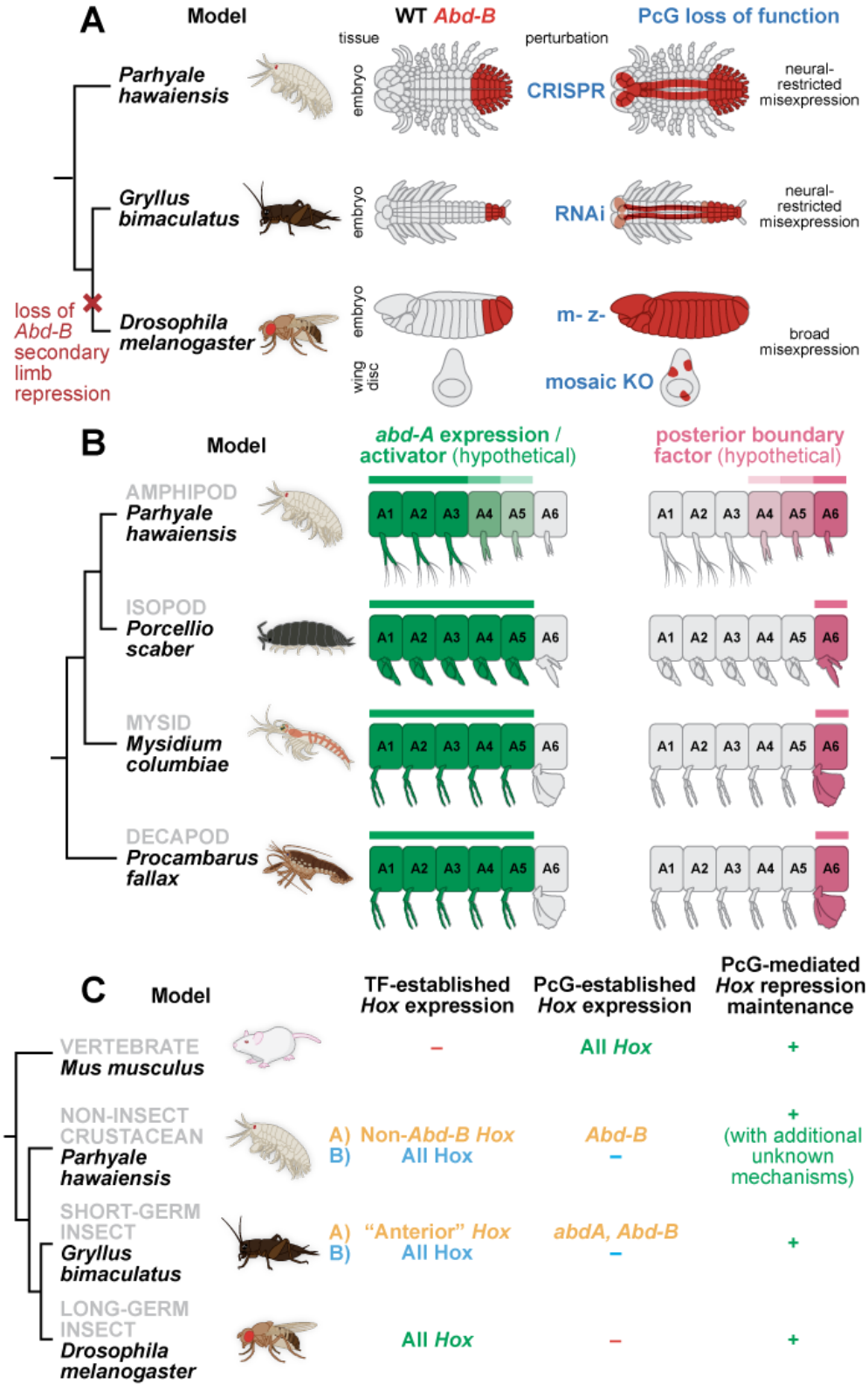
C) *Sce* CRISPR embryo at S21. This embryo is mounted as a lateral view, and is slightly more morphologically developed than in B. *Ubx* misexpression is much stronger by this point, and small patches of *abd-A* misexpression are beginning.

D) *Pcl* CRISPR embryo at S21. Strong *Ubx* and *Abd-B* misexpression are observed, and the beginning of *abd-A* misexpression is observed.


DAPI
Merged
elav
elav abd-A
elav Abd-B
S21
S23
A
WT
B
Sce CRISPR
C
Pcl CRISPR
D
WT
E
Sce CRISPR
F
Pcl CRISPR

Supp. Fig. 2.7.1: abd-A, Abd-B, and elav expression

A) Wildtype *Parhyale* embryo at S21 showing expression of *elav* (blue), *abd-A* (green), *Abd-B* (red), and DAPI (gray).
B) *Sce* CRISPR embryo at S21. Patches of *Abd-B* misexpression are observed at the lateral borders of *elav* expression.
C) *Pcl* CRISPR embryo at S21. Stronger *Abd-B* misexpression is observed than compared to an *Sce* mutant.
D) Wildtype *Parhyale* embryo at S23.
E) *Sce* CRISPR embryo at S23. Strong *Abd-B* misexpression is observed throughout the nervous system. *abd-A* misexpression is restricted to a few cells of the nervous system.
F) *Pcl* CRISPR embryo at S23. Strong *Abd-B* misexpression is observed throughout the nervous system. *abd-A* misexpression expands broadly through the nervous system and into the limbs.


A
Model
WT Abd-B
PcG loss of function
tissue
perturbation
Parhyale hawaiensis
embryo
CRISPR
neural-restricted misexpression
Gryllus bimaculatus
embryo
RNAi
neural-restricted misexpression
loss of Abd-B secondary limb repression
Drosophila melanogaster
embryo
m- z-
broad misexpression
wing disc
mosaic KO
B
Model
abd-A expression / activator (hypothetical)
posterior boundary factor (hypothetical)
AMPHIPOD
Parhyale hawaiensis
ISOPOD
Porcellio scaber
MYSID
Mysidium columbiae
DECAPOD
Procambarus fallax
A1
A2
A3
A4
A5
A6
C
Model
TF-established Hox expression
PcG-established Hox expression
PcG-mediated Hox repression maintenance
VERTEBRATE
Mus musculus
–
All Hox
+
NON-INSECT CRUSTACEAN
Parhyale hawaiensis
A) Non-Abd-B Hox
B) All Hox
Abd-B
–
+
(with additional unknown mechanisms)
SHORT-GERM INSECT
Gryllus bimaculatus
A) "Anterior" Hox
B) All Hox
abdA, Abd-B
–
+
LONG-GERM INSECT
Drosophila melanogaster
All Hox
–
+

Supp. Fig. 2.8.1: Hypothesized *Hox* regulatory mechanisms

A) An unknown limb-specific repression mechanism for *Abd-B* in arthropods. In both *Parhyale* CRISPR knockout of PcG genes and *Tribolium* RNAi, *Abd-B* misexpression extends into the nervous system, but not the limbs. This is in contrast to *Drosophila*, where *Abd-B* misexpression is possible across broad tissue types, including imaginal tissues such as wing discs. This suggests that a potentially ancestral limb-specific repression mechanism of *Abd-B* may have been lost in the lineage leading to flies. Future experiments using PcG knockout in more diverse organisms could reveal whether such a repression mechanism is commonly found across non-dipteran pancrustaceans.

B) A posterior boundary factor for *abd-A*. CRISPR-Cas9 knockout of *Ubx* and *Abd-B*, the *Hox* genes overlapping *abd-A*, do not result in disruption of the pattern of *abd-A* expression (Jarvis et al. 2022). This suggests that *abd-A* is not regulated by the other posterior *Hox* genes. In our study, PcG knockout resulted in misexpression of *abd-A* in the anterior, but no disruption to the pattern of *abd-A* expression in the posterior. This suggests that neither *Hox* cross-regulatory interactions or PcG repression sets the posterior boundary of *abd-A*. We propose a third mechanism, a posterior boundary factor, which represses *abd-A* at the posterior end of the abdomen. We hypothesize that such a factor may also be responsible for the shifts in the *abd-A* expression boundary between different groups of crustaceans.

C) Summary of *Hox* regulatory mechanisms at different phases of development across bilaterians. In *Mus*, *Hox* gene expression establishment is linked to PcG function, whereas in *Drosophila*, PcG repression appears to occur after an early transcription factor-dependent phase of regulation. In examining the *Parhyale* and *Gryllus* data, two possible interpretations emerge: either both *Parhyale* and *Gryllus* have TF-dependent regulation for some *Hox* genes and not others, or both *Parhyale* and *Gryllus* have concurrent TF-dependent and PcG-dependent regulatory mechanisms, which are distinguishable only at very early stages of *Hox* misexpression in PcG mutants. If the latter interpretation is true, this would suggest that arthropods and vertebrates have evolved fundamentally different mechanisms of early *Hox* boundary establishment. Additional experiments with representatives across bilaterians will be needed to determine whether the ancestral bilaterian used PcG- or TF-dependent *Hox* boundary establishment.

Supp. Table. 2.1: Select *Drosophila* PcG/TrxG homeotic phenotypes

Family	*Drosophila* Protein	Genotype	Phenotype(s)	Source(s)
PcG	Ph	*Ph-d*$^{m-z-}$ *Ph-p*$^{m-z-}$	A1-7 segments resemble A8 segment. Thoracic segments have denticles characteristic of abdominal segments.	McKeon et al., 1994; Dura et al., 1987; Doe et al., 1988
	Psc	*Psc*$^{m-z-}$	Anterior segments transformed into A8 segment. *Hox* misexpression. Other early developmental defects.	Martin and Adler, 1993; Morimoto et al., 2016
	Sce	*Sce*$^{m-z-}$	Anterior→posterior transformations. No segmentation defects. *Ubx* misexpression.	Breen and Duncan, 1986; Fritsch et al., 2003
	Pc	*Pc*$^{m+z-}$ or *Pc*$^{m-z-}$	Transformation of all thoracic and abdominal segments to A8 segment identity. *En* misexpression.	Chiang et al., 1995; Kassis and Brown, 2013; Haynie, 1983; Lawrence et al., 1983
	Scm	*Scm*$^{m-z-}$	Anterior→posterior homeotic transformation of most segments. No segmentation defects.	Breen and Duncan, 1986
	Su(z)12	*Su(z)12*$^{m+z-}$	All abdominal, thoracic, and some head segments are transformed into the A8 segment. *Abd-B* misexpression in every segment. Antennae partially transformed into legs. Smaller wings.	Birve et al., 2001
	E(z)	*E(z)*$^{m+z-}$	All thoracic and abdominal segments transformed to more posterior segments.	Phillips and Shearn, 1990
	Pcl	*Pcl*$^{m-z-}$	A7 segment partially transformed into A8 segment. Dorsal setal belt in head. Wart-like sense organs (normally in A8) in the anterior abdominal segments. Defects in even numbered segments.	Breen and Duncan, 1986
	Esc	*Esc*$^{m-z-}$	Transformation of all abdominal segments to A8 segment identity.	Wang et al., 2006; Kurzhals et al., 2008; Ohno et al., 2008
TrxG	Trx	*Trx*$^{m-z-}$	Transformation of T1-3 segments into T2 segment identity. Posterior→anterior transformations of abdominal segments. Reduced *Ubx* and *abd-A* protein levels. Haltere to wing transformation (due to decreased *Ubx* expression). Posterior abdominal and genital structures transformed to more anterior identities (due to decreased *abd-A* and *Abd-B* expression).	Gildea et al., 2000; Breen and Harte, 1991; Ingham, 1983; Klymenko and Müller, 2004; Tripoulas et al., 1994; Mazo et al., 1990; Ingham, 1981; Ingham and Whittle, 1980
	Ash2	*Ash2*$^{m-z-}$	Various posterior→anterior homeotic transformations including: haltere to wing transformation, leg 3 to leg 2 transformation, and genitalia to leg/antenna transformation.	Shearn, 1989
	Ash1	*Ash1*$^{m-z-}$	Various posterior→anterior homeotic transformations including: posterior wing to anterior wing transformation, haltere to wing transformation, leg 3 to leg 2 transformation, and genitalia to leg/antenna transformation. Loss of *En* function. Reduction of *Hox* expression.	Shearn, 1989; Tripoulas et al., 1994; Klymenko and Müller, 2004
	Kis	*Kis*$^{m+z-}$ clones	Homozygous clone of *Kis*$^{m+z-}$ tissue in A5 segment is transformed to a more anterior identity (loss of *Abd-B* function). Some clones showed transformations of first leg towards a second leg identity (loss of *Scr* function).	Daubresse et al., 1999
		Kis^{m-z-}	Pair-rule segmentation defects.	Daubresse et al., 1999
	Mor	*Mor*$^{z-}$	Embryonic lethal. Head defects that resemble defects seen in *Dfd*$^{z-}$ flies.	Harding et al., 1995
		Mor^{z-} clones	Clones of *Mor*$^{z-}$ cells caused transformations of metanotum and haltere to mesonotum and anterior wing, respectively. Posterior wing to anterior wing transformation (loss of *En* function).	Brizuela and Kennison, 1997
	Brm	*Brm hypomorph*	Reductions in the number of sex comb teeth. Transformations of A5 segment to a more anterior identity. Abnormalities similar to those seen in ANTP-C and BX-C loss-of-function mutants.	Tamkun et al., 1992
		Brm^{m-}	Early embryonic defects.	Brizuela et al., 1994
	Fs1h	*Fs1h*$^{m-z-}$ *temp. sen.*	Homeotic defects: anterior metanotum to anterior mesonotum transformation, and anterior haltere to anterior wing transformation. Mimics loss of *Ubx* function. Mutants often missing legs, halteres, or tergites.	Zalokar et al., 1975

Supp. Table. 2.2: Select *Mus* PcG/TrxG homeotic phenotypes

Family	*Drosophila* Protein	Genotype	Phenotype(s)	Source(s)
PcG	*Ph*	*PHC1-/-*	Perinatal lethality. Anterior→posterior skeletal homeotic transformations.	Isono et al., 2005; Takihara et al., 1997
	Psc	*PCGF2-/-*	Anterior→posterior transformation of axial skeleton. Growth retardation.	Akasaka et al., 2001
		PCGF4-/-	Anterior→posterior transformation of axial skeleton.	Akasaka et al., 2001
		PCGF2-/- *PCGF4-/-*	Embryonic lethal.	Akasaka et al., 2001
	Sce	*RING1A+/-* or *-/-*	Anterior→posterior transformations of vertebrae. Rib abnormalities.	del Mar Lorente et al., 2000
		RING1B-/-	Anterior→posterior homeotic transformations of the axial skeleton.	Suzuki et al., 2002
	Pc	*CDX2-/-*	Malformations of axial skeleton. Reduced viability. Poor growth. Male to female sex reversal. T7 vertebra transformed into T8 vertebra.	Coré et al., 1997
		CDX4-/-	Complete neonatal lethality. Severe thymus hypoplasia.	Coré et al., 1997
		CDX7-/-	Increased body length. Developed liver and lung adenomas and carcinomas.	Forzati et al., 2012
	Scm	*Scmh1*$^{SPM/-}$	Anterior→posterior vertebral transformations. Male infertility.	Takada et al., 2007
		Scmh1-/-	Abnormal hematopoiesis. Normal fertility and skeleton.	Yasunaga et al., 2013
	Su(z)12	*Su(z)12-/-*	Embryonic lethal.	Li et al., 2011
		Su(z)12+/-	Anterior tuberculums (characteristic of C6 vertebra) associated with C5 vertebra. Anteriorly shifted sternum. Spinous process (characteristic of T2 vertebra) associated with T1 vertebra.	Li et al., 2011
	E(z)	*EZH2 -/-*	Embryonic lethal.	O'Carroll et al., 2001
		EZH2 cond. loss	Severe growth retardation of neonate. Gastrulation failure.	O'Carroll et al., 2001
	Pcl	*Pcl2 -/-*	Anterior→posterior transformation of axial skeleton.	Wang et al., 2007
	Esc	*EED -/-*	Early gastrulation defect.	Wang et al., 2002
		EED hypomorph	Anterior expression boundaries of several *Hox* genes were shifted rostrally by one segment.	Wang et al., 2002
TrxG	*Trx*	*Mll -/-*	*Hoxa7* expression is activated normally but is not maintained. Branchial arch dysplasia. Aberrant segmental boundaries of spinal ganglia and somites.	Yu et al., 1998
	Ash2	*Ash2l -/-*	Die early during gestation.	Stoller et al., 2010
	Ash1	*Ash1l -/- (gene trap)*	Corpus to caput transformation. Widening of the cauda. Transformations similar to posterior→anterior transformations that occur in *Hoxa10-/-* mutant males.	Brinkmeier et al., 2015
	Kis	*CHD7 -/-*	Embryonic lethal.	Sperry et al., 2014
	Brm	*Brm -/-*	Increased cell proliferation.	Reyes et al., 1998
	Fs1h	*BRD2 -/-*	Embryonic lethal.	Shang et al., 2009

Supp. Table. 3: *Parhyale* PcG genes

Complex	*Parhyale* Gene Name	*Drosophila* Protein	*Mus* Proteins	Best *Parhyale* Transcript	Txome	*Parhyale* IGV Address
PRC1	Par-haw_ph	ph-d [Q9W523], ph-p [P39769]	PHC1 [Q64028], PHC2 [Q9QWH1], PHC3 [Q8CHP6]	mikado.phaw_50.283863eG7.1	Sun-Mikado	phaw_50.283863e:707,281-945,757
	Par-haw_Psc-1	Psc [P35820]	PCGF1 [Q8R023]	mikado.phaw_50.283028bG288.1	Sun-Mikado	phaw_50.283028b:22654336-22754322
	Par-haw_Psc-2	Su(z)2 [P25172]	PCGF2 [P23798]	mikado.phaw_50.283028bG712.1	Sun-Mikado	phaw_50.283028b:21,822,602-21,858,763
			PCGF3 [Q8BTQ0], PCGF4 [P25916], PCGF5 [Q3UK78], PCGF6 [Q99NA9]			
	Par-haw_Sce	Sce [Q9VB08]	RING1A [O35730], RING1B [Q9CQJ4]	mikado.phaw_50.283875aG27.1	Sun-Mikado	phaw_50.283875a:2670126-2804666
	Par-haw_Pc	Pc [P26017]	CBX2 [P30658], CBX4 [O55187], CBX6 [Q9DBY5], CBX7 [Q8VDS3], CBX8 [Q9QXV1]	mikado.phaw_50.000214fG102.1	Sun-Mikado	phaw_50.000214f:7,059,547-7,067,561
dRAF	includes Par-haw_Psc-1*, Par-haw_Psc-2*, Par-haw_Sce* (see above)					
	Par-haw_Kdm2-isoform_a	Kdm2 [Q9VHH9]	Kdm2a [P59997]	STRG.39002.1	Sun-StringTie2	-
	Par-haw_Kdm2-isoform_b			mikado.phaw_50.283869cG487.1	Sun-Mikado	phaw_50.283869c:39,738,428-39,800,237
PR-DUB	Par-haw_Calypso	Calypso [Q7K5N4]	Bap1 [Q99PU7]	comp166625_c1_seq3	Trinity-Limb	phaw_50.000214e:2,663,327-2,803,232
	Par-haw_Asx	Asx [Q9V727]	ASXL1 [P59598], ASXL2 [Q8BZ32], ASXL3 [Q8C4A5]	mikado.phaw_50.283821G75.2	Sun-Mikado	phaw_50.283821:4,766,576-4,819,515
Scm	Par-haw_Scm	Scm [Q9VHA0]	SCMH1 [Q8K214], SCML2 [B1AVB5], SCML4 [Q80VG1]	comp171618_c0	Trinity-Limb	phaw_50.007301a:14,931,316-14,976,242
PRC2.1	Par-haw_Su(z)12	Su(z)12 [Q9NJG9]	SUZ12 [Q80U70]	mikado.phaw_50.283867cG387.1	Sun-Mikado	phaw_50.283867c:24503623-24641721
	Par-haw_E(z)	E(z) [P42124]	EZH1 [P70351], EZH2 [Q61188]	mikado.phaw_50.001137bG316.1	Sun-Mikado	phaw_50.001137b:17254399-17416111
	Par-haw_Caf1	Caf1 [Q24572]	Rbbp4 [Q60972]	mikado.phaw_50.283826G445.1	Sun-Mikado	phaw_50.283826:29293007-29339138
	Par-haw_Pcl	Pcl [Q24459]	PCL2 [Q02395], PCL3 [Q9CXG9]	mikado.phaw_50.283817cG134.1	Kao-Mikado	phaw_50.283817c:3067340-3096584
	Par-haw_HDAC1	Rpd3 [Q94517]	Hdac1 [O09106]	comp160143_c0_seq1	Trinity-Limb	-
	Par-haw_Esc	Esc [Q24338]	EED [Q921E6]	mikado.phaw_50.283863aG40.1	Sun-Mikado	phaw_50.283863a:3,188,185-3,235,909
PRC2.2	includes Par-haw_Su(z)12*, Par-haw_E(z)*, Par-haw_Caf1*, Par-haw_Esc* (see above)					
	Par-haw_Jarid2	Jarid2 [Q9VT00]	Jarid2 [Q62315]	mikado.phaw_50.283865bG254.1	Kao-Mikado	-
	Par-haw_Jing	Jing [Q7KHG2]	Aebp2 [Q9Z248]	mikado.phaw_50.282260aG393.1	Sun-Mikado	-
PhoRC	Par-haw_pho	Pho [Q8ST83]	YY1 [Q00899]	mikado.phaw_50.000289aG44.1	Sun-Mikado	phaw_50.000289a:1971807-1992607
	Par-haw_Sfmbt	dSfmbt [Q9VK33]	SFMBT1 [Q9JMD1], SFMBT2 [Q5DTW2]	mikado.phaw_50.283867cG203.1	Sun-Mikado	phaw_50.283867c:14916319-14971921

Supp. Table. 2.4: *Parhyale* TrxG genes

Complex	*Parhyale* Protein Name	*Drosophila* Protein	*Mus* Protein	Best *Parhyale* Transcript	Txome	*Parhyale* IGV Address
dCOMPASS-like	Par-haw_Trx (SET)	Trx [P20659]	KMT2B [O08550]	mikado.phaw_50.282275aG400.1	Sun-Mikado	phaw_50.282275a:22,898,633-23,055,618
	Par-haw_Trx (DBD)			mikado.phaw_50.282275aG397.1	Sun-Mikado	-
	Par-haw_Ash2	Ash2 [Q94545]	Ash2l [Q91X20]	comp174477_c0_seq5	Trinity-Limb	phaw_50.282695g:2178005-2220752
dCOMPASS	includes Par-haw_Ash2 (see above)					
	Par-haw_Set1	Set1 [Q5LJZ2]	Setd1a [E9PYH6]	mikado.phaw_50.283355aG178.1	Sun-Mikado	phaw_50.283355a:9493556-9551696
Trr	Par-haw_Trr	Trr [Q8IRW8]	Kmt2c [Q8BRH4]	mikado.phaw_50.282861eG448.1	Kao-Mikado	phaw_50.282861e:9,250,785-9,312,955
TAC1	includes Par-haw_Trx (see above)					
	Par-haw_nej	nej [M9MS40]	Crebbp [P45481]	mikado.phaw_50.282695dG124.1	Sun-Mikado	phaw_50.282695d:7790068-7814604
	Par-haw_Sbf1	Sbf [Q9VGH9]	SBF1 [Q6ZPE2], SBF2 [E9PXF8]	mikado.phaw_50.283826G523.1	Sun-Mikado	phaw_50.283826:34087962-34200876
Ash1	Psr-haw_Ash1	Ash1 [Q9VW15]	Ash1l [Q99MY8]	mikado.phaw_50.007301aG268.1	Kao-Mikado	phaw_50.007301a:3,925,168-4,034,186
Kis	Par-haw_Kis	Kismet [Q9Y1L3]	Chd6 [A3KFM7], Chd7 [A2AJK6]	mikado.phaw_50.283870G403.1	Sun-Mikado	phaw_50.283870:21945760-22131964
PBAP	Par-haw_Osa	Osa [Q8IN94]	Arid1a [A2BH40], Arid1b [E9Q4N7]	mikado.phaw_50.283815cG377.1	Sun-Mikado	phaw_50.283815c:19852707-19939426
	Par-haw_Mor	Mor [Q7KPY3]	Smarcc2 [Q6PDG5]	mikado.phaw_50.283823aG160.1	Sun-Mikado	phaw_50.283823a:6311700-6337044
	Par-haw_Snr1	Snr1 [Q24090]	Smarcb1 [Q9Z0H3]	mikado.phaw_50.282275bG198.1	Kao-Mikado	phaw_50.282275b:3482023-3497731
	Par-haw_Brm	Brm [P25439]	Smarca2 [Q6DIC0], Smarca4 [Q3TKT4]	mikado.phaw_50.283873aG49.1	Sun-Mikado	phaw_50.283873a:2837385-2858196
BAP	includes Par-haw_Mor*, Par-haw_Snr1*, Par-haw_Brm* (see above)					
	Par-haw_SAYP	SAYP [Q9VWF2]	Phf10 [Q8WUB8]	mikado.phaw_50.000135aG190.1	Kao-Mikado	phaw_50.000135a:3600922-3684882
Other	Par-haw_Fs1h	Fs1h [P13709]	-	comp_163013	Trinity-Limb	phaw_50.283821:4,406,782-4,518,976
	Par-haw_Kto	Kto [Q9VW47]	Med12l [Q8BQM9]	comp_154627	Trinity-Limb	phaw_50.283867c:20,195,413-20,284,996
	Par-haw_Skd	Skd [Q7KTX8]	Med13 [Q5SWW4]	comp_160166	Trinity-Limb	phaw_50.283867c:20,195,413-20,284,996

Supp. Table. 2.5: CRISPR guide sequences

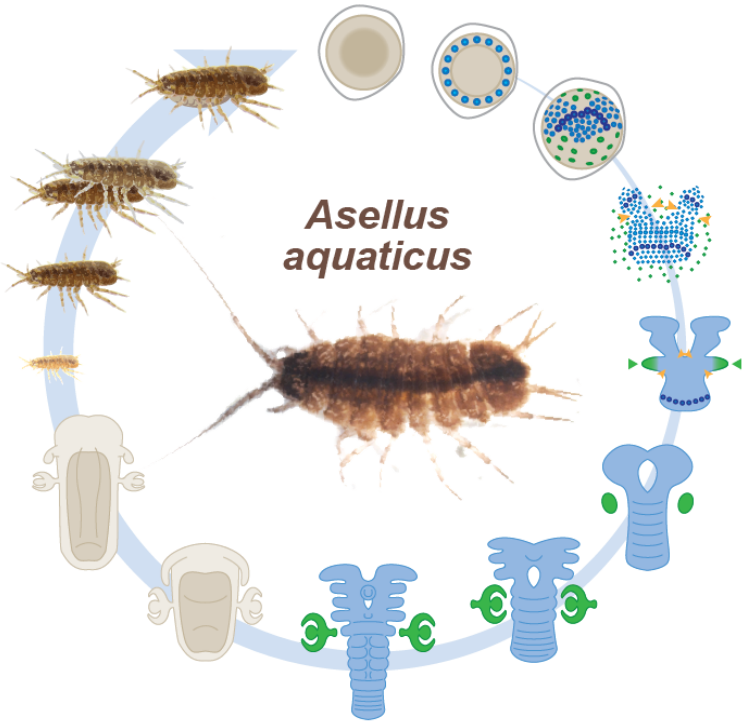
Family	Gene	CRISPR Guide ID	CRISPR Guide Sequence
PcG	Par-haw_Psc-1	Syn078_Psc-1_g1	AAAGCACACACACGAAGTGG
		Syn079_Psc-1_g2	AGCAACAGAGTGTCGCAGTG
	Par-haw_Psc-2	Syn039_PhawPsc2_g1	GTGAGAGGAATTCGGCGTAG
		Syn040_PhawPsc2_g2	CACAGATGAGGTGTGCGTTT
	Par-haw_Sce	Syn041_PhawSce_g1	AACAGCAGAACTACTGTTGT
		Syn042_PhawSce_g2	TGAACTGCATCGAACTCCTC
	Par-haw_Pc	Syn059_Pc_g1	GAATACCTTGTAAAATGGAA
		Syn060_Pc_g2	TTTAGATGGACGCCTGCTTG
	Par-haw_E(z)	Syn084_Ez_g1	ATCATCAATGAAGCCCGTCT
		Syn085_Ez_g2	GCCTCGGGCAACAAGGCCAG
	Par-haw_Pcl	Syn051_PhawPcl_g1	CTCGGGCGCACTTGCAAGCG
		Syn052_PhawPcl_g2	CAGCCAACACAAGAGCAGCT
	Par-haw_Esc	Syn082_Esc_g1	TTGAATGCCACACCAAACAG
		Syn083_Esc_g2	GTGTACCAGTGCGAGAAGCA
	Par-haw_pho	Syn086_pho_g1	CGACTATGCAGAGTACATGG
		Syn087_pho_g2	GGGTCACTGAGATCAACATC
TrxG	Par-haw_Trx	Syn043_PhawTrx_g1	GAGGATCCATGTGGTAATGG
		Syn044_PhawTrx_g2	GGGATCCGGACAGAGGAGGA
		Syn057_PhawTrx_rg1	TGTCAGCGACTGGTTGCCAT
		Syn058_PhawTrx_rg2	GACCAAGTGGTGAAGTCACT
		Syn061_Trx_rd3g1	TCGTAACACCCCTGACACAG
		Syn062_Trx_rd3g2	AGTGCGGTCAGTGCAAGGAA
		Syn076_Trx-SET_g1	GTCCGAGGATGGCTTCAAAG
		Syn077_Trx-SET_g2	GTGGCAGACGGTCTTTGATG
	Par-haw_Ash2	Syn080_Ash2_g1	GCCCCTTGAGACACCATGAG
		Syn081_Ash2_g2	AGCCTGATGGGTCACATGTA
	Par-haw_Kis	Syn055_PhawKis_g1	GGCAGACGTCTCGAGCACTG
		Syn056_PhawKis_g2	CTCAGACACCCAAAAAGAAG
		Syn063_Kis_rd2g1	ACGGAAGAGGAATTACTCAA
		Syn064_Kis_rd2g2	AACAGCAACCTTCTTTCCTG
	Par-haw_Brm	Syn053_PhawBRM_g1	TGGCCAGAGCCCAAGCCCTG
		Syn054_PhawBRM_g2	GGAGGAGGAGGCATTGGTGA
	Par-haw_Fs1h	Syn069_fs1h_g1	AAGCTGCCTGGCGATAAGCT
		Syn070_fs1h_g2	GGATTGGAATCTCTGAGAGA
	Par-haw_Kto	Syn065_kto_g1	TCCTGCTAAATCTTTGAACC
		Syn066_kto_g2	AACCAGATGGCACGCTGCGT
	Par-haw_Skd	Syn067_skd_g1	CTGCCACAAGACTTGGACTG
		Syn068_skd_g2	CAACTGCAGTATTTCCACCT

3 – Developmental staging and embryological techniques for the isopod crustacean *Asellus aquaticus*

Dennis A Sun[1], Lauren Zane[2], Meredith Protas[3], and Nipam H Patel[4,5]

1. Department of Molecular and Cell Biology, University of California, Berkeley
2. Department of Marine Biology and Biological Oceanography, University of Rhode Island
3. Dominican University of California, San Rafael, CA
4. Marine Biological Laboratory, Woods Hole, MA
5. Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL

Abstract



Isopod crustaceans are a group of animals with a distinctive and familiar body plan. Some commonly known isopods include wood lice, commonly referred to as “pill bugs” or “rolly-pollies,” as well as the charismatic giant isopods found roaming the deepest benthic reaches of the ocean. Among isopods, two species have been the focus of developmental biology studies: *Porcellio scaber* and *Asellus aquaticus*. *Porcellio scaber* is a terrestrial isopod commonly used in embryological studies, while *Asellus aquaticus* is a freshwater isopod used to study a variety of ecological factors. *Asellus* has evolved distinct yet genetically interfertile ecotypes which have been the subject of population genetic investigations, with a particular focus on differences between cave and surface populations. This chapter describes recent efforts to establish developmental and genetic tools for *Asellus*. We describe a developmental staging guide and compare the morphogenetic processes observed in *Asellus* embryos to those of *Porcellio* and other well-characterized crustaceans. We describe methods to improve the isolation of early stage *Asellus* embryos for microinjection through animal husbandry and embryo collection techniques, as well as microinjection trial experiments. These results and observations will provide a framework for the further development of *Asellus* as an emerging model system for the study of arthropod development and the evolution of cave traits.

3.0 – Introduction

Asellus aquaticus (Crustacea, Isopoda, Asellidae; (Linnaeus, 1758)) is a freshwater isopod crustacean found broadly across Europe. As an isopod crustacean, one of the defining characteristics of its body plan is the uniform walking leg identity of its thoracic appendages (T2-T8). Isopod crustaceans are phylogenetically grouped among others with amphipod crustaceans such as *Parhyale* in the clade Pericarida, based on their shared ventral brood pouch (Poore, 2002, 2005). Historically, isopods and amphipods had been considered sister taxa, although recent data suggests their relationship is more distant (Schwentner et al., 2018; Wilson, 2009). Some of the morphological differences between isopod and amphipod crustaceans, such as differences in the segmental identities of abdominal segments, have been attributed to changes in *Hox* expression between these groups of organisms (Abzhanov and Kaufman, 2000; Martin et al., 2015; Panganiban et al., 1995). In particular, amphipods have 3 pairs of swimming legs (pleopods) and 3 pairs of anchoring legs (uropods) along with a terminal segment (telson), while isopods have 5 pleopods and 1 pleotelson, a fusion between the A6 appendage and the telson (Fig. 3.1). The shift in the number of pleopods is attributed to changes to the expression of *abd-A* (Abzhanov and Kaufman, 2000; Martin et al., 2015)**.**

Among isopod crustaceans, *Asellus aquaticus* is relatively well-studied in a diverse range of fields. The first studies of *Asellus* began around 150 years ago, when researchers characterized the gross morphology of embryonic development and the ability of adult *Asellus* to regenerate amputated limbs (Dohrn, 1867; Wege, 1909; Zuelzer, 1908). In recent years, *Asellus* has become a platform for studies examining the effects of ecological variables on organismal development and physiology (reviewed in Lafuente et al., 2021). Researchers have probed the microbiome of *Asellus* in different ecological contexts, examined interspecies predator-prey and host-parasite interactions involving *Asellus*, and examined the effects of microbes on sex determination in this organism. The broad geographic distribution and ecological variation presented by *Asellus* positions it as an excellent emerging ecological model system.

Asellus have more recently been established as an emerging genetic model system for the evolution of cave traits. Different populations of *Asellus* found across Europe have independently colonized different freshwater cave ecosystems (reviewed in Verovnik and Konec, 2019). In doing so, these populations have convergently evolved common troglobitic traits, including loss of pigmentation, reduction or loss of eyes, and elongated limbs. Some cave *Asellus* populations remain interfertile with surface populations and can mate in laboratory conditions to give rise to cave-surface hybrid animals, allowing researchers to identify potential genetic determinants of cave traits. For example, previous studies have used quantitative trait locus (QTL) mapping to identify regions of the *Asellus* genome that are associated with specific traits (Protas et al., 2011). Other studies have used differential expression and allele-specific expression analyses to examine genes that differ in expression between cave and surface populations, or have allele-specific expression in hybrids (Gross et al., 2019; Stahl et al., 2015).

These genomic approaches have enabled researchers to identify potential candidate genes that may be responsible for driving cave trait development and evolution. However, genetic perturbations such as RNA interference (RNAi) or CRISPR-Cas9 mutagenesis would be needed to validate the function of such genes (Doudna and Charpentier, 2014; Hutvágner and Zamore, 2002). To date,

genetic perturbation techniques have not been reported for the embryos of any isopod crustacean (although adult RNAi has been reported for *Eurydice pulchra* (Zhang et al., 2013)). As such, establishing such techniques would enable the *Asellus* community to more directly test the function of genes during embryonic development in this organism, and would provide other isopod researchers with a tractable genetic model with the potential for generating transgenic lines. Such techniques would rely on the development of microinjection techniques, as have been established for organisms such as the amphipod crustacean *Parhyale hawaiensis* (Rehm et al., 2009).

In this manuscript, we describe efforts to establish embryological techniques in *Asellus*. We outline an animal husbandry approach that enables the collection of early stage *Asellus* embryos, as well as methods to improve embryo survival during and after collection. Furthermore, we describe early attempts at establishing microinjection for *Asellus*. During the course of this work, we realized that very early stage *Asellus* embryos lack obvious morphological markers. As such, it also became necessary to more carefully characterize the early stages of embryonic development in this organism in order to identify such markers and to determine the optimum timing for microinjections.

Much of previous *Asellus* biology has focused on features associated with adult animals, with fewer studies examining *Asellus* development. Previous work has roughly characterized the broad stages of *Asellus* development in the context of *Hox* expression and the development of cave-associated traits (Mojaddidi et al., 2018; Vick and Blum, 2010; Vick et al., 2009). Much older work has also described some of the developmental stages of *Asellus* and related species through hand illustrations (Dohrn, 1867; McMurrich, 1895). In contrast, detailed descriptions of embryonic development exist for the terrestrial isopod *Porcellio scaber* (Milatovič et al., 2010; Wolff, 2009). Previous work has relied on *Porcellio* development as a proxy for *Asellus* development, but the potential differences in development between these organisms have not been investigated. More careful descriptions of early *Asellus* development would fill a gap in the current literature, and provide the tools for more complete comparisons between these two isopod crustaceans.

In this chapter, we also examine the early development of *Asellus* using brightfield microscopy and nuclear staining. We provide a rough staging guide of *Asellus* development, with particular attention to the morphological features that can distinguish the early developmental stages. In addition, we compare the developmental progression in *Asellus* to that of *Porcellio* to reveal that the two isopod crustaceans have a number of notable differences during early developmental processes. These data will provide a useful framework for future studies of *Asellus* development and reveal a number of interesting evolutionary questions for future investigations.

By improving animal husbandry techniques and identifying the timing of cellularization, we also set the stage for microinjection experiments in *Asellus*. We describe trial injection experiments using mRNA for fluorescent proteins, with results suggesting that microinjection may be possible. This work will provide the foundation for establishing genetic manipulation in this emerging eco-evolutionary model organism.

3.1 – *Asellus* animal husbandry and reproductive biology

Asellus used in this study were surface ecotype animals in a population originally isolated from Rakov Škocjan in Slovenia. Mature males and females of the surface ecotype are sexually dimorphic,

with females having smaller, more oval-shaped bodies and males having larger, elongated, more triangular-shaped bodies and elongated posterior thoracic limbs (Fig. 3.1.1). Adult males also appear to have a genital appendage around the posterior end of the thorax that is not present in females.

Prior to mating, *Asellus* adults form mating pairs in a pose referred to as "amplexus." Larger males will grasp the smaller females from above using their elongated thoracic legs and maintain this coupled position until mating. After fertilization, the female is released and the fertilized embryos are deposited into a marsupium or "brood pouch" composed of medial plates (oostegites) located on the ventral side of the female which overlap to form a sealed compartment (Fig 3.1.1G, H). Females with embryos ("gravid" females) are easily identified when flipped so the ventral side of the animal is visible. Embryos can be seen through the semi-transparent brood pouch plates as small, semi-opaque beige spheres.

Adult animals used in this study were raised at 10°C in Artificial Pond Water (APW) and fed a diet of rotting leaves collected from Strawberry Creek on the University of California, Berkeley campus (see 3.1.6 – Materials and Methods for further details). As *Asellus* are highly cannibalistic of their juveniles and hatchlings, we regularly collected embryos from gravid female *Asellus* and raised hatchlings in cohorts of 20-30 individuals in isolated plastic containers (Fig. 3.1.1I). We raised embryos and hatchlings at 18°C to accelerate their development. After hatchlings appeared to have darker pigmentation and reached around 2mm in size, we classified them as "juveniles" and moved them to much larger tanks containing juveniles from different cohorts, which were kept at 10°C. Animals that grew large and appeared sexually dimorphic were regularly transferred into an "adult" tank for mating and embryo collection.

3.2 – Strategies for the isolation of early stage *Asellus* embryos

A major early challenge of collecting *Asellus* embryos for injection and embryological studies has been the unreliable pace of mating after amplexus. In our observations, amplexus pairs could take several weeks to separate after forming. The unreliability of mating can prove a major hindrance to performing research in this system, particularly in embryo injection experiments.

We determined that *Asellus* embryos and hatchlings could safely be raised at 18°C to accelerate development (see section 3.3 for further discussion). Given that the normal geographic distribution of *Asellus* is within temperate climate regions, and some populations of *Asellus* undergo seasonal mating (Andersson, 1969; Murphy and Learner, 1982; Økland, 1978), we wondered if shifting paired adults to higher temperatures might also "trick" them into mating faster by mimicking a temperature shift from winter to spring. We observed that shifting amplexus pairs from 10°C to 18°C in the evening usually resulted in animals mating overnight. This allowed us to collect very early embryos for attempts at microinjection.

When first deposited into the brood pouch, embryos have an irregular appearance similar to the tightly-packed honeycomb of cells in an epithelium. Embryos cannot easily be collected at this stage, in part because the brood pouch is tightly sealed, with the individual plates overlapping like an envelope, and in part because the embryos are extremely delicate and prone to bursting when any force is applied.

We attempted early embryo collection and injection after the embryos developed a more spherical appearance (Fig. 3.2). However, embryos collected at those stages do not appear to progress further through development and eventually die. When collecting embryos from this very early stage, we observed that opening the brood pouch results in the release of a slightly viscous liquid. The liquid released appeared to disperse in a similar fashion to the mixture of water of two different salinities/osmolarities, suggesting that the brood pouch may function similarly to the *Porcellio* marsupium in maintaining a specialized environment for embryonic survival.

Previous experiments in *Porcellio scaber* raised embryos in a slightly saline solution (isopod saline) (Surbida and Wright, 2001). We determined that *Asellus* embryos collected at very early stages and raised in a 10% isopod saline solution were able to progress through development until hatching. Upon hatching, animals were raised in APW or spring water.

We also developed improved techniques for embryo collection from the marsupium. Prior to embryo collection, gravid *Asellus* females were anesthetized in 20μL clove oil in 50mL spring water for 15-20 minutes, and then anesthetized adults were rinsed in spring water and transferred to spring water for collection. In the absence of rinsing with spring water, embryos did not survive to hatching.

Embryo collection from amphipod crustaceans such as *Parhyale* often uses a pulled glass Pasteur pipette with a rounded hook, referred to colloquially as "embryo scrapers." The *Asellus* marsupium is considerably more rigid than the *Parhyale* marsupium, and the larger plates of *Asellus* require multiple embryo scrapers: one to hold open the plates, and another to scrape out the embryos. This technique is very cumbersome, and runs the risk of accidentally killing the female if one uses too much force to hold her down during the scraping process.

Rather than using the embryo scraper approach, we have found that simply using a glass Pasteur pipette fitted with a rubber bulb allows one to "blow" embryos out of the brood pouch. To perform this technique, one first draws a small volume of liquid into the Pasteur pipette, then slides the tip of the pipette between the plates of the brood pouch using a gentle side-to-side motion. Once the Pasteur pipette is in place, rapidly injecting liquid into the brood pouch causes embryos to be jettisoned. This technique has the drawback of spraying embryos in random directions outside of the brood pouch, some of which might be crushed after moving the adult female. Moreover, it is usually quite difficult to recover all of the embryos from the brood pouch using this technique, as some will become jammed between plates or protrusions within the brood pouch. However, given that *Asellus* females can carry nearly 100 embryos per clutch, the loss of a few embryos is a small sacrifice compared to the potential to accidentally kill the female with a riskier technique.

3.3 – Stages of *Asellus aquaticus* development

During the early stages of my experiments with *Asellus*, prior to determining that temperature shift appeared to increase the frequency of mating, we collected many clutches of embryos of various stages and observed their development at 18°C. From the individual embryonic timelines we observed, we were able to characterize the general trajectory of *Asellus* development over time. As we raised these *Asellus* embryos, we fixed and DAPI-stained a few embryos each day. Compiling the data we generated from these short, semi-overlapping time course experiments, we have generated a broad picture of *Asellus* embryonic development.

Very early stages of *Asellus* embryos are morphologically difficult to distinguish using ambient or overhead illumination. During the course of our observations, we discovered that transillumination of *Asellus* embryos revealed morphological features stereotypical of different developmental stages, and would prove essential for the proper staging of early stage embryos. For most of the figures we display in our timecourse, we show both ambient/angled and transmitted illumination, which should assist future *Asellus* researchers in determining the developmental stage of freshly collected embryos.

Early cell divisions (0-1dpf)

The earliest stages of *Asellus* embryonic development are difficult to characterize, as the embryos are very delicate. When we performed my initial characterization experiments, we had also not yet determined that isopod saline was required for embryonic viability at this stage, and so embryos we collected at this stage did not progress through development.

Despite these drawbacks, we were able to collect very early *Asellus* embryos and determine some morphological landmarks that could distinguish them from later developmental stages. The earliest *Asellus* embryos we were able to collect appeared not to have completed the first cell division (Fig. 3.2). These embryos appeared to have a nuclear "shadow" when transilluminated. Sibling embryos, some of which were likely slightly more developed, often showed multiple nuclear shadows upon transillumination. These results are consistent with observations in *Porcellio scaber* that the region around the nucleus of the embryo is noticeably more opaque, casting a shadow upon transillumination (Wolff, 2009).

DAPI stains of these embryos were not able to distinguish any obvious nuclear morphology (data not shown), and appeared to show uniform background staining of the yolk. A potential explanation for this difficulty could be that DAPI also binds tubulin, which may be abundant in the early embryo. Early embryos may also be more likely to die upon removal and exhibit degraded morphology. Moreover, upon fixation and methanol treatment of *Asellus* embryos, the normally semi-opaque yolk becomes opaque and white, potentially interfering with the proper imaging of the embryonic nuclei, which would be located deep within the center of the opaque tissue.

Future work should attempt to optically "clear" the early embryonic tissue for improved imaging and will benefit from collecting embryos into 10% isopod saline, to facilitate embryonic viability after collection. Alternatively, using DAPI or other stains in living embryos, along with other vital markers, could better characterize the earliest developmental stages.

From around 0.5-1dpf, multiple nuclear shadows can be observed in transillumination views. At this stage, indentations can be seen along the surface of the embryo, suggesting that cell boundaries may be forming. If true, this would mean that *Asellus* embryos undergo cellularization at a much earlier stage than is reported for *Porcellio* embryos, which appear to develop distinct cell membranes at around the germ cap stage. The presence of early cellularization means that *Asellus* embryos should be injected within the first 12 hours of development for genetic perturbations to be able to affect all cells of the embryo.

"Soccerball" or blastoderm stage (2dpf)

As it was previously not possible to raise embryos collected at the earliest developmental stages, we determined the relative timing of this developmental stage by collecting embryos about 2 days after amplexus pairs separated. Embryos collected at this stage appeared to display a visible "ring" in transillumination, which is barely noticeable in the overhead illumination view (Fig. 3.3). DAPI staining of embryos at this stage show a uniform distribution of nuclei along the surface of the embryo, which is morphologically similar to the "soccerball" stage (S6) in *Parhyale* (Browne et al., 2005), or early stage 3 embryos of *Porcellio* (Milatovič et al., 2010; Wolff, 2009). The gross morphology of a central yolk surrounded by a uniform sheet of cells also resembles the blastoderm stage of *Drosophila* development. Moreover, the distinct boundary separating the cell and yolk layers suggests that cellularization has completed by this stage, again much earlier than has been reported for *Porcellio*.

Gastrulation/ germ cap (3dpf)

A day after the "soccerball" stage, embryos appear to show a very small transparent region on one side of the embryo, which is barely perceptible as a cloudy indentation in the overhead illumination view (Fig. 3.3). DAPI staining of embryos at this stage reveals that nuclei of different sizes can be found in a crude morphological organization. Sibling embryos of slightly younger and older stages appear to show the organization of a row of cells with larger nuclei ("large cell row"). These larger cells are bounded "above" by the two gradually developing lobes, and "below" by a small focus of cells. Further "below" that region appears to be a region of sparse and uniformly spaced extraembryonic cells, reminiscent of the extraembryonic tissue found in the *Parhyale* germ cap stages of development (S8-S11).

Based on the morphology of the *Asellus* embryo in comparison to *Porcellio* embryos, it is likely that this large cell row represents the ectoteloblasts previously described in *Porcellio*, which give rise to the parasegmental rows of the body. This cell row at early stages is not a straight line, but instead a semicircle around the site of gastrulation, which gradually straightens into a row as more ectoteloblasts are recruited. At this developmental stage, the ectoteloblast row consists of 10-12 cells.

In *Porcellio* embryos at this stage, the condensation of cells in the presumptive posterior is the result of the gastrulation process, wherein some cells migrate from the surface of the embryo into the interior to form the mesoderm. An analogous process likely also occurs at this stage in *Asellus* development.

Early germband (4-5dpf)

Following the germ cap stage, *Asellus* embryos develop morphology that is more obviously recognizable as a malacostracan embryo, with morphologically obvious rows of cells that will become the trunk and abdomen, as well as two lobes of cells that will become the head (Fig. 3.4). While the embryo at this stage is very hard to distinguish from the yolk in both transillumination and overhead illumination, DAPI stains of embryos at this stage are much easier to compare to those of *Parhyale* and *Porcellio*. *Asellus* embryos develop two lobes of cells in the anterior and progressively add parasegmental rows to the trunk and abdomen in the posterior of the embryo. At ~4dpf, the large cell row is

morphologically identifiable as an ectoteloblast row similar to what has been previously described in *Porcellio*.

A notable difference in *Asellus* embryos in comparison to *Parhyale* or *Porcellio* embryos is the presence of additional large cells in the head lobes. These cells are larger than their neighbors, positioned towards the anterior-most region of the head lobes, and resemble the presumptive neural structures found in the anterior of branchiopod crustaceans such as *Daphnia magma* (Mittmann et al., 2014). In addition, the apparent head lobes of *Asellus* embryos appear morphologically discontinuous with the rest of the body, with a gap separating the head lobes from the trunk parasegmental rows. This gap is most obvious from the lateral and anterior views of the embryo, where a region of the embryo that is only sparsely populated by cells separates the head lobes from the rest of the body. The two head lobes are also separated from one another by a large gap, similar to what has been described in *Parhyale* and *Porcellio*.

The morphological gap between head structures and the trunk parasegmental rows continues to be visible at 5dpf, although the two regions of the body appear to gradually become more contiguous. At this stage, the gap between the two head lobes appears to become more obvious, and appears grossly similar to the split between head lobes in *Porcellio* stage 6 embryos. The posterior ectoteloblast region remains morphologically distinct from surrounding cells as a row of large cells. There appear to be 19 ectoteloblasts in the posterior of the embryo, consistent with the number of cells found in other crustacean groups (Dohle and Scholtz, 1997).

Moreover, at 5dpf, a large group of cells appears to be condensing into two rounded structures on the lateral sides of the embryo, slightly anterior to the gap between the head lobes and the trunk parasegmental rows. These cells appear to be organized in a circular pattern around a focus, and are contiguous with the embryonic head lobes, suggesting that these cells may be embryonic in origin. The position of this condensation of cells is consistent with the position of the *Asellus* lateral organ, which becomes more obvious at later developmental stages. Notably, the anterior-posterior position of the *Asellus* lateral organ bud appears to be around the mandibular segment (Mn), which is the final segment of the anterior that develops without contribution from the trunk ectoteloblast rows.

Middle germband (6dpf)

By around 6dpf, the lateral organ primordia appear to be fully separated from the rest of the embryo as a pair of cell condensations to the distal and lateral sides of the embryo (Fig 3.5). At this stage, the anterior-most regions of the head lobes appear to connect along the embryonic midline, creating a "gap" in the middle of the head. Moreover, the parasegments of the body appear to become more morphologically obvious as indentations along the body axis. The ectoteloblast row remains visible at this stage in the embryonic posterior.

Late germband/ limb bud elongation (7-9dpf)

At ~7dpf, the limb primordia in the head of the embryo become obvious as limb buds, and the lateral organ appears to become fully developed with its distinctive and unusual shape, composed of a central circular lobe and two pointed lobes that together resemble a crescent moon (Fig 3.5). In the posterior of the embryo, the ectoteloblast row that was morphologically obvious at earlier stages is now

not easily distinguishable, although a group of cells with slightly larger nuclei and slightly dimmer DAPI staining is visible.

By ~8dpf, the limb primordia of the trunk become more obvious as limb buds. In contrast to the limb buds of the head segments, which appear to grow out along the proximal-distal axis of the embryo, the thoracic limb buds appear to grow from the lateral edge of the embryo towards the midline. In addition, the gap in the head of the embryo appears nearly sealed, and the presumptive stomodeum/foregut becomes obvious with a distinctive U-shaped pucker-like morphology. In the posterior, a ring-like condensation of cells appears to form the presumptive anus.

The embryo also fully wraps the yolk at this stage, with anterior and posterior portions of the embryo nearly coming into contact. Between the anterior and posterior ends of the embryo, the yolk appears to begin a process of invagination, which will eventually morphologically separate the anterior and posterior ends of the embryo.

At around this developmental stage, embryos begin to undergo their first "hatching" event, wherein the loose outer chorion of the embryo is shed, and the lateral organs become released, similar to a pair of wings around a small spherical ball. Colloquially, we referred to this stage as the "snitch" stage.

By ~9dpf, the limbs of the embryo have substantially elongated, and an additional membrane becomes obvious in the gap between the invaginating anterior and posterior halves of the embryo. In the posterior of the embryo, cells begin invaginating towards the core of the embryo, resulting in a condensation of cells and brighter DAPI staining.

The developmental stages of the embryo beyond this stage were considerably more difficult to profile with nuclear staining, as the embryo became increasingly difficult to dissect. I observed at least 3 different additional "ecdyses" beyond this stage, which contributed to the difficulty dissecting the embryos and laying them flat, and may have contributed to uneven DAPI staining along the body axis. In the interest of providing a reference for future studies, I have included these images from my dissection attempts in this thesis chapter, but the remaining morphological stages of development will be described primarily based on the morphology of the embryos under brightfield microscopy.

Body and limb elongation (10-12dpf)

In the later stages of embryonic development, the embryo gradually unfurls, pushing the anterior and posterior regions away as the body axis elongates and straightens (Fig 3.6). Around 10dpf, the embryo undergoes a second "hatching" to reveal a pair of antennae. The lateral organs also rotate over the course of body axis elongation so that they become perpendicular to the body axis, rather than parallel.

"Angel" stage (14dpf) to hatching (16dpf)

Sometime between 12-14dpf, the embryo undergoes a third "hatching" that exposes the remaining limbs (Fig 3.6). At this point, the embryo appears to be a nearly fully-developed manca, with the exception of the retained and rotated lateral organs, which resemble a pair of wings. The limbs are freely exposed at this developmental stage, and can move separately from each other. We colloquially referred to this late pre-hatching stage as the "angel" stage.

Upon hatching at around 16dpf, the embryo sheds a final membrane that includes the lateral organs to undergo its true ecdysis. This membrane tightly encases each of the limbs of the hatchling, such that during the hatching process, the manca appears as though it is trying to worm its way out of a very complex set of pants.

Comparisons to previous studies

Previous work has examined *Asellus* development on a coarse scale, marking a handful of major morphological events in this organism (Mojaddidi et al., 2018; Vick and Blum, 2010). By cataloging *Asellus* development along a timecourse, we generated a more comprehensive view of *Asellus* development and the morphogenetic processes occurring within the embryo (Fig. 3.7).

We compared our timecourse analysis to previous studies to examine how development might vary under different temperature conditions (Fig. 3.8). While the Vick and Blum study did not provide information about the precise number of days or the temperature at which embryos were raised, we were able to benchmark development using the "angel" stage. The Mojaddidi et al. study raised embryos at a lower temperature (12°C), which resulted in slower development. Embryos reached the "angel" stage at around 32 dpf in the Mojadiddi study, in comparison to around 14 days in our study, suggesting that a shift in temperature from 12°C to 18°C caused development to complete in around half as much time. These results may prove useful for examining developmental timing in different populations of *Asellus* which preferentially develop at different temperatures.

3.4 – Injection trials for *Asellus* embryos

In addition to characterizing the early stages of *Asellus* development, we also performed preliminary injection experiments in *Asellus* embryos. For these experiments, we used borosilicate glass needles pulled as specified for *Parhyale* injection experiments. We injected DsRed-NLS mRNA (see methods) into early embryos, for which we could see the nuclear shadows suggesting very early embryonic stages (Fig. 3.9).

One day post-injection (1dpi), we imaged the embryos using fluorescent microscopy. We observed in some embryos apparent red fluorescence that was markedly distinct from autofluorescence found in green or blue fluorescent channels. Some of the morphology of the red fluorescence was reminiscent of the nuclear shadows observed in 1dpf embryos, suggesting that DsRed-NLS mRNA may have been transcribed in those embryos. After 1dpi, we recovered no surviving embryos. We repeated these experiments in three separate trials, and observed similar results across the experiments: we observed DsRed fluorescence in some embryos, but no embryos survived after 1dpi.

These results suggest that embryonic injection is possible in *Asellus* embryos, but survivorship post-injection remains a major challenge. When we first conducted these experiments, we had not determined that embryos collected at very early stages may require isopod saline for survival. It is possible that the embryos died not as a result of injection trauma, but instead as a result of lack of isopod saline. Future experiments should attempt collection and injection in isopod saline, which may allow for embryos to survive until hatching. If these injection experiments had indeed been successful, microinjection of mRNAs would allow for live imaging of *Asellus* embryonic development or introduction of CRISPR-Cas9 complexes for targeted mutagenesis.

3.5 – Discussion

The isopod crustacean *Asellus aquaticus* shows great promise as an emerging ecological and evolutionary-developmental model system. To further empower *Asellus* researchers, we have generated a timecourse of *Asellus* development, revealing unique processes that distinguish its development from that of other isopod crustaceans such as *Porcellio scaber*. We determine that, unlike *Porcellio*, which appears to cellularize its embryos around the time of gastrulation, *Asellus* appears to generate cell membranes as early as within the first 12hrs after egg laying. In addition, we have described animal husbandry practices and injection trials that will serve as a foundation for the development of genetic perturbations such as RNAi and CRISPR-Cas9 mutagenesis in this system.

By analyzing the developmental progression of *Asellus* in comparison with other crustaceans, we suggest that *Asellus* has several unique developmental features that deserve further investigation: the separated development of the head and thorax during early embryogenesis; the developmental and evolutionary origin of the *Asellus* lateral organ; and the origin of the numerous false "hatching" events during the late stages of development.

3.5.1 – The potential for dual growth zones during *Asellus* germband elongation

The head and trunk regions of the *Asellus* embryo appear to be separated by a morphological gap during the early stages of development, from 4-5dpf. In other malacostracan crustaceans, the head segments and trunk segments form through distinct developmental mechanisms: the head segments appear to develop from cells that do not have a stereotypical organization, while the trunk segments develop from an ectodermal germband grid. In the case of *Asellus* and other isopods such as *Porcellio*, this ectodermal germband grid appears to emerge as a result of the division of a population of stem cells known as ectoteloblasts, which form a row of cells in the posterior-most region of the embryo.

Despite this bifurcation in mechanisms, a physical separation between the two sets of structures has not been previously reported in any malacostracan, indicating that *Asellus* may have unique early developmental processes that result in a morphological gap between head and trunk structures. This may be a direct consequence from the heterochronic shifts in developmental timing that have been described across diverse crustacean groups, which are hypothesized to result from the transition from an ancestral indirect development to direct development (Jirikowski et al., 2015).

In the anterior of the *Asellus* embryo, we observed two groups of large cells, one in each of the two bilobed sections of the separated head. In the branchiopod crustacean *Daphnia magna*, a similar pair of kidney-shaped structures composed of large cells develops at early developmental stages. These structures are purported to develop into neural structures. Such large cells have not been characterized in *Porcellio*, or in amphipods such as *Parhyale* and *Orchestia*. It is possible that the large cells found in the anterior of the *Asellus* embryo have a similar developmental function to the large cells found in *Daphnia*, and that these cells become neural structures later in development. Such a developmental trajectory would be unique among the isopod and decapod crustaceans studied to date.

Another hypothesis would be that the large cells in the anterior of the *Asellus* embryo are, in fact, ectoteloblasts of the head. The large anterior cells appear morphologically similar to the ectoteloblasts of the posterior of the embryo in their size and diffuse DAPI staining in comparison to

their smaller neighbors. The morphological gap that exists between the head lobes and the trunk at 4dpf, which shrinks as development progresses towards 5dpf and 6dpf, might suggest that the trunk and head structures are growing towards each other. In such a model, the *Asellus* embryo might have two separate growth and segmentation processes: the posterior growth zone, which gives rise to the ectodermal germband grid of the trunk, and which is conserved to other isopods and amphipods; and an anterior growth zone, which gives rise to the segments of the head.

To test these hypotheses, it will be necessary for additional tools to be developed and applied in the *Asellus* embryo. For example, nuclear labeling using injectable mRNAs would allow for cell lineage analysis over live developmental timecourses. Such experiments could enable researchers to determine how cells in the anterior and posterior of the embryo are organized, and whether the large cells of the anterior develop into neural structures, or give rise to the segments of the head. Other approaches, such as improved *in situ* hybridization or antibody staining techniques, would allow researchers to determine the genetic identity of the large cells of the anterior and posterior to better reveal their function during development.

3.5.2 – The developmental and evolutionary origin of the *Asellus* lateral organ

Over the developmental timecourse presented in this chapter, we reveal that the *Asellus* lateral organ appears to emerge laterally from the head region of the embryo. Moreover, this region falls in the same anterior-posterior axis region as the developing mandible. In the stages presented, the lateral organ appears to progressively bud off from the ectodermal cells of the body beginning at 4dpf to become its own distinct, fully-formed structure by 7dpf. These data, when considered in comparison to previous hypotheses and the development of other crustaceans, suggests that the *Asellus* lateral organ may share an evolutionary-developmental origin with other lateral structures, such as the *Daphnia* carapace.

The developmental origin of the *Asellus* lateral organ has not been previously characterized, and its function during development remains unclear. Previous work has suggested that the lateral organs may serve as either an embryonic kidney or respiratory structure. Such hypotheses would suggest that the lateral organ is an extraembryonic structure, and would likely develop from the uniformly distributed extraembryonic cells of the embryo. Another hypothesis might be that the lateral organ is homologous to the dorsal organ found in *Parhyale*, *Orchestia*, or other amphipod crustaceans, which is also formed from extraembryonic cells. The data presented in this section suggests that the lateral organ is actually embryonic and ectodermal in origin, budding directly off from the mandibular segment, condensing laterally, and then growing outward to form its unique trident shape.

The position of the lateral organ is particularly notable when considered in the context of the development of other crustaceans groups. In particular, the lateral organs appear to emerge from the same anterior-posterior position as the carapace of the branchiopod crustacean *Daphnia magna* (Bruce and Patel, 2021; Shiga et al., 2017). In *Daphnia*, the carapace develops from the region of the embryo dorsal to the mandibular segment, as evidenced by expression of *Scr* and *Dfd*. In *Asellus*, this same region appears to bud off to become the lateral organ. Given the apparent positional similarity between these structures, it is possible that a conserved developmental program may pattern these two processes.

Previous work by Shiga et al. suggested that the *Daphnia* carapace may be a novel, co-opted structure that borrows the genetic network responsible for developing insect wings (Shiga et al., 2017). A more recent study by Bruce and Patel contradicts this view by showing that the *Daphnia* carapace may be a structure known as an "exite," or a protruding structure emerging from an arthropod limb (Bruce and Patel, 2021). In the latter study, the authors posit a model of "cryptic persistence" as an explanation for how such seemingly novel structures might arise. In a cryptic persistence model, developmental fields and genetic programs are retained in intermediate lineages from which apparently novel structures then develop. These developmental fields do not give rise to obvious structures in intermediate lineages, but the persistence of developmental potential allows seemingly new structures to form in later-branching organisms.

The *Asellus* lateral organ may serve as a platform to examine these two hypotheses of gene network co-option versus conserved, cryptic developmental fields. By performing *in situ* hybridization experiments on *Hox*, wing, and leg gap genes in *Asellus* embryos, it should be possible to determine the exact proximal-distal register of the lateral organ. Moreover, live imaging experiments using nuclear dyes or injectable mRNAs could more precisely reveal the morphogenetic processes leading to lateral organ outgrowth. By examining another type of seemingly novel structure in a different crustacean, it should be possible to establish whether the *Asellus* lateral organ and *Daphnia* carapace share an evolutionary-developmental origin.

3.5.3 – Multiple false "hatching" events during *Asellus* development and similarities to nauplii

During the course of *Asellus* development, the embryo undergoes multiple "hatching" events that progressively reveal developmental structures. The timing of these hatching events is interesting when considering the development of the naupliar and postnaupliar appendages of this organism. In many crustaceans, including many clades of Malacostracan crustaceans, of which isopods are a member, embryonic development includes a naupliar stage, wherein the embryo develops its first three segments – An1, An2, Mn – and then hatches as a free-floating nauplius, which grows additional body segments and limbs as it develops (Jirikowski et al., 2013, 2015).

In the case of *Asellus*, the timing of the multiple false "hatching" events seems to mirror to some degree the phases of naupliar development found in other malacostracans. For example, the second hatching of *Asellus* reveals the second antennae and head, while the third hatching reveals the entirety of the developing limbs. Moreover, even the first hatching results in the release of limb-like structures – the lateral organs – reminiscent of the hatching of a naupliar larvae. The parallels between the progressive naupliar ecdyses and the embryonic hatchings of *Asellus* could be a result of a conserved set of developmental processes shared between true naupliar crustaceans and direct-developing crustaceans.

Asellus has several features that suggest that it may be an excellent candidate for examining the evolution of naupliar development. These include: the unusual morphological separation between the head and trunk, which is the same separation as the naupliar and post-naupliar appendages; the emergence of the lateral organs, which seem to share a positional origin with the carapace of *Daphnia*; and the presence of multiple limb-revealing "hatchings" prior to the final, true hatching that results in

a free-moving manca. Each of these features show an interesting intermediate developmental morphology to naupliar and direct-developing crustaceans, suggesting that the embryonic development of *Asellus* may represent a "bridge" stage between these two life history approaches.

Future studies will need to more precisely examine the early development of *Asellus* in comparison to the development of naupliar embryos, both in terms of morphogenetic movements and gene expression. By examining *Asellus* development more closely in comparison to better-studied organisms such as *Parhyale* or *Daphnia*, it might be possible to understand how these distinct crustacean life histories have evolved.

3.5.4 – Next steps and future directions

The data presented in this chapter provide the groundwork for future embryological and genetic manipulations in this organism. Notably, the preliminary injection trials suggest that it is possible to introduce exogenous nucleic acids, such as mRNAs, and that such injected materials can be transcribed to generate fluorescent proteins. However, these experiments resulted in total embryonic lethality, which could be a result of the unique osmolar requirements of early *Asellus* embryogenesis. Future attempts at injections could use 0.5X isopod saline to improve early embryonic viability, and would benefit from the improved animal husbandry techniques that enable the more reproducible isolation of early stage embryos.

In addition, since the early investigations of *Asellus* development, more advanced techniques have been developed that could enable investigations into gene expression in this organism. In particular, past experiments used colorimetric *in situ* hybridization to examine gene expression, in whole mount and on slices of embryonic tissue, to variable success. One major challenge for such experiments could be the numerous layers of membranes that emerge during development.

Recent work using the *in situ* hybridization chain reaction approach (*in situ* HCR) have enabled staining through the cuticle of adult crustaceans such as *Daphnia magna* (Bruce and Patel, 2021; Bruce et al., 2021). Given the success of such approaches, establishing *in situ* HCR for *Asellus* could enable the characterization of expression patterns for numerous developmental genes, such as leg gap genes or *Hox* genes. Moreover, *in situ* HCR could also enable the direct staining of genes of interest from cave and surface population embryos, allowing researchers to more precisely identify the developmental-genetic basis of cave trait evolution, and to validate the hypotheses that have been generated using currently available data.

3.5.5 – Conclusion

By collecting and staining timecourses of *Asellus* embryos, I have filled in a gap in the current knowledge of this emerging research organism and provided a guide for future embryological research. I demonstrate a number of interesting developmental features of *Asellus* that may prove fruitful for future evolutionary-developmental comparisons. Moreover, through improved animal husbandry and early injection trials, I demonstrate the potential for genetic studies in *Asellus*. Future work to establish genetic perturbation and provide more effective gene expression visualization tools in this organism will help establish *Asellus* as a novel and exciting genetic system for evolutionary research.

3.6 – Materials and Methods

Asellus aquaticus animal husbandry

Adult surface animals were received as a gift from Meredith Protas, and were originally collected from Rakov Škocjan in Slovenia. Animals were raised in large tanks of artificial pond water (APW) consisting of 5.32g Tropic Marin Sea Salt, 36.8 mL of CaCl2 solution (5.88g CaCl2 in 50mL H2O), 35.2mL of NaHCO3 solution (1.26g NaHCO3 in 50mL water) in 16L of distilled water. Animals were fed with decaying leaves collected from Strawberry Creek on the University of California, Berkeley campus. Decayed leaves were dredged up from the creek and visually cleaned of any macroscopic organisms (mosquito larvae, amphipods, trichopteran larvae, etc.) before being rinsed and stored in APW before feeding. We attempted to culture rotting leaves by taking dried leaves and mixing them with the debris leftover from the rotting leaf tank, but the animals did not seem to eat these cultured leaves. All tanks were oxygenated using a standard aquarium air pump and airstone and water was changed every few months, or added as evaporation reduced the water level. Water removed during water changes was bleached to a final volume of 10% bleach to avoid any potential contamination to local water systems.

Embryo collection and culture

Adult tanks were checked periodically for pairs in amplexus, which were then separated into individual, smaller tanks. Tanks were inspected each morning for separated pairs and separated females were flipped ventrally to see if they were gravid. Gravid females were anesthetized in 20μL clove oil in 50mL APW for around 20-30 minutes prior to embryo collection. Embryos were collected initially using a pair of flamed glass pasteur pipettes similar to those used for *Parhyale* embryo collection, but eventually were collected simply by jetting water into the brood pouch to eject embryos into solution. Embryos were raised in APW or spring water (e.g. Poland Spring) at 18°C until hatching. Hatchlings were raised in clutch-specific tanks of no more than 30 animals. Juveniles were transferred to a large juvenile tank and mature adults were moved from the juvenile tank into the adult tank. The separation of animals of different ages was a measure to decrease cannibalism, as suggested by Prof. Protas.

Embryo fixation, dissection, and staining

Embryos were fixed in 9 parts APW to 1 part 37% paraformaldehyde for either 1-2hrs at room temperature (~21°C) or overnight at 4°C. During fixation, the most well-separated embryonic membranes were removed using a pair of sharpened tungsten needles in an analogous fashion to *Parhyale* embryo dissections. At early stages, after removing the chorion, additional membranes were present but could not be removed easily without destroying the embryo; for those stages, a puncture was inserted randomly into the embryo, or when possible, into the region of the embryo without obvious embryonic tissue. Fixed embryos were washed 3x in PT and stained in 0.1 μg/mL DAPI in PT for up to 2 weeks at 4°C. After staining, embryos were washed in PT and transferred to 50% glycerol/ 50% 1X PBS and 70% glycerol/ 30% 1X PBS for mounting.

Microscopy

Live embryos, hatchlings, and adults were imaged using a Leica (model no.) trinocular microscope or a Keyence (model no.) microscope. The Leica microscope had a glass platform with an adjustable mirror, allowing for total transmission of light or angled illumination settings. Transmitted light and angled illumination were essential for revealing the morphological characteristics of early (0-4dpf) embryos, and should be used for future experiments. DAPI-stained embryos were imaged using a Zeiss (model no.) compound fluorescence microscope. For DAPI-stained embryos, a z-stack was manually performed by slowly taking images of different focal planes and focus-stacking using the Helicon Focus software. For images used in most figures, a single embryo was imaged from multiple angles by rolling the embryo between a glass cover slide and the microscope slide while the embryo was embedded in a 70% glycerol/ 30% PBS solution. For injection experiments, fluorescence was examined using a Zeiss (model no.) fluorescent dissecting microscope.

Capped messenger RNA transcription

Capped DsRed-NLS messenger RNA for injections was made using the mMESSAGE mMACHINE kit (Life Technologies, now a part of Thermo-Fisher Scientific). After digestion with DNAse and LiCl precipitation, mRNA was quantified using a Nanodrop and mixed to a final concentration of 1μg/μL in nuclease-free water and filter-sterilized 0.05% phenol red. Injection mixes were aliquoted in 1.5μL volumes, stored at -80°C, and used once before being discarded, as capped mRNA tends to degrade within 24hrs after being thawed.

Microinjection

Embryos were injected using needles used for *Parhyale hawaiensis* microinjections, following the standard procedure for those injections. Needles were able to penetrate the embryo very easily, as evidenced by obvious phenol red staining entering and remaining within the embryo after injection. Early injected embryos (injected around 0-0.5dpf) never survived injection, although it remains unclear whether the lack of survival was due to injection trauma, the need for a more saline environment, or another issue such as contamination. mRNA-injected embryos were imaged at 1 day post injection. We did not observe any surviving embryos at 2dpi for the injection trials performed.

3.7 – References, Chapter 3

Abzhanov, A., and Kaufman, T. (2000). Crustacean (malacostracan) Hox genes and the evolution of the arthropod trunk. Development.

Andersson, E. (1969). Life-cycle and Growth of Aselus Aquaticus (L.) with Special Reference to the Effects of Temperature (Institute of Freshwater Research Drottningholm).

Browne, W.E., Price, A.L., Gerberding, M., and Patel, N.H. (2005). Stages of embryonic development in the amphipod crustacean, Parhyale hawaiensis. Genesis *42*, 124–149. https://doi.org/10.1002/gene.20145.

Bruce, H.S., and Patel, N.H. (2021). The Daphnia carapace and the origin of novel structures. BioRxiv 2021.09.29.462403. https://doi.org/10.1101/2021.09.29.462403.

Bruce, H., Jerz, G., Kelly, S., McCarthy, J., Pomerantz, A., Senevirathne, G., Sherrard, A., Sun, D., Wolff, C., and Patel, N. (2021). Hybridization Chain Reaction (HCR) In Situ Protocol. Protocols.Io.

Dohle, W., and Scholtz, G. (1997). How far does cell lineage influence cell fate specification in crustacean embryos? Semin. Cell Dev. Biol. *8*, 379–390. https://doi.org/10.1006/scdb.1997.0162.

Dohrn, A. (1867). Die embryonale Entwicklung des Asellus aquaticus (Wilhelm Engelmann).

Doudna, J.A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096. https://doi.org/10.1126/science.1258096.

Gross, J.B., Sun, D.A., Carlson, B.M., Brodo-Abo, S., and Protas, M.E. (2019). Developmental Transcriptomic Analysis of the Cave-Dwelling Crustacean, Asellus aquaticus. Genes *11*, 42. https://doi.org/10.3390/genes11010042.

Hutvágner, G., and Zamore, P.D. (2002). RNAi: nature abhors a double-strand. *12*, 225–232. https://doi.org/10.1016/S0959-437X(02)00290-3.

Jirikowski, G.J., Richter, S., and Wolff, C. (2013). Myogenesis of Malacostraca – the "egg-nauplius" concept revisited. Front. Zool. *10*, 76. https://doi.org/10.1186/1742-9994-10-76.

Jirikowski, G.J., Wolff, C., and Richter, S. (2015). Evolution of eumalacostracan development—new insights into loss and reacquisition of larval stages revealed by heterochrony analysis. EvoDevo *6*, 4. https://doi.org/10.1186/2041-9139-6-4.

Lafuente, E., Lürig, M.D., Rövekamp, M., Matthews, B., Buser, C., Vorburger, C., and Räsänen, K. (2021). Building on 150 Years of Knowledge: The Freshwater Isopod Asellus aquaticus as an Integrative Eco-Evolutionary Model System. Front. Ecol. Evol. *9*. https://doi.org/10.3389/fevo.2021.748212.

Linnaeus, C. (1758). Systema naturae (Stockholm Laurentii Salvii).

Martin, A., Serano, J.M., Jarvis, E., Bruce, H.S., Wang, J., Ray, S., Barker, C.A., C, O., Liam, and Patel, N.H. (2015). CRISPR/Cas9 Mutagenesis Reveals Versatile Roles of Hox Genes in Crustacean Limb Specification and Evolution. Curr Biol *26*, 14–26. https://doi.org/10.1016/j.cub.2015.11.021.

McMurrich, J.P. (1895). Embryology of the isopod Crustacea. J. Morphol. *11*, 63–154. .

Milatovič, M., Kostanjšek, R., and Štrus, J. (2010). Ontogenetic Development of Porcellio Scaber: Staging Based on Microscopic Anatomy. J. Crustac. Biol. *30*, 225–235. https://doi.org/10.1651/09-3189.1.

Mittmann, B., Ungerer, P., Klann, M., Stollewerk, A., and Wolff, C. (2014). Development and staging

of the water flea Daphnia magna (Straus, 1820; Cladocera, Daphniidae) based on morphological landmarks. EvoDevo *5*, 12. https://doi.org/10.1186/2041-9139-5-12.

Mojaddidi, H., Fernandez, F.E., Erickson, P.A., and Protas, M.E. (2018). Embryonic origin and genetic basis of cave associated phenotypes in the isopod crustacean Asellus aquaticus. Sci. Rep. *8*, 16589. https://doi.org/10.1038/s41598-018-34405-8.

Murphy, P.M., and Learner, M.A. (1982). The life history and production of Asellus aquaticus (Crustacea: Isopoda) in the River Ely, South Wales. Freshw. Biol. *12*, 435–444. .

Økland, K.A. (1978). Life history and growth of Asellus aquaticus (L.) in relation to environment in a eutrophic lake in Norway. Hydrobiologia *59*, 243–259. .

Panganiban, G., Sebring, A., Nagy, L., and Carroll, S. (1995). The Development of Crustacean Limbs and the Evolution of Arthropods. Science *270*, 1363–1366. https://doi.org/10.1126/science.270.5240.1363.

Poore, G.C.B. (2002). Superorder: Peracarida Calman, 1904. In Zoological Catalogue of Australia, (CSIRO Publishing), pp. 24–25.

Poore, G.C.B. (2005). Peracarida: monophyly, relationships and evolutionary success. Nauplius *13*, 1–27. .

Protas, M.E., Trontelj, P., and Patel, N.H. (2011). Genetic basis of eye and pigment loss in the cave crustacean, Asellus aquaticus. Proc Natl Acad Sci USA *108*, 5702–5707. https://doi.org/10.1073/pnas.1013850108.

Rehm, E.J., Hannibal, R.L., Chaw, R.C., Vargas-Vila, M.A., and Patel, N.H. (2009). Injection of Parhyale hawaiensis Blastomeres with Fluorescently Labeled Tracers. Cold Spring Harb. Protoc. *2009*, pdb.prot5128. https://doi.org/10.1101/pdb.prot5128.

Schwentner, M., Richter, S., Rogers, D.C., and Giribet, G. (2018). Tetraconatan phylogeny with special focus on Malacostraca and Branchiopoda: highlighting the strength of taxon-specific matrices in phylogenomics. Proc. R. Soc. B Biol. Sci. *285*. https://doi.org/10.1098/rspb.2018.1524.

Shiga, Y., Kato, Y., Aragane-Nomura, Y., Haraguchi, T., Saridaki, T., Watanabe, H., Iguchi, T., Yamagata, H., and Averof, M. (2017). Repeated co-option of a conserved gene regulatory module underpins the evolution of the crustacean carapace, insect wings and other flat outgrowths. BioRxiv 160010. https://doi.org/10.1101/160010.

Stahl, B.A., Gross, J.B., Speiser, D.I., Oakley, T.H., Patel, N.H., Gould, D.B., and Protas, M.E. (2015). A Transcriptomic Analysis of Cave, Surface, and Hybrid Isopod Crustaceans of the Species Asellus aquaticus. PLOS ONE *10*, e0140484. https://doi.org/10.1371/journal.pone.0140484.

Surbida, K., and Wright, J.C. (2001). Embryo Tolerance and Maternal Control of the Marsupial Environment in Armadillidium vulgare (Isopoda: Oniscidea). Physiol. Biochem. Zool. *74*, 894–906. https://doi.org/10.1086/324474.

Verovnik, R., and Konec, M. (2019). Asellus aquaticus: A model system for historical biogeography. In Encyclopedia of Caves, (Elsevier), pp. 76–84.

Vick, P., and Blum, M. (2010). The isopod Asellus aquaticus: a novel arthropod model organism to study evolution of segment identity and patterning. Palaeodiversity.

Vick, P., Schweickert, A., and Blum, M. (2009). Cloning and expression analysis of the homeobox gene abdominal-A in the isopod Asellus aquaticus.

Wege, W. (1909). Über die Insertionsweise der Arthropodenmuskeln nach Beobachtungen an Asellus aquaticus. Zool Anz. Leipz. *35*, 124–129. .

Wilson, G.D. (2009). The phylogenetic position of the Isopoda in the Peracarida (Crustacea: Malacostraca). Arthropod Syst. Phylogeny *67*, 159–198. .

Wolff, C. (2009). The embryonic development of the malacostracan crustacean Porcellio scaber (Isopoda, Oniscidea). Dev. Genes Evol. *219*, 545–564. https://doi.org/10.1007/s00427-010-0316-6.

Zhang, L., Hastings, M.H., Green, E.W., Tauber, E., Sladek, M., Webster, S.G., Kyriacou, C.P., and Wilcockson, D.C. (2013). Dissociation of Circadian and Circatidal Timekeeping in the Marine Crustacean Eurydice pulchra. Curr. Biol. *23*, 1863–1873. https://doi.org/10.1016/j.cub.2013.08.038.

Zuelzer, M. (1908). Über den Einfluß der Regeneration auf die Wachstums. Wilhelm RouxArchiv Für Entwicklungsmechanik Org. *25*, 361. .

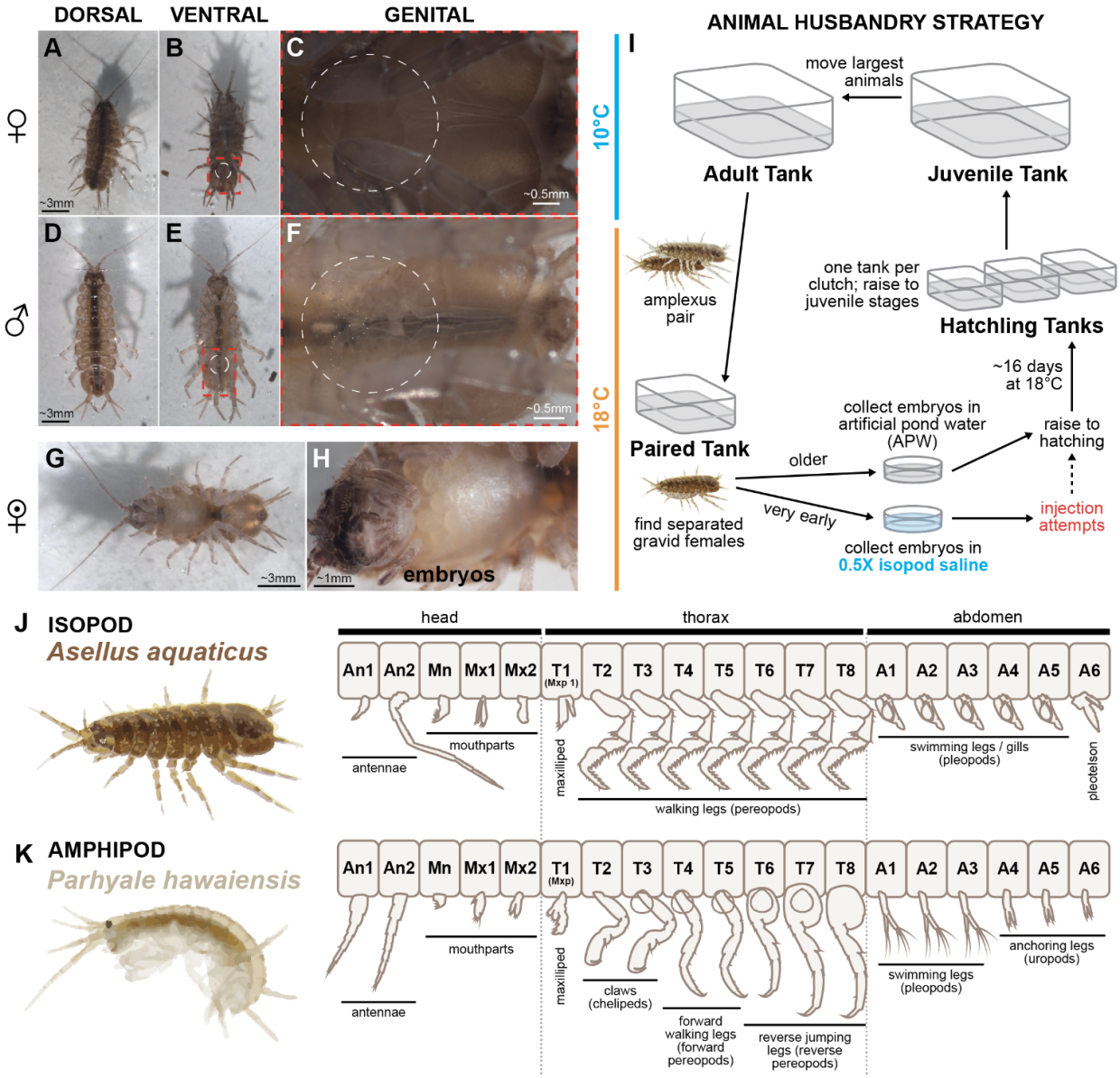


Figure 3.1: Animal husbandry and body plan of *Asellus aquaticus*

Scale bars are approximate.

A-C) *Asellus* adult female. Females tend to be darker in coloration and have a more oval-shaped body, with a widening in the anterior around the region where the brood pouch is located.

D-F) *Asellus* adult male. Males tend to be lighter in coloration and around 1.5-2x the size of females. Males also tend to have a more triangular body shape, with the width of the body increasing along the anterior-posterior axis. In the posterior of the animal, around the juncture between the thorax and abdomen, a genital structure can be found.

G-H) Gravid *Asellus* female, ventral view. Developing embryos are clearly visible through the transparent brood pouch.

I) Optimal animal husbandry strategy for obtaining early embryos. Adults are raised at 10°C. Pairs of adults in amplexus are collected regularly and moved into individual paired tanks, which are kept at 18°C. Frequently after temperature shift, mating occurs. Gravid females are isolated and anesthetized in artificial pond water (APW) with 0.2% clove oil. Embryos that appear to be at very early stages of development are collected in 0.5X isopod saline, while embryos at later stages are collected in APW. Very early stage embryos are used for injection attempts. Older embryos are raised to hatching, which takes about 16 days at 18°C. Hatchlings are raised in individual tanks per clutch of embryos to protect

against cannibalism. Once animals appear light brown in coloration, they are transferred to large juvenile tanks. Large, mature-looking animals are transferred from the juvenile tank to the adult tank at regular intervals after manual inspection of tanks every 3-6 weeks.

J) Body plan of *Asellus aquaticus*. Major differences between *Asellus* and other isopods in comparison to amphipods such as *Parhyale hawaiensis* include a lack of claws (chelipeds) and reverse jumping legs (reverse pereopods) in the thorax, as well as a difference in the number of pleopods in the abdomen.

K) Body plan of *Parhyale hawaiensis*.

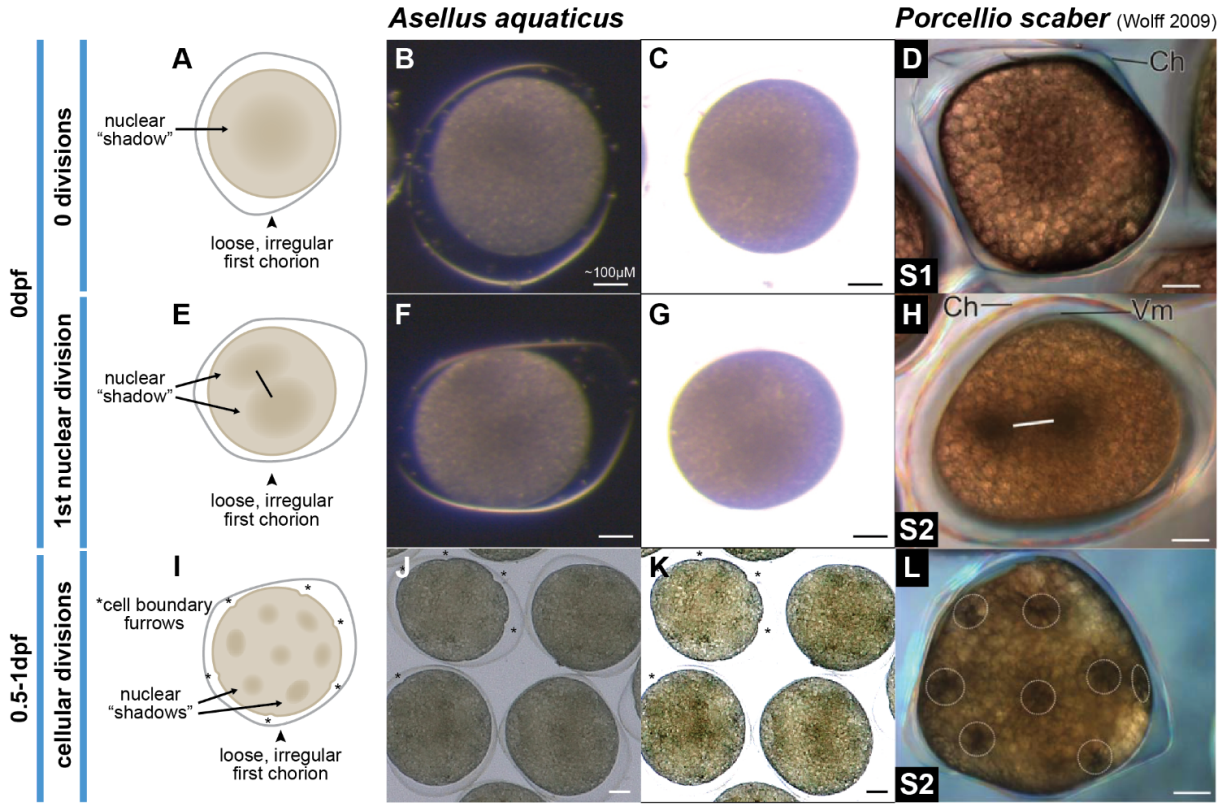


Figure 3.2: Earliest stages of *Asellus* embryogenesis

Scale bars represent approximately 100μM.

A) Schematic illustration of *Asellus* 1-nucleus embryo. A nuclear "shadow" can be observed in transmitted and oblique light microscopy.

B-C) Angled (B) and transmitted (C) light microscopy of a 1-nucleus (0dpf) embryo.

D) *Porcellio scaber* embryo (image from Wolff 2009) showing an analogous early stage of development as in A-C with a visible nuclear shadow.

E) Schematic illustration of *Asellus* 2-nucleus embryo. Two nuclear "shadows" can be observed.

F-G) Angled (F) and transmitted (G) light microscopy of a 2-nucleus (0dpf) embryo.

H) *Porcellio scaber* embryo (image from Wolff 2009) showing an analogous early stage of development as in E-G with a visible pair of nuclear shadows.

I) Schematic illustration of *Asellus* embryo at 0.5-1dpf. Multiple nuclear "shadows" are visible. Asterisks mark cell boundary furrows observed at this stage.

J-K) Transmitted light microscopy of a 2-nucleus (0dpf) embryo. (K) is a contrast-enhanced view of J. Asterisks mark apparent furrows between cells at this stage, suggesting that cellularization is occurring.

L) *Porcellio scaber* embryo (image from Wolff 2009) showing an analogous early stage of development as in I-K with multiple condensations.

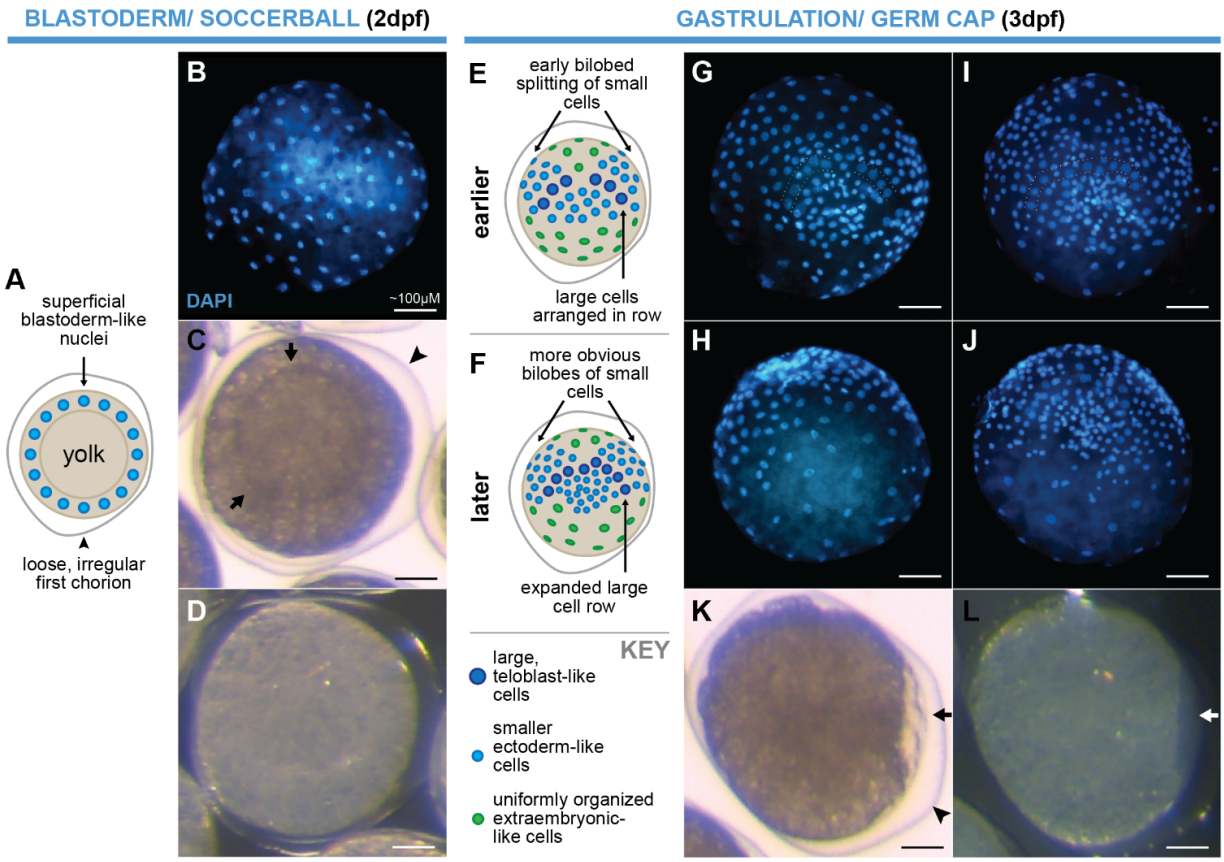


Figure 3.3: "Soccerball" and gastrulation stages

Scale bars represent approximately 100μM.

A) Schematic illustration of *Asellus* "soccerball" embryo.

B) DAPI stain of fixed 2dpf embryo.

C-D) Transmitted (C) and ambient (D) light microscopy of 2dpf embryo. In the transmitted image, a circle (boundaries marked by arrows) is visible within the embryo, suggesting a partition between the outer layer of cells and interior yolk. Magenta arrow marks visible boundary of the vitelline membrane (VM).

E-F) Schematic illustration of *Asellus* embryos at stages analogous to gastrulation stages found in *Porcellio*. Three populations of cells can be observed at this stage: small cells forming the bilobed structures and condensed "posterior" structure (light blue); large cells forming a cell row (light blue); and uniformly scattered smaller cells, which in other malacostracans (*Porcellio scaber*, *Parhyale hawaiensis*, *Orchestia cavimana*) form extraembryonic structures.

G-H) Two views of the same embryo at early stages of gastrulation. Cells have begun to form bilobed structures in the "anterior" and a row of large cells is beginning to form (dotted lines). A lateral view (H) shows that cells appear to be condensing at one end of the embryo. In *Porcellio*, cells begin migrating internally at this stage to begin the gastrulation process.

I-J) Two views of the same embryo at slightly later stages of gastrulation. The two bilobed structures are more obvious in the "anterior", and a full row of large cells is formed (dotted lines). The cell condensation in the "posterior" appears more diffuse, suggesting that most of the gastrulation process may have completed by this point.

K-L) Transmitted (K) and ambient (L) light microscopy of 3dpf embryo. At one end of the embryo, a slight transparent region can be observed, which is likely the position where cells are condensing during the gastrulation process.

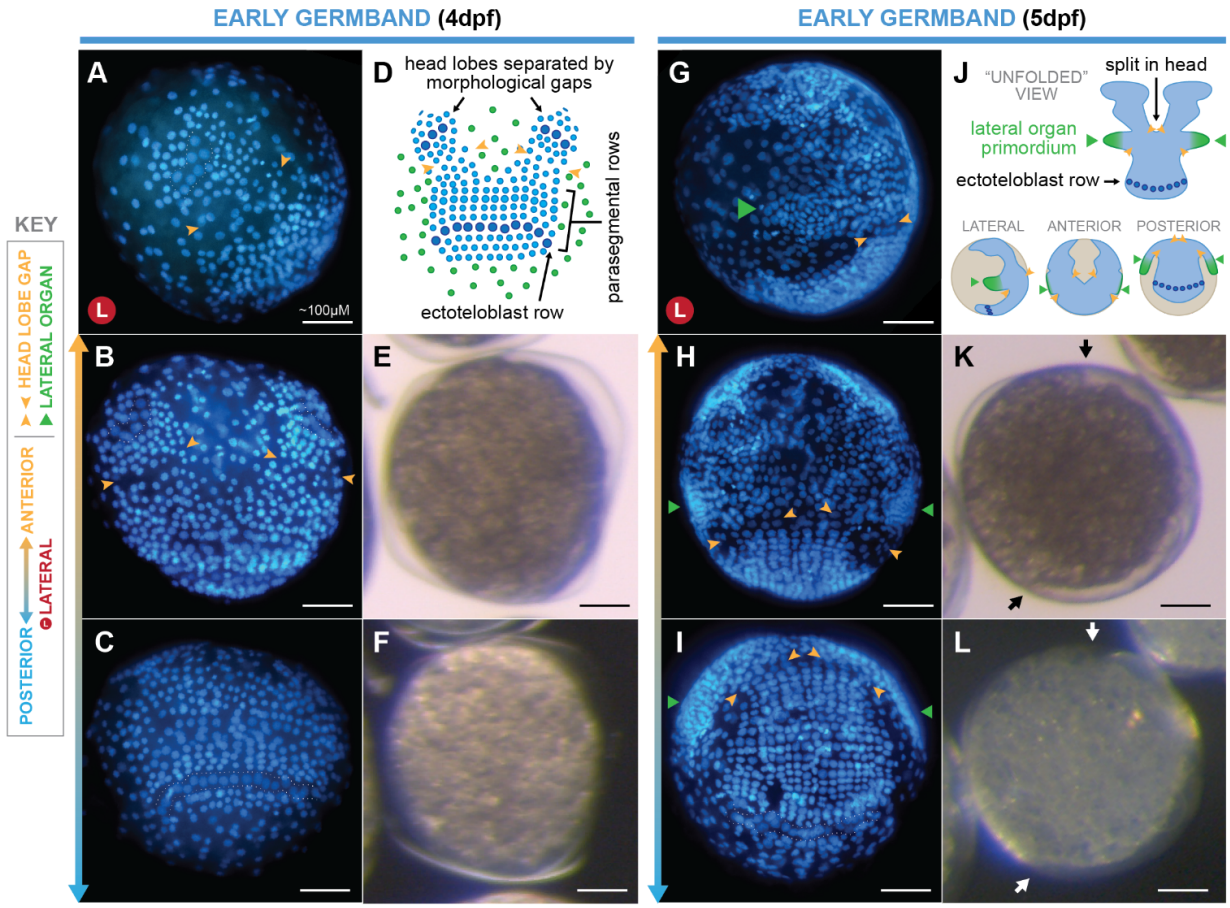


Figure 3.4: Early germband and lateral organ bud stages

Scale bars represent approximately 100μM.

A-C) Three views of the same 4dpf embryo: lateral (A), anterior (B), posterior (C). The embryo has developed an obvious pair of head lobes, analogous to those observed in other crustaceans. These head lobes are separated by a large morphological gap with the trunk (yellow arrows in A and B). Within each head lobe is a population of large cells (dotted lines) which resemble the ectoteloblasts in the posterior. In the posterior, the ectoteloblast row is well-formed (demarcated by dotted lines in C). In the posterior, the parasegment rows are beginning to form.

D) Schematic view of 4dpf embryo showing bilobed head structures separated by a morphological gap; uniformly distributed extraembryonic cells; and ectoteloblast row.

E-F) Transmitted (E) and ambient (F) light microscopy of 4dpf embryo. The embryo is a very thin sheet at this stage, and no obvious structures can be observed aside from a mottled yolk.

G-I) Three views of the same 5dpf embryo: lateral (G), anterior (H), posterior (I). The embryo has developed obvious parasegmental rows, as well as anterior lateral outgrowths. In (G), an outgrowth of cells anterior to the gap between head and trunk regions is observed (green triangle marks outgrowth, yellow arrows mark head-trunk gap). The cells in this region appear to bud off from the embryonic ectoderm and condense in a radial pattern. At this position in later stages, the lateral organ emerges. This structure is thus likely the lateral organ bud. In the posterior (I), the ectodermal germband grid is highly organized and obvious. The ectoteloblast row persists in the posterior (dotted lines).

J) Schematic view of 5dpf embryo showing bilobed head structures separated by a split in the head, as well as a gap between the head and trunk structures (yellow arrows), the ectoteloblast row (large blue nuclei), and the lateral organ bud (green arrow). An "unfolded" view, as well as lateral, anterior, and posterior schematics are shown.

K-L) Transmitted (K) and ambient (L) light microscopy of 4dpf embryo. The wrapping of the embryo around the yolk can be observed as a thin region of transparency, marked in the anterior and posterior by arrows.

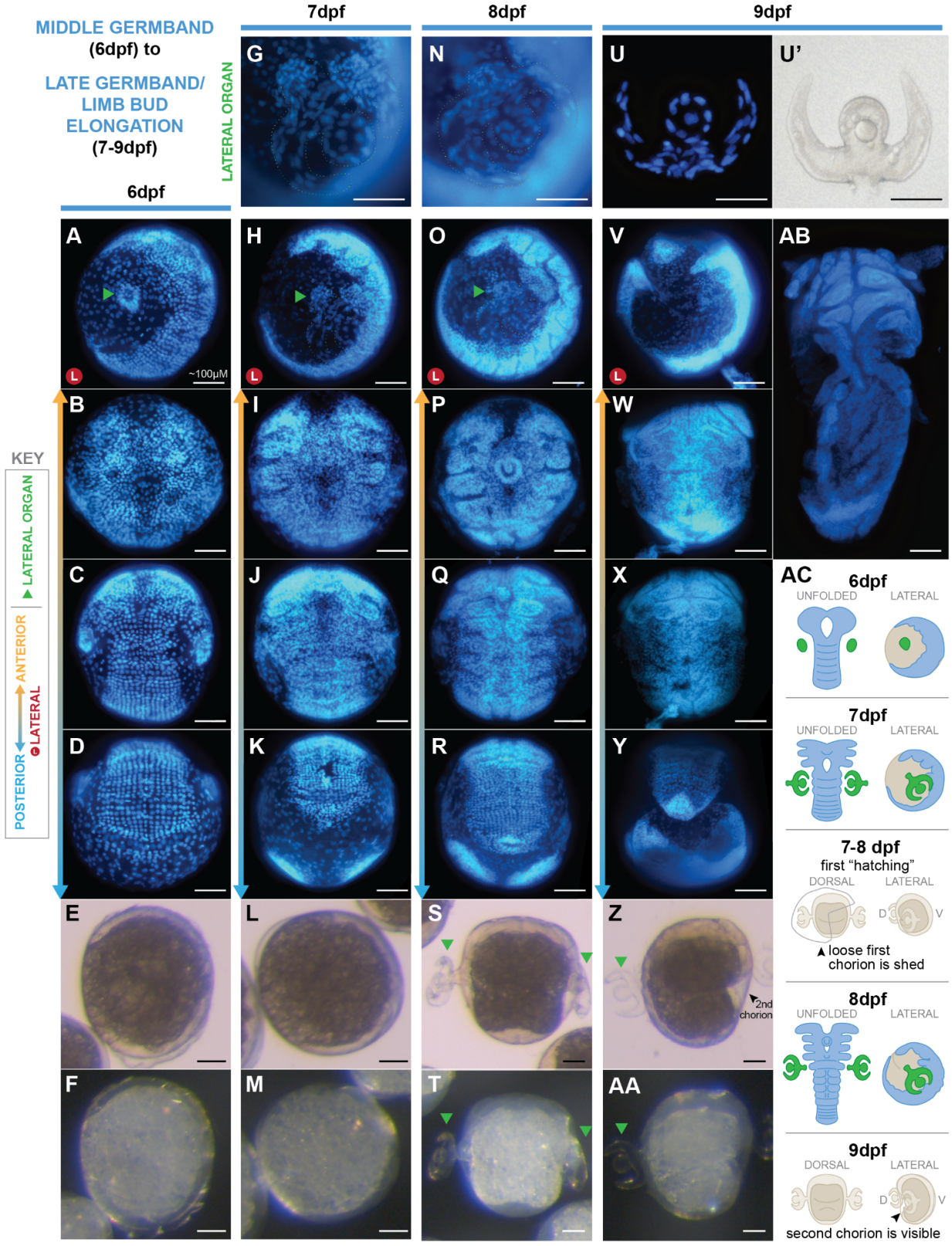

MIDDLE GERMBAND
(6dpf) to
LATE GERMBAND/
LIMB BUD
ELONGATION
(7-9dpf)
6dpf
7dpf
8dpf
9dpf
LATERAL ORGAN
KEY
LATERAL ORGAN
ANTERIOR
LATERAL
POSTERIOR
~100µM
AC
6dpf
UNFOLDED
LATERAL
7dpf
UNFOLDED
LATERAL
7-8 dpf
first "hatching"
DORSAL
LATERAL
loose first
chorion is shed
8dpf
UNFOLDED
LATERAL
9dpf
DORSAL
LATERAL
second chorion is visible
2nd
chorion

Figure 3.5: Mid germband to late germband and limb bud elongation stages.

Scale bars represent approximately 100μM.

A-D) Four views of the same 6dpf embryo: lateral (A), and three rotations along the anterior-posterior axis (B-D). In a lateral view, the developing lateral organ is visible as a circular condensation of cells (A, green arrow). Additionally, the segments of the body are becoming more obvious through indentations along the AP grid of the trunk. In the anterior (B), the head lobes have joined at the anterior-most region, resulting in a "gap" where the mouth and gut tube will likely later form. At the juncture of the head and trunk regions (C), the morphological gap observed at previous stages is less obvious. The condensations to the left and right sides of the trunk are the lateral organ buds (marked with a green triangle in A). The curvature of the cell regions anterior of the lateral organ buds indicates their origin anterior to the head-trunk gap. In the posterior (D), the ectoderm germband grid continues to elongate, and the ectoteloblasts are still visible.

E-F) Transmitted (E) and ambient (F) light microscopy of 6dpf embryo. The wrapping of the embryo around the yolk can be observed as a thin region of transparency, as at the 5dpf stage.

G) Zoom of the lateral organ from a lateral view at 7dpf. The trident shape of the lateral organ is fully developed at this stage, although the lateral organ itself remains underneath the first false "ecdysis" membrane.

H-K) Four views of the same 7dpf embryo: lateral (H), and three rotations along the anterior-posterior axis (I-K). In the anterior, the antennal and mandibular limb buds appear to be formed. The ectoteloblast row remains visible in the posterior.

L-M) Transmitted (L) and ambient (M) light microscopy of 7dpf embryo.

N) Zoom of the lateral organ from a lateral view at 8dpf. The lateral organ is released from the first "ecdysis" at this stage. However, it does not appear to undergo any change in cell number or morphology from the 7dpf stage.

O-R) Four views of the same 8dpf embryo: lateral (O), and three rotations along the anterior-posterior axis (P-R). In the anterior, a U-shaped pucker forms at the presumptive labrum. The thoracic limb buds begin to appear, and grow inwards towards the midline, in contrast to the head appendages, which grow outward. The ectoteloblast row is no longer visible in the posterior of the embryo.

S-T) Transmitted (S) and ambient (T) light microscopy of 8dpf embryo. Between 7-8dpf, most embryos undergo a first false "ecdysis," which reveals the lateral organs. We referred to this stage colloquially as the "snitch" stage.

U-U') Dissected lateral organ at 9dpf in brightfield and DAPI views. The trident shape of the organ is preserved across developmental stages. Comparisons of the gross nuclear morphology of the lateral organ across 7-9dpf suggest that the organ appears to be fully formed by 7dpf and does not undergo additional cell divisions or morphological changes.

V-Y) Four views of the same 9dpf embryo: lateral (V), and three rotations along the anterior-posterior axis (W-Y). Staining of embryos at this stage was not uniform, even after long incubation times in DAPI-PT, potentially due to the numerous membranes surrounding the embryo by this stage. Across the body axis, limb buds are visible and elongating.

Z-AA) Transmitted (Z) and ambient (AA) light microscopy of 9dpf embryo. The dorsal portion of the embryo is beginning to unfold, resulting in a growing separation of the developing head and abdomen and an elongation of the body axis. A gap between the membrane and the unfolding dorsal side of the embryo is visible.

AB) Unfolded 9dpf embryo, ventral view, showing elongated limbs. Due to the curvature of the embryo and the large number of membranes at this stage, mounting, dissection, and staining of the embryo became considerably more challenging.

AC) Schematic illustration of days 6-9dpf. The embryo closes in the anterior, forming a hole where the gut tube presumably forms. The lateral organs develop and are released. Segmentation progresses and completes, and the limb buds begin to grow and elongate. The embryo also undergoes its first false "ecdysis."

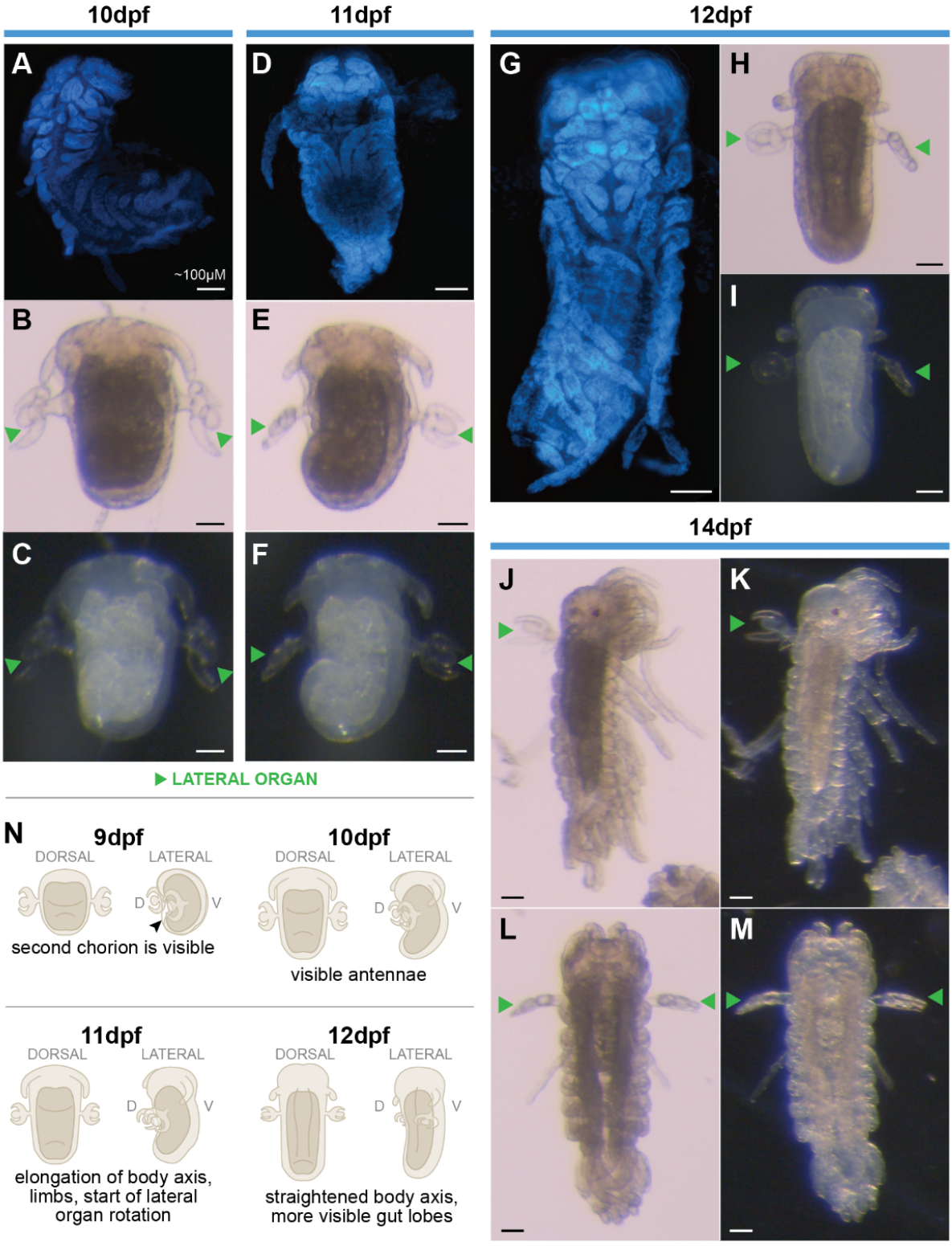

10dpf
11dpf
12dpf
A
B
C
D
E
F
G
H
I
~100µM
14dpf
J
K
L
M
▶ LATERAL ORGAN
N
9dpf
DORSAL
LATERAL
D
V
second chorion is visible
10dpf
DORSAL
LATERAL
D
V
visible antennae
11dpf
DORSAL
LATERAL
D
V
elongation of body axis, limbs, start of lateral organ rotation
12dpf
DORSAL
LATERAL
D
V
straightened body axis, more visible gut lobes

Figure 3.6: Late limb elongation stages to "angel" stage

Scale bars represent approximately 100μM.

A) DAPI stain of a 10dpf embryo.

B-C) Transmitted (B) and ambient (C) light microscopy of 10dpf embryo. Around the 9-10dpf stage, the embryo undergoes a second false "ecdysis" and reveals a pair of antennae. The embryo begins to elongate further, and the lateral organ begins to rotate towards its final position, with the trident pointing outward to the sides of the embryo.

D) DAPI stain of an 11dpf embryo.

E-F) Transmitted (E) and ambient (F) light microscopy of 11dpf embryo. The body axis elongates further at this stage, and the lateral organs continue their rotation.

G) DAPI stain of a 12dpf embryo.

H-I) Transmitted (H) and ambient (I) light microscopy of 12dpf embryo. The anterior-most portions of the head of the embryo further develop at this stage, such that the antennae no longer appear to be the anterior-most structures, but instead flank the left and right sides of the "neck" of the embryo. The lateral organs have nearly completed their rotation. Through the yolk, it is possible to see the formation of two parallel gut tubes running down the anterior-posterior axis.

J-M) Transmitted (J, L) and ambient (K, M) light microscopy of 12dpf embryo from dorsal and lateral views. We referred to embryos at this stage as "angel" embryos. Around 12-14dpf, the embryo undergoes its third false "ecdysis" to reveal all of the embryonic limbs. The animal resembles a hatchling or manca at this stage, with the exception of the obvious "wings" that are the lateral organs. The lateral organs are shed with the final, fourth "ecdysis" which is when the embryo truly hatches.

N) Schematic illustration of later developmental stages, indicating those features that most distinguish the different stages.

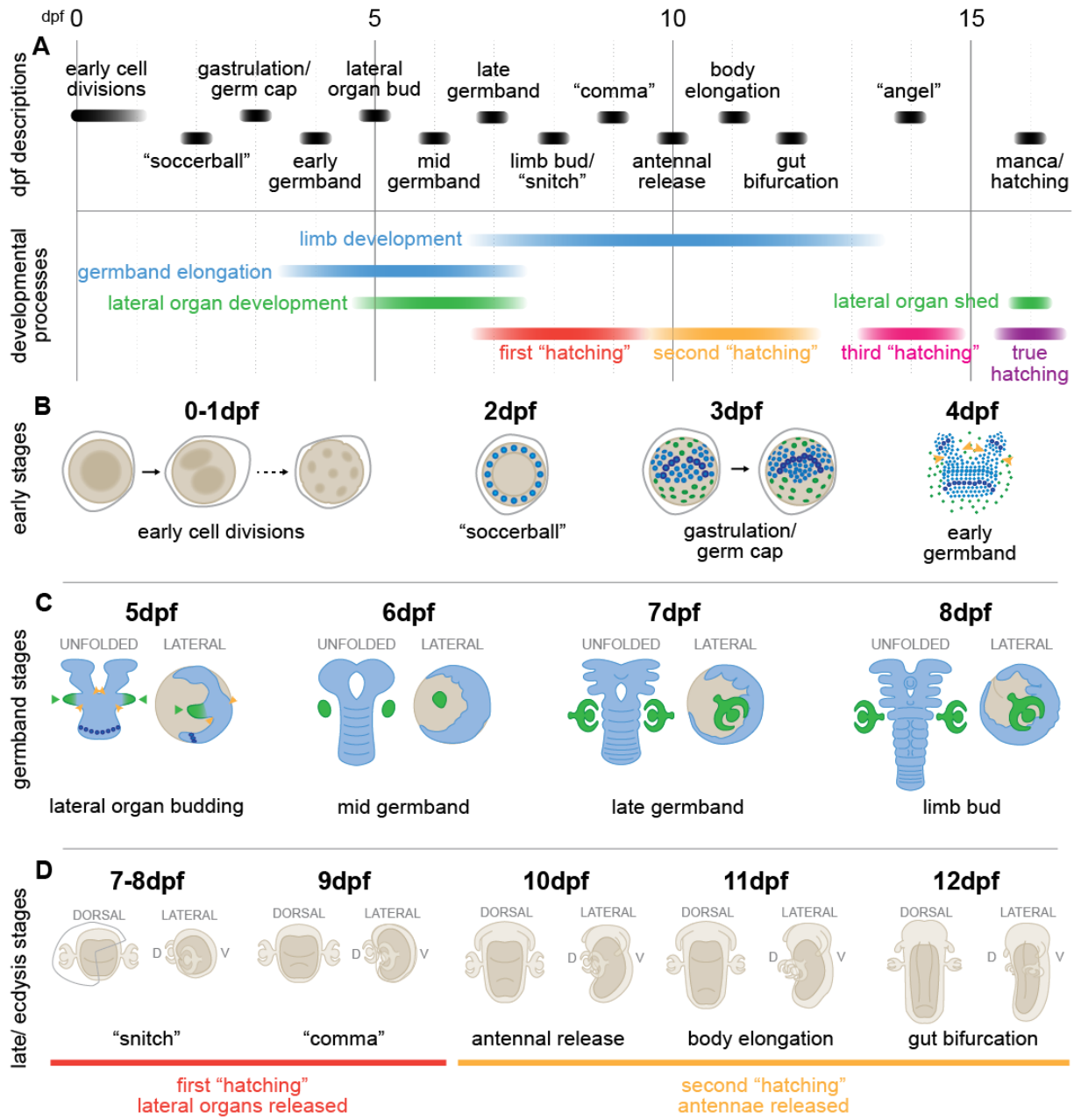


Figure 3.7: Schematic timeline of *Asellus* development

A) Developmental timeline for Asellus, indicating the most prominent features or processes for most of the ~16 days of development at 18°C (top row) and more general developmental processes (bottom row).

B) Early developmental stages. From 0-4dpf, the embryo undergoes early cell divisions, cellularization, gastrulation, and the start of germband elongation.

C) Middle developmental stages. The embryo buds and develops its lateral organs, forms limb buds, closes the gap between the anterior head lobes, and undergoes its first "ecdysis."

D) Late developmental stages. The embryo undergoes multiple false "ecdyses" and the anterior and posterior halves of the embryo unfold as the embryo elongates. The lateral organ is released and rotates. The gut tubes emerge at late stages. The limbs progressively elongate and develop.

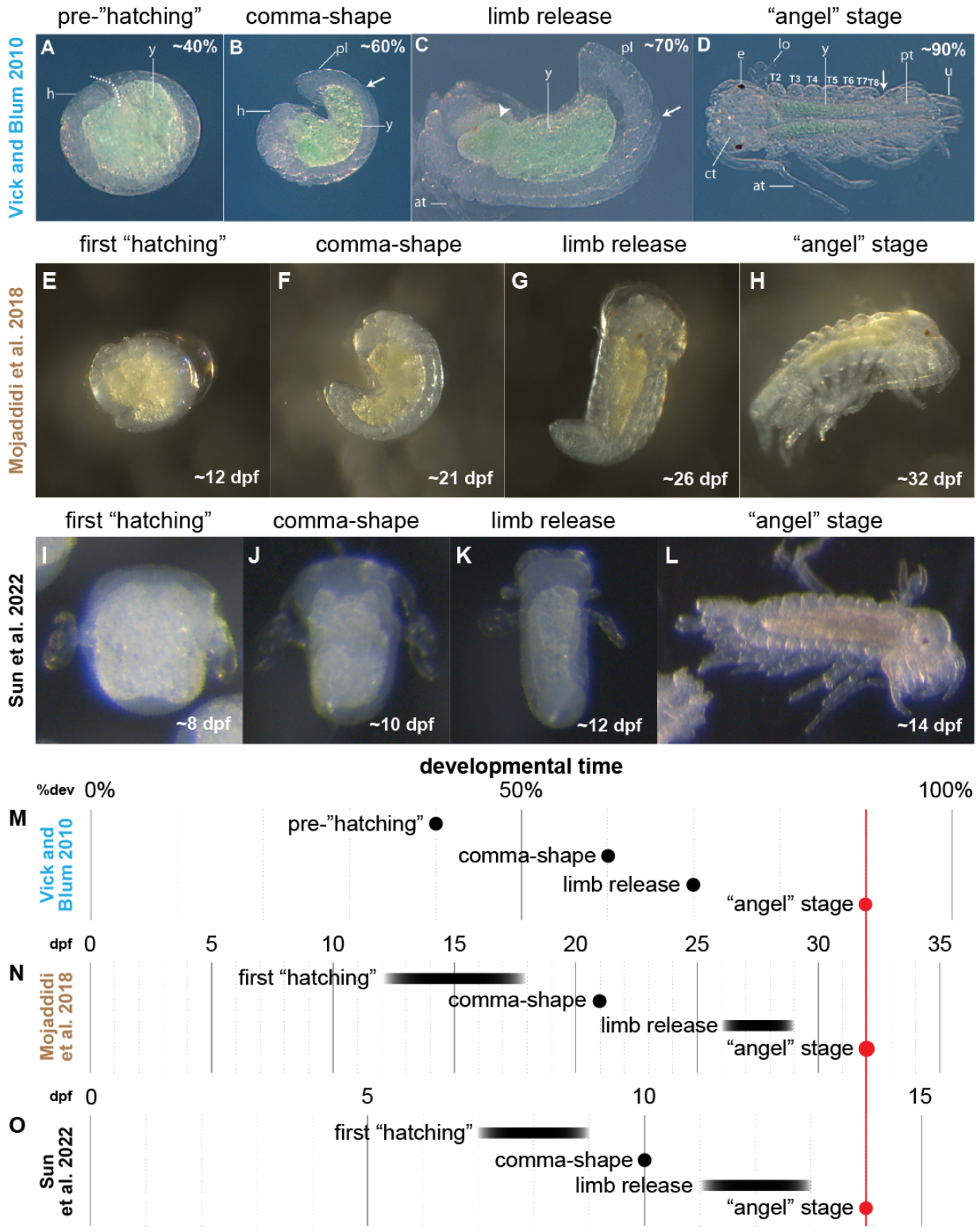

pre-"hatching"
comma-shape
limb release
"angel" stage
Vick and Blum 2010
A
~40%
y
h
B
~60%
pl
h
y
C
~70%
pl
y
at
D
~90%
e
lo
y
pt
u
T2 T3 T4 T5 T6 T7 T8
ct
at
first "hatching"
comma-shape
limb release
"angel" stage
Mojaddidi et al. 2018
E
~12 dpf
F
~21 dpf
G
~26 dpf
H
~32 dpf
first "hatching"
comma-shape
limb release
"angel" stage
Sun et al. 2022
I
~8 dpf
J
~10 dpf
K
~12 dpf
L
~14 dpf
developmental time
%dev
0%
50%
100%
M
Vick and Blum 2010
pre-"hatching"
comma-shape
limb release
"angel" stage
dpf
0
5
10
15
20
25
30
35
N
Mojaddidi et al. 2018
first "hatching"
comma-shape
limb release
"angel" stage
dpf
0
5
10
15
O
Sun et al. 2022
first "hatching"
comma-shape
limb release
"angel" stage

Figure 3.8: Comparison of developmental timelines to previous studies.
A-D) Images from Vick and Blum, 2010 showing four developmental stages. The authors did not report the precise timings of these images, but instead reported percentages of developmental time.
E-H) Images from Mojaddidi et al. 2018, showing four developmental stages. The ages shown are approximate.
I-J) Comparable stages to (A-H) from the current study.
M-O) Developmental time for the three studies examined, anchored by the timing of the "angel" stage. Each time bar has been scaled so that 0dpf and 8dpf occupy the same horizontal space in the figure. In the Mojaddidi et al. study, development took around twice as long as in the present study, likely owing to the colder temperatures used to raise animals in that study.

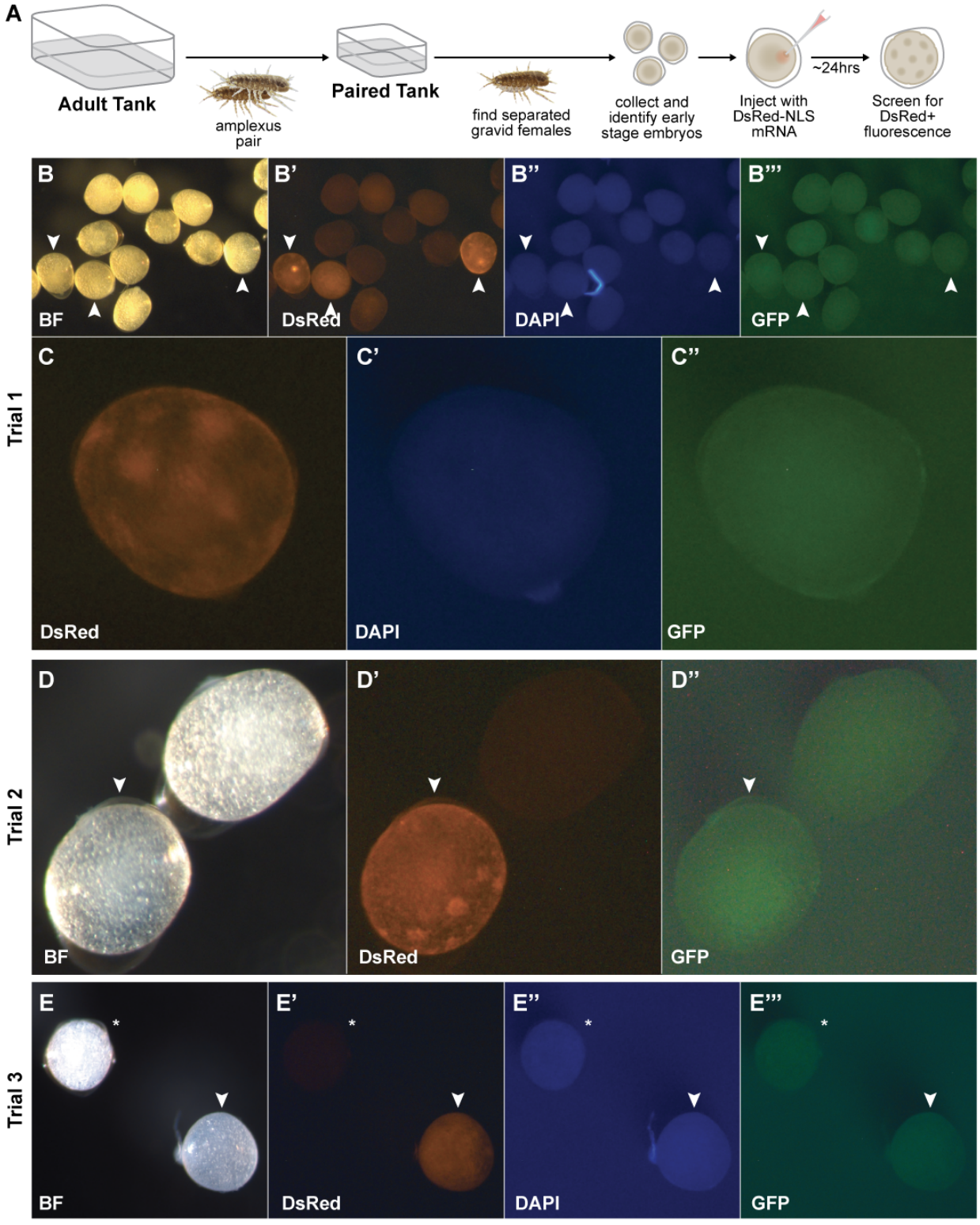

A
Adult Tank
amplexus pair
Paired Tank
find separated gravid females
collect and identify early stage embryos
Inject with DsRed-NLS mRNA
~24hrs
Screen for DsRed+ fluorescence
B
BF
B'
DsRed
B''
DAPI
B'''
GFP
Trial 1
C
DsRed
C'
DAPI
C''
GFP
Trial 2
D
BF
D'
DsRed
D''
GFP
Trial 3
E
BF
E'
DsRed
E''
DAPI
E'''
GFP

Figure 3.9: Trial injections of DsRed mRNA in *Asellus* embryos.

A) Schematic of injection approach. Embryos were collected from females that appeared to have mated within the last 12 hours. Injected embryos showed the nuclear "shadow" morphology. Selected embryos were injected with 1μg/μL capped DsRed mRNA + 0.05% phenol red and raised for 1 day, then imaged.

B-B''') Multiple brightfield and fluorescence views of embryos at 1dpf. In the DsRed channel, apparent fluorescence can be observed which is distinct from any autofluorescent signal found in the DAPI or GFP channels. Not all embryos appear to show this fluorescence, as is consistent with other mRNA injection experiments such as those in *Parhyale*.

C-C'') View of a single embryo from the first trial experiment that appeared to show multiple diffuse red fluorescent foci, suggestive of the multiple nuclear shadows observed at around 1dpf in uninjected embryos.

D-D'') Views of two embryos from the second trial experiment. Arrow marks an embryo at 1dpi that shows more condensed apparent nuclear morphology and fluorescence.

E-E''') Views of two embryos from the third trial experiment. Arrow marks an embryo at 1dpi that shows diffuse background fluorescence, while asterisk marks an embryo that is dead, as indicated by its highly opaque yolk. The apparent fluorescence observed in these injection experiments is not likely to be driven by dead embryos, as evidenced by the lack of strong red fluorescence in the dead embryo. The absence of obvious nuclear staining in the fluorescent embryo may be a result of capturing an image as the embryo is in the midst of a synchronous cell division, as has previously been shown to occur in *Porcellio*.

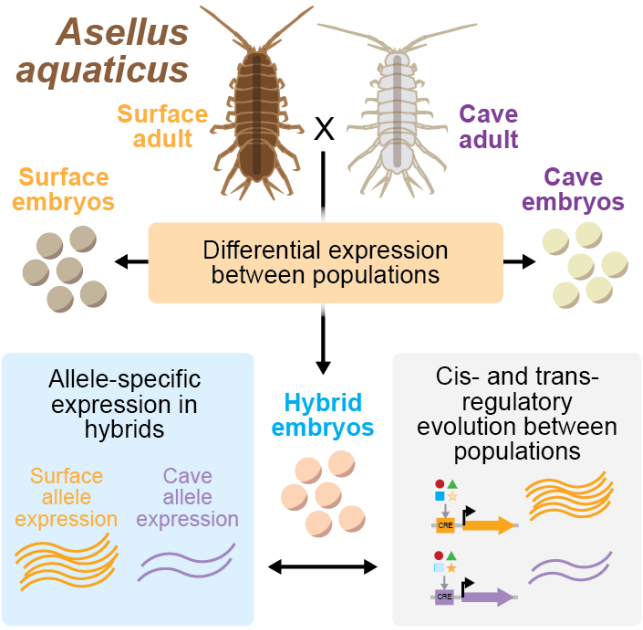
4 – Developmental transcriptomic analysis of the cave-dwelling crustacean, *Asellus aquaticus*

Joshua B. Gross[1,†], Dennis A. Sun[2,†], Brian M. Carlson[3], Sivan Brodo-Abo[4], and Meredith E. Protas[5,*]
1. Department of Molecular and Cell Biology, University of California, Berkeley
2. Department of Marine Biology and Biological Oceanography, University of Rhode Island
3. Dominican University of California, San Rafael, CA
4. Marine Biological Laboratory, Woods Hole, MA
5. Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL
† equal contribution
* corresponding author

Previously published in Genes (Basel)

Abstract

Cave animals are a fascinating group of species often demonstrating characteristics including reduced eyes and pigmentation, metabolic efficiency, and enhanced sensory systems. Asellus aquaticus, an isopod crustacean, is an emerging model for cave biology. Cave and surface forms of this species differ in many characteristics, including eye size, pigmentation, and antennal length. Existing resources for this species include a linkage map, mapped regions responsible for eye and pigmentation traits, sequenced adult transcriptomes, and comparative embryological descriptions of the surface and cave forms. Our ultimate goal is to identify genes and mutations responsible for the differences between the cave and surface forms. To advance this goal, we decided to use a transcriptomic approach. Because many of these changes first appear during embryonic development, we sequenced embryonic transcriptomes of cave, surface, and hybrid individuals at the stage when eyes and pigment become evident in the surface form. We generated a cave, a surface, a hybrid, and an integrated transcriptome to identify differentially expressed genes in the cave and surface forms. Additionally, we identified genes with allele-specific expression in hybrid individuals. These embryonic transcriptomes are an important resource to assist in our ultimate goal of determining the genetic underpinnings of the divergence between the cave and surface forms.



4.0 – Introduction

Cave animals are fascinating organisms that frequently share a common suite of characteristics, including reduced eyes, reduced pigmentation, metabolic differences, and enhanced sensory systems. Questions that have long fascinated cave biologists include how and why these characteristics have evolved and whether the same underlying mechanisms mediate trait loss between different cave populations and different cave species.

Historically, it has been challenging to understand how and why cave characteristics have evolved, due to difficulties with rearing cave organisms in captivity and a lack of contemporary experimental resources (e.g., genomic, genetic, and functional molecular tools) for most cave species. In recent years, however, there have been vast expansions of, and improvements in, resources and tools for emerging model organisms. Obtaining genomic information is now possible for most systems, and the complete genomic sequence is available for a limited number of cave-dwelling species (McGaugh et al., 2014; Yang et al., 2016). In addition, many studies have involved transcriptome sequencing projects for cave dwellers such as crayfish, salamanders, amphipods, isopods, and fish (Carlini et al., 2013; Meng et al., 2013; Passow et al., 2017; Stahl et al., 2015; Stern and Crandall, 2018a, 2018b). For the vast majority of these projects, adult samples have been utilized due to the challenge of obtaining embryonic samples of natural, cave-dwelling species. However, for several cave species, many trait differences are established early in embryonic development, underscoring the importance (and value) of analyzing gene expression differences across embryonic development.

The most widely studied cave model species is Astyanax mexicanus, where it is possible to work with embryos and obtain embryonic samples, as well as perform genetic analyses (reviewed in (Jeffery, 2005, 2009)). Both adult and embryonic transcriptomes, as well as a draft genome sequence for cave and surface morphs (Gross et al., 2013; Hinaux et al., 2013; McGaugh et al., 2014), have been generated. Additionally, contemporary genomic tools, such as gene editing, provide the ability to functionally analyze candidate genes discovered through transcriptome sequencing (Klaassen et al., 2018; Kowalko et al., 2016; Ma et al., 2015). Because of the wealth of data provided by these emerging resources, historical questions impacting on the evolution of cave animals can now be addressed (reviewed in (Casane and Rétaux, 2016; Gross et al., 2015)).

Despite the great deal of information provided by decades of research in A. mexicanus, studies from additional cave organisms are necessary to understand the convergence of regressive loss across animals that inhabit the cave biome. Specifically, the mechanisms that mediate regressive loss in A. mexicanus may differ from those mechanisms operating in other cave-adapted species. Thus, it is important to develop other species in a similar way to A. mexicanus in order to widen our perspective and to gain a broader understanding of how cave evolution occurs across diverse taxa.

Unfortunately, not every cave animal is amenable to develop as a model in the same way as A. mexicanus. There are many considerations, foremost of which is the ability to raise and breed a species in the lab. This feature greatly reduces the number of cave-adapted species for which genetic and developmental studies are feasible. Another important feature for these investigations is an extant surface-dwelling form capable of interbreeding with cave morphs. Owing to the divergence times between cave and surface morphs, the ability to produce viable hybrid offspring is very unusual among studied cave organisms.

Asellus aquaticus is a freshwater crustacean that has two morphs, a cave and surface form, both of which can be raised in the lab and can interbreed (Protas and Jeffery, 2012). Much of the historical work on Asellus aquaticus has included comparative morphology between the surface and cave forms, and population genetic analyses of several cave and surface populations throughout Europe (Pérez-Moreno et al., 2017; Turk et al., 1996; Turk-Prevorčnik and Blejec, 1998; Verovnik et al., 2003, 2004; VEROVNIK et al., 2005). Recently, a classical genetics approach has been made possible by multiple crossing strategies to create F1, F2, and backcross pedigrees between cave and surface populations. These studies have resulted in production of a linkage map, insight into the genetic architecture of this species, and identification of genomic regions associated with different cave-associated phenotypes (Mojaddidi et al., 2018; Protas et al., 2011; Re et al., 2018).

Though advances have been made in genomic mapping alongside the development of genetic resources, the identity of genes responsible for these trait differences between cave and surface forms remains unknown. A powerful approach to identifying genetic differences between cave-and surface-dwelling forms is comparative transcriptomics. Transcriptomes have been characterized for the Pivka Channel of Planina Cave population, the Molnár János Cave population, and nearby surface populations to both caves (Pérez-Moreno et al., 2018; Stahl et al., 2015). Though these studies have been useful in generating genetic resources, the causative genes mediating differences between cave and surface populations have not been established. Part of the issue, as discussed above, is that adult samples are not the most appropriate, as many different characteristics between cave and surface individuals are established during embryonic development (Mojaddidi et al., 2018; Sket, 1956). For example, eye loss and pigment loss are established by the end of embryogenesis. To investigate the genetic pathways responsible for eye and pigment loss, the most appropriate samples to sequence would be those obtained at this timepoint in embryonic development.

To address this gap in knowledge, we generated de novo embryonic transcriptomes from one cave and one surface population, as well as from hybrid individuals. We hypothesized that many genes would be differentially expressed between cave and surface forms, including those involved in neurogenesis, pigment development, eye development, and metabolism. Furthermore, we expected that a subset of these differentially expressed genes would also show allele-specific expression, suggesting that regulatory mutations result in altered transcriptional abundance for those genes.

4.1 – Results

4.1.1 – Characterization of Surface, Cave, and Hybrid Transcriptomes

Following the optimization of our assembly parameters, we retrieved highly similar results for all three of our assembled transcriptomes (Table 1; Supplementary Figure S1). The total number of reads that were assessed for each transcriptome was very similar between surface morphs (~364 M), cave morphs (~361 M), and hybrids (~394 M). The total number of assembled reads for surface (~155 M), cave (~164 M), and hybrids (~132 M) were similarly comparable, although a higher proportion of assembled reads were utilized in cave morphs (45.5%) compared to surface morphs (42.4%) and hybrids (33.5%). The reduced proportion of assembled reads used in the hybrid transcriptome assembly may reflect the sequence divergence between cave and surface morphs.

All assemblies were completed in roughly the same amount of time (50.4–54.2 h) and yielded comparable numbers of transcripts (surface = 113 K; cave = 119 K; hybrid = 143 K), or comparable average lengths (surface = 1061 bp; cave = 1069 bp; hybrid = 952 bp). Our goal, however, was to annotate the best-characterized transcripts in each dataset. We reasoned that the longest transcripts represented the best individual transcript assemblies, and therefore proceeded to annotate those assembled transcripts that were 1000 bp or longer. This value was similar across all three assemblies: surface = 49,233; cave = 51,822; hybrids = 52,390 (Table 2).

Using these assemblies as a starting point, we subjected each transcriptome to comprehensive annotation using Blast2GO (Methods). This BLAST-based approach was performed against the Tribolium castaneum genome and SwissProt database, in order to compare the quality of each database. We chose these databases because Tribolium castaneum is an arthropod with a comprehensive genome database, and the SwissProt database is an open-access and manually annotated resource of protein sequence and functional information. Overall, we found that the average percentage of failed BLAST hits was higher when we used the SwissProt database (mean = 61.3%) compared to the Tribolium castaneum database (mean = 58.4%). Consequently, our final transcriptome size was larger when we annotated against the Tribolium (mean = 19,727 transcripts) compared to the SwissProt database (mean = 18,110 transcripts). In sum, our results indicated that the Tribolium castaneum database provided better results (Table 2), and therefore our downstream analyses utilized these annotated transcriptomes.

4.1.2 – Differential RNA-Seq Analysis Between Cave and Surface Morphs

We mapped the cave and surface reads separately to each of the four different transcriptomes: cave, surface, hybrid, and integrated transcriptomes. We selected all genes that had at least a two-fold change in the same direction (increased or decreased expression) between cave and surface in all four experiments and had a standard deviation of less than or equal to 8. Then we selected the top 50 genes that were underexpressed in the cave form and the top 50 genes that were overexpressed in the cave form to analyze further (Figure 1E; Supplementary Table S1).

Several of the genes that were underexpressed in the cave made biological sense, as they are involved in eye or pigment function such as long-wavelength sensitive opsin, cell cycle control protein 50A-like, membrane-bound transcription factor site 1 protease-like protein, scarlet-like protein, protein pygopus-like, and atonal. Genes that were overexpressed in the cave form include those involved in metabolism, such as solute carrier family 35 member F6-like protein, gamma-glutamyltransferase 7-like protein, and inositol oxygenase-like protein. Also overexpressed in the cave samples was annulin-like protein, which is expressed in stripes in each limb bud segment (Bastiani et al., 1992) and could be a candidate for differential antennal characteristics in the cave form.

4.1.3 – Allele-Specific Expression Analysis in Hybrid Individuals

Genes that display differential expression between populations may arrive at this difference through both cis- and trans-regulatory mechanisms. In cis-regulatory changes to gene expression, a

change to the DNA sequence either within a gene or in regulatory elements thereof is responsible for an observed expression difference between populations (Figure 2A). When trans-regulatory factors change gene expression, the regulatory sequence of a gene may not change, but instead, a change to the expression of a trans-regulatory factor (an activator, repressor, etc.) between populations drives the difference in expression of a downstream gene. By examining the expression of alleles of a given gene in hybrid organisms, one can determine mechanisms of gene expression difference, whether they be cis-regulatory, trans-regulatory, or a combination of both. In hybrid animals, trans-regulatory effects are normalized across alleles, as both alleles existing in the same nucleus are subjected to the same input by activators and repressors. As such, when expression differences in alleles are observed in hybrids, one possible explanation is that cis-regulatory changes contribute to differential expression between populations (Figure 2A). Allele-specific differences can also come about due to parent-of-origin effects (see Discussion).

We wanted to examine the mechanism of differential gene expression for the genes we identified as differentially expressed between cave and surface populations. To do this, we performed allele-specific expression (ASE) analysis using the ASE-TIGAR software package (Nariai et al., 2016). This software, given a FASTA file containing both isoforms of a gene and FASTQ reads from hybrid animals, generates transcript abundance estimates for each allele (Figure 2B). We identified pairs of alleles for the most differentially expressed genes and generated a log fold change value for the usage of surface vs. cave alleles in hybrid animals (Figure 2). We then used a two-sample Kolmogorov–Smirnov test with a Benjamini–Hochberg multiple hypothesis testing correction to call significance of observed ASE, using intra-allele log fold change as our null distribution (see Methods; Figure 2C,D). Overall, genes with significant ASE tended to have larger log2 fold change between the two alleles (Figure 2E; Supplementary Table S1).

Many of the genes we identified as highly differentially expressed (DE) between individuals of different populations also appeared to show ASE between alleles in hybrid animals (Figure 2E, Figure 3A,B; Supplementary Table S1). For example, the long-wavelength sensitive opsin gene was found to be about four-fold (mean log2 fold change across transcriptomes) underexpressed in cave than surface animals (mean = 4.15 (log2 scale), SEM = 0.33), and was the most surface-biased gene by DE analysis. In hybrid animals containing one surface and one cave allele, we observed that the same gene showed a 10-fold (mean log2 fold change) difference between alleles (mean = 10.497, SEM = 2.79). Cis-regulatory changes may contribute to differences in long-wavelength sensitive opsin expression between populations.

By examining all genes with significant ASE, we observed that most of the genes likely had some cis-regulatory component to their change in expression between populations. We inferred this result because genes that showed DE in favor of surface animals, on the whole, tended to also show ASE in favor of the surface allele (23 out of 26 genes, Figure 3A,B). Meanwhile, genes that showed DE in favor of cave animals also tended to have ASE in favor of the cave allele (16 out of 19 genes, Figure 3A,B). For six genes (Figure 3B, marked with asterisks), we observed significant ASE that showed a strong bias in the opposite direction from what we expected from the DE analysis. For example, C-terminal binding protein-like protein and maltase A1-like protein were found to be more highly

expressed in surface animals, but by ASE the cave allele appeared to be more expressed. Such results can be explained through models of competing cis-by-trans effects.

4.2 – Discussion

4.2.1 – Candidate Genes

Typical features of cave animals include loss of eyes, loss of pigment, differences in metabolism, and enhanced sensory structures. Specifically in Asellus aquaticus, the cave form can show loss of eyes, loss of pigment, and increased appendage length (Prevorcnik et al., 2004; Turk et al., 1996; Turk-Prevorčnik and Blejec, 1998). Less is known about metabolic and behavioral differences between the cave and surface populations, but a recent study showed that acetylcholinesterase and glutathionine S transferase had lower activity in cave individuals as compared to the surface individuals, supporting the idea that the cave form has lower metabolic and locomotor activity (Jemec et al., 2017). Differences in allele-specific expression were also seen in glutathione S transferase mu 5 from a previous analysis (Stahl et al., 2015). In addition, shelter-seeking behavior has been shown to be different between some cave and surface populations (Fišer et al., 2019). Overall, we expected to find differential expression and allele-specific expression in genes involved in eye development, pigmentation, appendage development, and metabolism. As expected, some of the differentially expressed genes that we found to be differentially expressed have been shown to play a role in phototransduction, photoreceptor development, and/or eye development, such as atonal, long-wavelength sensitive opsin, cell cycle control protein 50A-like, membrane-bound transcription factor site 1 protease-like protein, protein EFR3 homolog cmp44E-like protein, pygopus-like protein, and domeless. Furthermore, a subset of the above (long-wavelength sensitive opsin, cell cycle control protein 50A-like, membrane-bound transcription factor site 1 protease-like protein, pygopus-like protein, and protein EFR3 homolog cmp44E-like protein) also showed allele-specific expression indicating that cis-regulatory changes may be responsible for the differential expression of those genes. Fewer genes with known involvement in pigmentation were observed. Scarlet, a gene involved in pigment transport (Tearle et al., 1989), was overexpressed in the surface form; however, scarlet was not shown to have allele-specific expression and therefore is unlikely to have a cis-regulatory change. Annulin-like protein was another gene of interest that was overexpressed in the cave form as compared to the surface form and had higher cave allele expression in the hybrids. Interestingly, this gene is expressed in grasshoppers in stripes along the forming limb segments and could be a candidate for appendage length changes in A. aquaticus (Bastiani et al., 1992). Another gene of interest that had showed higher ASE for the cave allele was myotubularin-related protein 9-like protein (MTMR9). Polymorphisms in this gene have been shown to be associated with obesity and glucose tolerance in Genome-wide association studies (GWAS) in humans (Tang et al., 2014; Yanagiya et al., 2007). One study in Asellus aquaticus found that the surface form had a greater feeding activity than the cave form (Mösslacher and Creuzé des Châtelliers, 1996), but little else is known regarding differences in food acquisition in the cave environment for Asellus aquaticus. However, studies in the cavefish Astyanax mexicanus have shown that some cave populations are insulin resistant and able to binge eat (Aspiras Ariel C. et al., 2015; Riddle et al., 2018).

Another interesting gene that showed both expression differences between populations and allele-specific differences is gamma-glutamyl transferase 7-like protein (GGT7). Elevated GGT is

commonly seen in individuals with non-alcoholic fatty liver disease (Rajput and Ahlawat, 2019). Interestingly, one of the cave populations of the cavefish Astyanax mexicanus develops fatty livers when exposed to high-nutrient conditions (Aspiras Ariel C. et al., 2015). Lipid, carbohydrate, and protein amount have been examined in wild-caught cave and surface specimens of Asellus aquaticus, and little difference was observed (Zidar et al., 2018). Future studies can examine whether lab-reared cave and surface forms of Asellus aquaticus differ in fat storage, insulin resistance, and starvation resistance similar to cave and surface populations of Astyanax mexicanus.

4.2.2 – Involvement of Regulatory Mutation Versus Coding Mutation in Evolution of Cave Traits

When working with species with limited genomic and genetic resources, most studies that discover the causative genes for particular phenotypes involve coding mutations. This may be due to ascertainment bias, as coding mutations are much easier to identify than cis-regulatory mutations, which could be in much larger (and uncharacterized) regions of the genome. Furthermore, cis-regulatory changes can be more difficult to test functionally than coding mutations. Because of these challenges, most of the mutations and genes identified as causative for cave-related traits in the model system of Astyanax mexicanus have been coding mutations (Aspiras Ariel C. et al., 2015; Berning et al., 2019; Elipot et al., 2014; Gross and Wilkens, 2013; Gross et al., 2009; Protas et al., 2006; Riddle et al., 2018) though there are some exceptions (Stahl and Gross, 2015). Allele-specific expression studies in hybrids are a powerful way of identifying cis-regulatory differences. Here, we have identified many genes with allele-specific expression, some of which likely have cis-regulatory changes, as inferred through a positive correlation between allele-specific expression and differential expression. Studies have indicated that much of evolutionary change occurs via cis-regulatory mutations (reviewed in (Wray, 2007)), and therefore the establishment of techniques to identify such changes in species that have limited genomic and genetic resources is crucial for identifying the genetic/genomic substrate of evolutionary change.

4.2.3 – Cis-Versus Trans-Regulation

Another major question in evolutionary biology regards whether cis- or trans-regulatory changes dominate in driving evolutionary change. In trans-regulatory changes, modifications to the expression or function of trans-regulatory factors, such as transcription factors, have a cascading effect on the expression of many other downstream target genes, driving evolutionary changes. Cis-regulatory changes, on the other hand, are more restricted, tend to occur in regulatory regions, and affect the expression of a particular gene. Trans-regulatory change might be expected to cause more drastic and pleiotropic effects, whereas cis-regulatory change would be less likely to have pleiotropic consequences (reviewed in (Signor and Nuzhdin, 2018)). Previous studies comparing species and interspecific hybrids have shown input of both trans-and cis-regulatory change (reviewed in (Signor and Nuzhdin, 2018)).

In our study, we have observed possible cases of both cis- and trans- regulation between Asellus aquaticus populations. An example of a likely cis-regulatory change is in the case of the long-wavelength sensitive opsin gene, for which cave samples showed lower expression than surface

samples. In hybrid samples, the cave allele also showed significantly lower expression as compared to the surface allele; the shared directionality of the DE and ASE results for this gene suggests that cis-regulatory effects are responsible for expression differences between the populations. On the other hand, the scarlet gene is a likely example of a trans-regulatory change in our dataset. Here, though the cave samples showed lower expression as compared to the surface samples, in hybrid samples, the cave allele was not significantly reduced in expression compared to the surface allele. When both scarlet alleles were placed in an identical trans-regulatory environment, the alleles expressed at indistinguishable levels, suggesting that differences in a trans-regulatory factor between the populations is responsible for expression differences. However, we cannot exclude cis-regulation for this gene as it is possible that certain genes show allele-specific expression only in specific tissues and sequencing transcriptomes of entire bodies dilutes out any tissue-specific allele-specific expression (Buchberger et al., 2019). In addition, we observed cases wherein cave samples showed lower expression compared to surface samples, but in hybrids the cave allele had higher expression. This might indicate both trans and cis modes of regulation, which may be evidence for compensatory mutations (reviewed in (Signor and Nuzhdin, 2018)). However, a recent study indicated that cases of compensatory cis-and trans-regulation are often overestimated as a result of correlated errors that occur when estimating ASE (Fraser, 2019). Our data were not amenable to the analysis presented in this paper, due to the methods we used to estimate ASE. The interplay between cis-and trans-regulation can ideally be examined by comparing the fold change of DE versus ASE. If the fold change of ASE is equal to the fold change of DE, cis-regulation likely explains the differential expression fully (Signor and Nuzhdin, 2018). If the fold change of ASE is less than the fold change of DE, a combination of cis-regulation and trans-regulation likely explains the differential expression. Our DE and ASE analyses used different measures of transcript abundance, and therefore the fold changes of each are not directly comparable. Future analyses with greater sample sizes and different measures of transcript abundance may enable finer examination of the differences in ASE and DE for other genes.

4.2.4 – Parent-of-Origin Effects Versus Cis-Regulation

Allele-specific expression in hybrid organisms can result from cis-regulatory change or because of parent-of-origin effects, in which the maternal and paternal copies of the gene are expressed differently, as has been observed in different organisms, including mammals, insects, and plants (Lawson et al., 2013). We cannot exclude parent-of-origin effects in the genes we found to have significant allele-specific expression, but it is likely that many of these genes have cis-regulatory changes. Future studies can eliminate potential parent-of-origin effects by generating hybrid samples from both cave female/surface male and surface female/cave male matings. As the former crosses are considerably more difficult to generate, our study was restricted to samples from the latter type of cross. Future investigations may tease apart the genes that are truly expressed as a result of cis-regulatory changes versus those with parent-of-origin effects, once it is more tractable to generate crosses with cave female and surface male animals.

4.2.5 – Comparison to Adult Asellus Transcriptome

There are two previously published transcriptomes from Asellus aquaticus, both on mostly adult samples (Pérez-Moreno et al., 2018; Speiser et al., 2014; Stahl et al., 2015). The first transcriptome utilized Roche/454 sequencing technology and was more limited in terms of actual sequence generated, though some surface embryonic samples were sequenced (Speiser et al., 2014; Stahl et al., 2015). This transcriptome was generated from individuals from the Pivka Channel of Planina Cave and a nearby surface population, Planina Polje, both found in Slovenia. More recently, a transcriptome was generated from Hungarian populations of Asellus aquaticus, including the Molnár Janós Cave population. This study found that genes involved in phototransduction were still expressed in this cave population (Pérez-Moreno et al., 2018). The authors found two expressed opsins, and neither seemed to have drastic coding changes. Consequently, they hypothesized that if vision loss has occurred in this population, it is likely due to the decreased expression of opsins. This idea is supported by our study, which uses a different cave population. Specifically, we found both differential and allele-specific expression in long-wavelength sensitive opsin. Our studies have expanded the transcriptomic resources for this species by generating a transcriptome for an additional cave population, the Rak Channel of Planina Cave. This is a useful cave population to examine as comparative embryology, as well as genetic mapping studies, have both been performed for this cave population (Mojaddidi et al., 2018; Re et al., 2018). In addition, this is the first study generating an embryonic transcriptome of a cave population of Asellus aquaticus and examining differential and allele-specific expression between cave and surface embryonic samples, giving us a window into the developmental mechanisms resulting in population-specific differences.

4.2.6 – Comparison to Other Cave-Dwelling Animal Transcriptomes

Transcriptomes of many cave-dwelling organisms have now been sequenced. Examples include other populations of Asellus aquaticus, Gammarus minus (an amphipod crustacean) (Carlini and Fong, 2017), Niphargus hrabei (another amphipod crustacean; (Pérez-Moreno et al., 2018)), cave crayfishes (Stern and Crandall, 2018a, 2018b), Poecilia mexicana (Passow et al., 2017), Sinocycloheilus species (Huang et al., 2019; Meng et al., 2013), multiple species of cave beetles (Friedrich et al., 2011; Tierney et al., 2018), multiple isopod species (Lefébure et al., 2017), and Astyanax mexicanus (Gross et al., 2013; Hinaux et al., 2013). Transcriptome studies of these cave animals often look to see whether genes involved in phototransduction are still expressed and whether there are any obvious mutations in genes involved with vision. The majority of the transcriptomes described above are from adult samples, owing to difficulties with breeding or otherwise obtaining embryonic samples. However, embryonic samples have been examined in Astyanax mexicanus (Hinaux et al., 2013).

One approach that has been lacking in studies of cave transcriptomes is using hybrid transcriptomes to evaluate allele-specific expression. A previous study in Asellus aquaticus examined allele-specific expression in a limited number of genes from a single adult hybrid sample (Stahl et al., 2015). In most other cave-dwelling animals, it is not possible to examine allele-specific expression because it requires both a cave and surface form, and they must be capable of interbreeding. However, here we show that where this approach is possible, it is a powerful way to investigate genes that might have cis-regulatory mutations. In the future, this approach can be applied to other species that have

surface and cave forms, even those that might not have fertile hybrids or viable hybrids (as long as the hybrids can start development). Potential species to examine include Poecilia mexicana and Gammarus minus.

4.2.7 – Future Steps

Here, we examined comparative expression and allele-specific expression in whole bodies of groups of individuals at a particular developmental timepoint. In the future, we aim to expand our analysis to additional timepoints and potentially specific tissues, as these two factors are known to influence both comparative and allele-specific expression (Buchberger et al., 2019). Additionally, now that methods are established to investigate differential expression and allele-specific expression in embryonic samples of cave versus surface morphs of Asellus aquaticus, one next step is to expand the analysis to other cave populations. One of the advantages of working with this species is the number of populations that are thought to be independently evolved (Konec et al., 2015; Pérez-Moreno et al., 2017). By examining gene expression and allele expression differences in these different cave populations, it should be possible to better understand how these cave-specific traits have evolved and determine if the independently evolved populations have evolved similarly or differently. Furthermore, now that a number of candidates with putative cis-regulatory changes have been identified, we can investigate them by placing them to the linkage map to determine if they coincide with mapped regions responsible for eye and pigment variation. Future work developing functional methods in Asellus aquaticus, such as genome editing and gene expression visualization, will enable testing of these genes to validate whether they are causative for associated cave-related traits.

4.3 – Materials and Methods

Animals

Animals were collected from Rakov Škocjan (surface) and the Rak Channel of Planina Cave (cave) (Figure 1A). Animals were reared in water, lighting, and food conditions as previously described (Mojaddidi et al., 2018; Protas et al., 2011; Re et al., 2018). Briefly, the animals were kept in an incubator at 12 °C with no lights and were only exposed to light when they were removed from the incubator or the incubator door was opened. Surface animals were raised in tanks with around 10 individuals per tank. Similarly, cave animals were raised in tanks with around 10 individuals per tank. Hybrid crosses were generated by mating a single cave male to a single surface female. When a female with embryos was observed in any of the above tanks, the females were monitored until the embryos were around 70% of the way through development. They were then removed from the female using a clove oil solution of 20 µL in 50 mL of fresh water as previously described (Mojaddidi et al., 2018). Embryos were kept in a small dish with commercial spring water (Crystal Geyser) until they reached 90% of embryonic development, when both pigmentation and incipient ommatidia were present in the surface, but not cave, embryos (Figure 1B,C) (Mojaddidi et al., 2018).

RNA Extraction, Library Preparation, and Sequencing

An entire brood at 90% of embryonic development was used for a single sample, which ranged from 25–89 embryos. Embryos were extracted in 200 µL of TRIzol (Thermofisher, Waltham, MA,

USA) and mechanically disrupted using an Eppendorf pestle. Samples were sent to the Functional Genomics Lab, Vincent J. Coates Genomics Sequencing Laboratory, California Institute for Quantitative Biosciences (QB3) University of California, Berkeley. Total RNA was extracted using the TRIzol Thermofisher protocol. PolyA selection was performed, and library preparation was performed using the low input protocol of the Nugen kit. Sequencing was performed using 150 bp paired end reads on both the Illumina HiSeq 4000 and the HiSeq 2500 sequencing machines to ensure a quality dataset by the sequencing facility.

De Novo Transcriptome Assembly and Annotation

A total of nine samples, which produced 36 fastq files, were processed for transcriptome assembly and annotation. We evaluated three Asellus cave embryonic samples (MPD1, MPD5, MPD6), three surface embryonic samples (MPD2, MPD3, MPD8), and three hybrid embryonic samples (MPD4, MPD7, MPD9) subjected to pair-read sequencing and processed in duplicate (total = 36 files). To achieve the most accurate mapping for downstream RNA-seq studies, we built morphotype-specific transcriptomes using SeqMan NGen (DNAStar, Madison, WI, USA). Initial de novo assemblies utilized the default assembly parameters for SeqMan NGen, including a mer size of 21, a minimum match percentage of 80%, and a maximum cluster size of 100,000, resulting in incompletely assembled contigs. This was based on the fact that default parameters yielded far fewer transcripts >1 kb in length. We sought to increase the average transcript lengths of our assemblies, and we tested a variety of parameters. We found the optimal results when we adjusted the mer size (19) and increased the minimum match percentage (to 97%) and maximum cluster size (to 300,000). This approach resulted in a 30% increase in the number of transcripts >1 kb in length (from 37,769 to 49,146 in the surface assembly). This approach also provided the longest mean transcript lengths (surface = 1061 bp, cave = 1069 bp, hybrids = 952 bp), as well as the most assembled transcripts >1 kb in length (surface = 49,233, cave = 51,822, hybrids = 52,390; Table 1). Additionally, we performed a BUSCO analysis and found highly similar results across all three transcriptomes, as well as an "integrated transcriptome" that utilized every read we generated (Supplementary Figure S1). We felt that the longest transcripts represented the best individual transcript assemblies, and therefore proceeded to annotate those assembled transcripts that were 1000 bp or longer.

All annotations were carried out using Blast2GO (v.5.2.5) running Java v.1.8.0_144. To capture the most comprehensive information, we performed two rounds of BLAST-associated annotations for each of three transcriptomes—one using the Tribolium castaneum genome as a reference and one using the SwissProt database (Table 2). The latest SwissProt database was downloaded from within Blast2GO, using the following link: https://ftp.ncbi.nlm.nih.gov/blast/db. The Tribolium database used for our blast-based annotation was downloaded from ftp.ncbi.nlm.nih.gov. We used the Tribolium reference that was uploaded in March 2016. Although the SwissProt database is perpetually being updated, both databases were retrieved (and all annotation tasks were performed) in May 2018. In brief, we submitted a fasta-formatted file containing all de novo-assembled sequences to Blast2GO, specified our database of interest, and proceeded through all default annotation steps. We implemented a script to remove all annotated transcripts associated with ribosomal or mitochondrial sequences, which ranged between 734–1066 sequences with an identified

blast hit. For all three transcriptomes (surface, cave, and hybrids), we obtained comparable results for both databases; however, the Tribolium castaneum reference provided the largest number of successful annotations.

RNA-Sequencing and Expression Analyses

Once annotation was completed using Blast2GO, we performed RNA-sequencing analyses using ArrayStar (v.13; DNAStar, Madison, WI, USA). We performed RNA-seq analyses for all transcriptome references (i.e., Tribolium castaneum and SwissProt). Specifically, we mapped reads from all of our embryonic samples to all of our transcriptome references. All of our analyses retrieved very similar results. Our workflow involved mapping sequencing reads from all three morphotypes (cave, surface, and hybrids). Gene expression results were normalized using reads per kilobase per million mapped (RPKM). This normalization strategy is necessary to control for differences in sequencing depth between samples and to compare expression levels of transcripts that differ in length. Our resulting dataset included a measure of linear total RPKM, which provides a statistical metric of expression that could be compared across datasets. This metric was then used to compare expression (based on fold change difference) between groups (e.g., cave versus surface).

We compared assemblies of different transcriptome reference files to evaluate the consistency of this calculated expression metric and found them to be highly similar. However, our process of annotation using Blast2Go periodically yielded more than one blast hit to a single, orthologous reference transcript. To deal with this issue, we averaged the expression values (i.e., RPKM values) for contigs that blasted to the same reference transcript. This yielded the most accurate expression value for each annotated gene. This calculation enabled us to correct for multiple blast hits to the same reference, however it may have inadvertently collapsed the expression for different isoforms (or paralogues) into a single transcript. This project could not evaluate the possibility of Asellus aquaticus-specific isoforms or paralogous genes, a caveat that will need to be addressed in future genome sequencing projects. Finally, given the inaccessibility of fresh tissues (with which to extract RNA for quantitative PCR validation), we used a variety of filters to maximize the validity of our reported differentially-expressed genes.

Allele-Specific Expression Using ASE-TIGAR

To assess allele-specific expression of differentially expressed genes, pairs of transcripts were identified across cave and surface transcriptomes if they had the same Tribolium castaneum Uniprot ID. For a given pair of alleles, transcripts were manually trimmed to be similar in length, based on sequence identity (Figure 2B; Supplementary File 1). We then used the ASE-TIGAR software [30] to generate transcript abundances for each allele. The software was supplied a single FASTA file containing both trimmed alleles from the cave and surface transcriptomes, as well as paired-end reads from the MPD4, MPD7, and MPD9 hybrid embryo transcriptomes. The output of this software was a file containing the expected number of fragments mapped by ASE-TIGAR, an FPKM value, and a THETA value, which was the estimated transcript abundance. We used this THETA value as our metric of expression for each allele. Given that the list of genes we selected for allele-specific expression analysis could be biased towards genes that might show allele-specific expression, we determined that it

was important to have a statistically rigorous approach to identifying genes with true allele-specific expression differences. An ideal null distribution for hypothesis testing in this scenario would be the distribution of all log fold change values for all pairs of genes. However, generating such a dataset was neither practical nor computationally feasible. Instead, we chose to simulate a null distribution that represented the intra-allele variance using the THETA values calculated for each allele in each replicate (MPD4, MPD7, MPD9). This null distribution would convolve noise arising from technical differences (batch effects, sequencing errors, etc.) and biological differences (gene expression variability between samples, gene expression noise, etc.). We generated an intra-allele null distribution by comparing inter-replicate log fold changes for all replicates within a given allele, e.g., gene X, surface allele replicate 1 vs. gene X, surface allele replicate 2, etc., using a custom Python script. Genes that had THETA = 0 in one or more replicates of one or more alleles were filtered out of the analysis. We then compared the distributions of intra-allele variations for surface alleles and cave alleles using a two-sample Kolmogorov–Smirnov (K–S) test and found that the two distributions were indistinguishable (K–S statistic = 0.0289, p-value = 0.9643). We merged the surface and cave allele null distributions and used this total distribution as a null distribution for assessing significance, also using a two-sample Kolmogorov–Smirnov (K–S) test. For each pair of alleles, we generated a distribution of log fold changes by comparing each replicate of one allele to each replicate of the other allele, for a total of nine values per pair of alleles. We used the two-sample K–S test implemented in the Pandas Python package to generate a K–S statistic and a p-value, and then performed the Benjamini–Hochberg (B–H) multiple hypothesis testing correction procedure to that p-value using a Scipy.stats Python package and $\alpha = 0.05$. Genes for which significant log fold change differences were observed based on this B–H corrected p-value were called genes with true allele-specific expression differences.

Data Deposition

All sequences analyzed in this report have been provisionally submitted to the National Center for Biotechnology Information, Sequencing Reads Archive (BioProject ID:PRJNA597080).

Acknowledgements

Thanks to Peter Trontelj, Žiga Fišer, Teo Delić, and Gregor Bračko for animal collection and/or advice throughout the project. Thanks to Žiga Fišer for critical reading of the manuscript. Thanks to Hafasa Mojaddidi, Hana Whitehorn, and Emily Klein for technical assistance. Thanks to Vincent J. Coates Genomics Sequencing Laboratory, California Institute for Quantitative Biosciences (QB3) University of California, Berkeley for the transcriptome sequencing services. In addition, we would like to thank Nicholas Lammers and Nicholas Everetts for their helpful advice on the statistical methods performed in this manuscript. Thanks also to Rachel Brem and Tal Ronnen Oron for advice about allele-specific expression analysis. Thanks also for advice and assistance from the National Center for Genome Analysis Support (NSF ABI-1458641 to Indiana University; NSF ABI-1759914 to the Pittsburgh Supercomputing Center.

Author Contributions

J.B.G. and D.A.S. designed the experiments, carried out the experiments, performed the analyses, and wrote the manuscript. B.M.C. carried out the experiments, performed analyses, and wrote the manuscript. S.B.-A. carried out the experiments and reviewed the manuscript. M.E.P. designed the experiments, carried out the experiments, and wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

M.E.P. is supported by funding from the US National Institutes of Health (NEI R15 EY029499-01A1). Research reported in this publication was supported by the National Eye Institute of the National Institutes of Health under Award Number R15EY029499. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. D.A.S. is supported by an NSF GRFP Fellowship (Division of Graduate Education, Grant/Award Number: NSF GRFP 2016230010). J.B.G. is supported by funding from the US National Institute of Health (NIDCR R01-DE025033) and the US National Science Foundation (DEB-1457630).

Conflicts of Interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

4.4 – References, Chapter 4

Aspiras Ariel C., Rohner Nicolas, Martineau Brian, Borowsky Richard L., and Tabin Clifford J. (2015). Melanocortin 4 receptor mutations contribute to the adaptation of cavefish to nutrient-poor conditions. Proc. Natl. Acad. Sci. *112*, 9668–9673. https://doi.org/10.1073/pnas.1510802112.

Bastiani, M.J., de Couet, H.G., Quinn, J.M.A., Karlstrom, R.O., Kotrla, K., Goodman, C.S., and Ball, E.E. (1992). Position-specific expression of the annulin protein during grasshopper embryogenesis. Dev. Biol. *154*, 129–142. https://doi.org/10.1016/0012-1606(92)90054-K.

Berning, D., Adams, H., Luc, H., and Gross, J.B. (2019). In-Frame Indel Mutations in the Genome of the Blind Mexican Cavefish, Astyanax mexicanus. Genome Biol. Evol. *11*, 2563–2573. https://doi.org/10.1093/gbe/evz180.

Buchberger, E., Reis, M., Lu, T.-H., and Posnien, N. (2019). Cloudy with a Chance of Insights: Context Dependent Gene Regulation and Implications for Evolutionary Studies. Genes *10*. https://doi.org/10.3390/genes10070492.

Carlini, D.B., and Fong, D.W. (2017). The transcriptomes of cave and surface populations of Gammarus minus (Crustacea: Amphipoda) provide evidence for positive selection on cave downregulated transcripts. PLOS ONE *12*, e0186173. https://doi.org/10.1371/journal.pone.0186173.

Carlini, D.B., Satish, S., and Fong, D.W. (2013). Parallel reduction in expression, but no loss of functional constraint, in two opsin paralogs within cave populations of Gammarus minus

(Crustacea: Amphipoda). BMC Evol. Biol. *13*, 1–14. .

Casane, D., and Rétaux, S. (2016). Chapter Five - Evolutionary Genetics of the Cavefish Astyanax mexicanus. In Advances in Genetics, N.S. Foulkes, ed. (Academic Press), pp. 117–159.

Elipot, Y., Hinaux, H., Callebert, J., Launay, J.-M., Blin, M., and Rétaux, S. (2014). A mutation in the enzyme monoamine oxidase explains part of the Astyanax cavefish behavioural syndrome. Nat. Commun. *5*, 3647. https://doi.org/10.1038/ncomms4647.

Fišer, Ž., Prevorčnik, S., Lozej, N., and Trontelj, P. (2019). No need to hide in caves: shelter-seeking behavior of surface and cave ecomorphs of Asellus aquaticus (Isopoda: Crustacea). Zoology *134*, 58–65. https://doi.org/10.1016/j.zool.2019.03.001.

Fraser, H.B. (2019). Improving Estimates of Compensatory cis–trans Regulatory Divergence. Trends Genet. *35*, 3–5. https://doi.org/10.1016/j.tig.2018.09.003.

Friedrich, M., Chen, R., Daines, B., Bao, R., Caravas, J., Rai, P.K., Zagmajster, M., and Peck, S.B. (2011). Phototransduction and clock gene expression in the troglobiont beetle Ptomaphagus hirtus of Mammoth cave. J. Exp. Biol. *214*, 3532–3541. https://doi.org/10.1242/jeb.060368.

Gross, J.B., and Wilkens, H. (2013). Albinism in phylogenetically and geographically distinct populations of Astyanax cavefish arises through the same loss-of-function Oca2 allele. Heredity *111*, 122–130. https://doi.org/10.1038/hdy.2013.26.

Gross, J.B., Borowsky, R., and Tabin, C.J. (2009). A Novel Role for Mc1r in the Parallel Evolution of Depigmentation in Independent Populations of the Cavefish Astyanax mexicanus. PLOS Genet. *5*, e1000326. https://doi.org/10.1371/journal.pgen.1000326.

Gross, J.B., Furterer, A., Carlson, B.M., and Stahl, B.A. (2013). An integrated transcriptome-wide analysis of cave and surface dwelling Astyanax mexicanus. PloS One *8*, e55659. .

Gross, J.B., Meyer, B., and Perkins, M. (2015). The rise of Astyanax cavefish. Dev. Dyn. *244*, 1031–1038. https://doi.org/10.1002/dvdy.24253.

Hinaux, H., Poulain, J., Da Silva, C., Noirot, C., Jeffery, W.R., Casane, D., and Rétaux, S. (2013). De novo sequencing of Astyanax mexicanus surface fish and Pachón cavefish transcriptomes reveals enrichment of mutations in cavefish putative eye genes. PloS One *8*, e53553. .

Huang, Z., Titus, T., Postlethwait, J.H., and Meng, F. (2019). Eye Degeneration and Loss of otx5b Expression in the Cavefish Sinocyclocheilus tileihornes. J. Mol. Evol. *87*, 199–208. https://doi.org/10.1007/s00239-019-09901-8.

Jeffery, W.R. (2005). Adaptive evolution of eye degeneration in the Mexican blind cavefish. J. Hered. *96*, 185–196. .

Jeffery, W.R. (2009). Evolution and development in the cavefish Astyanax. Curr. Top. Dev. Biol. *86*, 191–221. .

Jemec, A., Škufca, D., Prevorčnik, S., Fišer, Ž., and Zidar, P. (2017). Comparative study of acetylcholinesterase and glutathione S-transferase activities of closely related cave and surface Asellus aquaticus (Isopoda: Crustacea). PLOS ONE *12*, e0176746. https://doi.org/10.1371/journal.pone.0176746.

Klaassen, H., Wang, Y., Adamski, K., Rohner, N., and Kowalko, J.E. (2018). CRISPR mutagenesis confirms the role of oca2 in melanin pigmentation in Astyanax mexicanus. Dev. Biol. *441*, 313–318. .

Konec, M., Prevorčnik, S., Sarbu, S.M., Verovnik, R., and Trontelj, P. (2015). Parallels between two geographically and ecologically disparate cave invasions by the same species, Asellus aquaticus (Isopoda, Crustacea). J. Evol. Biol. *28*, 864–875. https://doi.org/10.1111/jeb.12610.

Kowalko, J.E., Ma, L., and Jeffery, W.R. (2016). Genome editing in Astyanax mexicanus using transcription activator-like effector nucleases (TALENs). JoVE J. Vis. Exp. e54113. .

Lawson, H.A., Cheverud, J.M., and Wolf, J.B. (2013). Genomic imprinting and parent-of-origin effects on complex traits. Nat. Rev. Genet. *14*, 609–617. https://doi.org/10.1038/nrg3543.

Lefébure, T., Morvan, C., Malard, F., François, C., Konecny-Dupré, L., Guéguen, L., Weiss-Gayet, M., Seguin-Orlando, A., Ermini, L., Sarkissian, C.D., et al. (2017). Less effective selection leads to larger genomes. Genome Res. *27*, 1016–1028. .

Ma, L., Jeffery, W.R., Essner, J.J., and Kowalko, J.E. (2015). Genome editing using TALENs in blind Mexican cavefish, Astyanax mexicanus. PloS One *10*, e0119370. .

McGaugh, S.E., Gross, J.B., Aken, B., Blin, M., Borowsky, R., Chalopin, D., Hinaux, H., Jeffery, W.R., Keene, A., and Ma, L. (2014). The cavefish genome reveals candidate genes for eye loss. Nat. Commun. *5*, 1–10. .

Meng, F., Braasch, I., Phillips, J.B., Lin, X., Titus, T., Zhang, C., and Postlethwait, J.H. (2013). Evolution of the eye transcriptome under constant darkness in Sinocyclocheilus cavefish. Mol. Biol. Evol. *30*, 1527–1543. .

Mojaddidi, H., Fernandez, F.E., Erickson, P.A., and Protas, M.E. (2018). Embryonic origin and genetic basis of cave associated phenotypes in the isopod crustacean Asellus aquaticus. Sci. Rep. *8*, 16589. https://doi.org/10.1038/s41598-018-34405-8.

Mösslacher, F., and Creuzé des Châtelliers, M. (1996). Physiological and behavioural adaptations of an epigean and a hypogean dwelling population of Asellus aquaticus (L.) (Crustacea, Isopoda). Arch. Für Hydrobiol. *138*, 187–198. https://doi.org/10.1127/archiv-hydrobiol/138/1996/187.

Nariai, N., Kojima, K., Mimori, T., Kawai, Y., and Nagasaki, M. (2016). A Bayesian approach for estimating allele-specific expression from RNA-Seq data with diploid genomes. BMC Genomics *17*, 2. https://doi.org/10.1186/s12864-015-2295-5.

Passow, C.N., Brown, A.P., Arias-Rodriguez, L., Yee, M.-C., Sockell, A., Schartl, M., Warren, W.C., Bustamante, C., Kelley, J.L., and Tobler, M. (2017). Complexities of gene expression patterns in natural populations of an extremophile fish (Poecilia mexicana, Poeciliidae). Mol. Ecol. *26*, 4211–4225. .

Pérez-Moreno, J.L., Balázs, G., Wilkins, B., Herczeg, G., and Bracken-Grissom, H.D. (2017). The role of isolation on contrasting phylogeographic patterns in two cave crustaceans. BMC Evol. Biol. *17*, 247. https://doi.org/10.1186/s12862-017-1094-9.

Pérez-Moreno, J.L., Balázs, G., and Bracken-Grissom, H.D. (2018). Transcriptomic Insights into the Loss of Vision in Molnár János Cave's Crustaceans. Integr. Comp. Biol. *58*, 452–464. https://doi.org/10.1093/icb/icy071.

Prevorcnik, S., Blejec, A., and Sket, B. (2004). Racial differentiation in Asellus aquaticus (L.) (Crustacea: Isopoda: Asellidae). Arch. Für Hydrobiol. *160*, 193–214. https://doi.org/10.1127/0003-9136/2004/0160-0193.

Protas, M., and Jeffery, W.R. (2012). Evolution and development in cave animals: from fish to

crustaceans. Wiley Interdiscip. Rev. Dev. Biol. *1*, 823–845. https://doi.org/10.1002/wdev.61.

Protas, M.E., Hersey, C., Kochanek, D., Zhou, Y., Wilkens, H., Jeffery, W.R., Zon, L.I., Borowsky, R., and Tabin, C.J. (2006). Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. Nat. Genet. *38*, 107–111. https://doi.org/10.1038/ng1700.

Protas, M.E., Trontelj, P., and Patel, N.H. (2011). Genetic basis of eye and pigment loss in the cave crustacean, Asellus aquaticus. Proc Natl Acad Sci USA *108*, 5702–5707. https://doi.org/10.1073/pnas.1013850108.

Rajput, R., and Ahlawat, P. (2019). Prevalence and predictors of non-alcoholic fatty liver disease in prediabetes. Diabetes Metab. Syndr. Clin. Res. Rev. *13*, 2957–2960. https://doi.org/10.1016/j.dsx.2019.07.060.

Re, C., Fišer, Ž., Perez, J., Tacdol, A., Trontelj, P., and Protas, M.E. (2018). Common Genetic Basis of Eye and Pigment Loss in Two Distinct Cave Populations of the Isopod Crustacean Asellus aquaticus. Integr. Comp. Biol. *58*, 421–430. https://doi.org/10.1093/icb/icy028.

Riddle, M.R., Aspiras, A.C., Gaudenz, K., Peuß, R., Sung, J.Y., Martineau, B., Peavey, M., Box, A.C., Tabin, J.A., McGaugh, S., et al. (2018). Insulin resistance in cavefish as an adaptation to a nutrient-limited environment. Nature *555*, 647–651. https://doi.org/10.1038/nature26136.

Signor, S.A., and Nuzhdin, S.V. (2018). The Evolution of Gene Expression in cis and trans. Trends Genet. *34*, 532–544. https://doi.org/10.1016/j.tig.2018.03.007.

Sket, B. (1956). Taxonomische Problematik tier Art Asellus aquaticus (L.) Rac. etc. Razprave-Dissertationes SAZU, CI 4.8.

Speiser, D.I., Pankey, M.S., Zaharoff, A.K., Battelle, B.A., Bracken-Grissom, H.D., Breinholt, J.W., Bybee, S.M., Cronin, T.W., Garm, A., Lindgren, A.R., et al. (2014). Using phylogenetically-informed annotation (PIA) to search for light-interacting genes in transcriptomes from non-model organisms. BMC Bioinformatics *15*, 350. https://doi.org/10.1186/s12859-014-0350-x.

Stahl, B.A., and Gross, J.B. (2015). Alterations in Mc1r gene expression are associated with regressive pigmentation in Astyanax cavefish. Dev. Genes Evol. *225*, 367–375. https://doi.org/10.1007/s00427-015-0517-0.

Stahl, B.A., Gross, J.B., Speiser, D.I., Oakley, T.H., Patel, N.H., Gould, D.B., and Protas, M.E. (2015). A Transcriptomic Analysis of Cave, Surface, and Hybrid Isopod Crustaceans of the Species Asellus aquaticus. PLOS ONE *10*, e0140484. https://doi.org/10.1371/journal.pone.0140484.

Stern, D.B., and Crandall, K.A. (2018a). Phototransduction gene expression and evolution in cave and surface crayfishes. Integr. Comp. Biol. *58*, 398–410. .

Stern, D.B., and Crandall, K.A. (2018b). The evolution of gene expression underlying vision loss in cave animals. Mol. Biol. Evol. *35*, 2005–2014. .

Tang, L., Tong, Y., Cao, H., Xie, S., Yang, Q., Zhang, F., Zhu, Q., Huang, L., Lü, Q., Yang, Y., et al. (2014). The MTMR9 rs2293855 polymorphism is associated with glucose tolerance, insulin secretion, insulin sensitivity and increased risk of prediabetes. Gene *546*, 150–155. https://doi.org/10.1016/j.gene.2014.06.028.

Tearle, R.G., Belote, J.M., McKeown, M., Baker, B.S., and Howells, A.J. (1989). Cloning and characterization of the scarlet gene of Drosophila melanogaster. Genetics *122*, 595–606. https://doi.org/10.1093/genetics/122.3.595.

Tierney, S.M., Langille, B., Humphreys, W.F., Austin, A.D., and Cooper, S.J.B. (2018). Massive Parallel Regression: A Précis of Genetic Mechanisms for Vision Loss in Diving Beetles. Integr. Comp. Biol. *58*, 465–479. https://doi.org/10.1093/icb/icy035.

Turk, S., Sket, B., and Sarbu, Ş. (1996). Comparison between some epigean and hypogean populations of Asellus aquaticus (Crustacea: Isopoda: Asellidae). Hydrobiologia *337*, 161–170. https://doi.org/10.1007/BF00028517.

Turk-Prevorčnik, S., and Blejec, A. (1998). Asellus Aquaticus Infernus, New Subspecies (Isopoda: Asellota: Asellidae), From Romanian Hypogean Waters. J. Crustac. Biol. *18*, 763–773. https://doi.org/10.1163/193724098X00647.

Verovnik, R., Sket, B., Prevorčnik, S., and Trontelj, P. (2003). Random Amplified Polymorphic DNA Diversity among Surface and Subterranean Populations of Asellus aquaticus (Crustacea: Isopoda). Genetica *119*, 155–165. https://doi.org/10.1023/A:1026085716777.

Verovnik, R., Sket, B., and Trontelj, P. (2004). Phylogeography of subterranean and surface populations of water lice Asellus aquaticus (Crustacea: Isopoda). Mol. Ecol. *13*, 1519–1532. https://doi.org/10.1111/j.1365-294X.2004.02171.x.

VEROVNIK, R., SKET, B., and TRONTELJ, P. (2005). The colonization of Europe by the freshwater crustacean Asellus aquaticus (Crustacea: Isopoda) proceeded from ancient refugia and was directed by habitat connectivity. Mol. Ecol. *14*, 4355–4369. https://doi.org/10.1111/j.1365-294X.2005.02745.x.

Wray, G.A. (2007). The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. *8*, 206–216. https://doi.org/10.1038/nrg2063.

Yanagiya, T., Tanabe, A., Iida, A., Saito, S., Sekine, A., Takahashi, A., Tsunoda, T., Kamohara, S., Nakata, Y., Kotani, K., et al. (2007). Association of single-nucleotide polymorphisms in MTMR9 gene with obesity. Hum. Mol. Genet. *16*, 3017–3026. https://doi.org/10.1093/hmg/ddm260.

Yang, J., Chen, X., Bai, J., Fang, D., Qiu, Y., Jiang, W., Yuan, H., Bian, C., Lu, J., and He, S. (2016). The Sinocyclocheilus cavefish genome provides insights into cave adaptation. BMC Biol. *14*, 1–13. .

Zidar, D., Primo\vz and \vSkufca, Prevor\vcnik, S., Kal\vcikova, G., and Jemec Kokalj, A. (2018). Energy reserves in the water louse Asellus aquaticus (Isopoda, Crustacea) from surface and cave populations: seasonal and spatial dynamics. Fundam. Appl. Limnol. *191*, 253–265. https://doi.org/10.1127/fal/2018/1118.

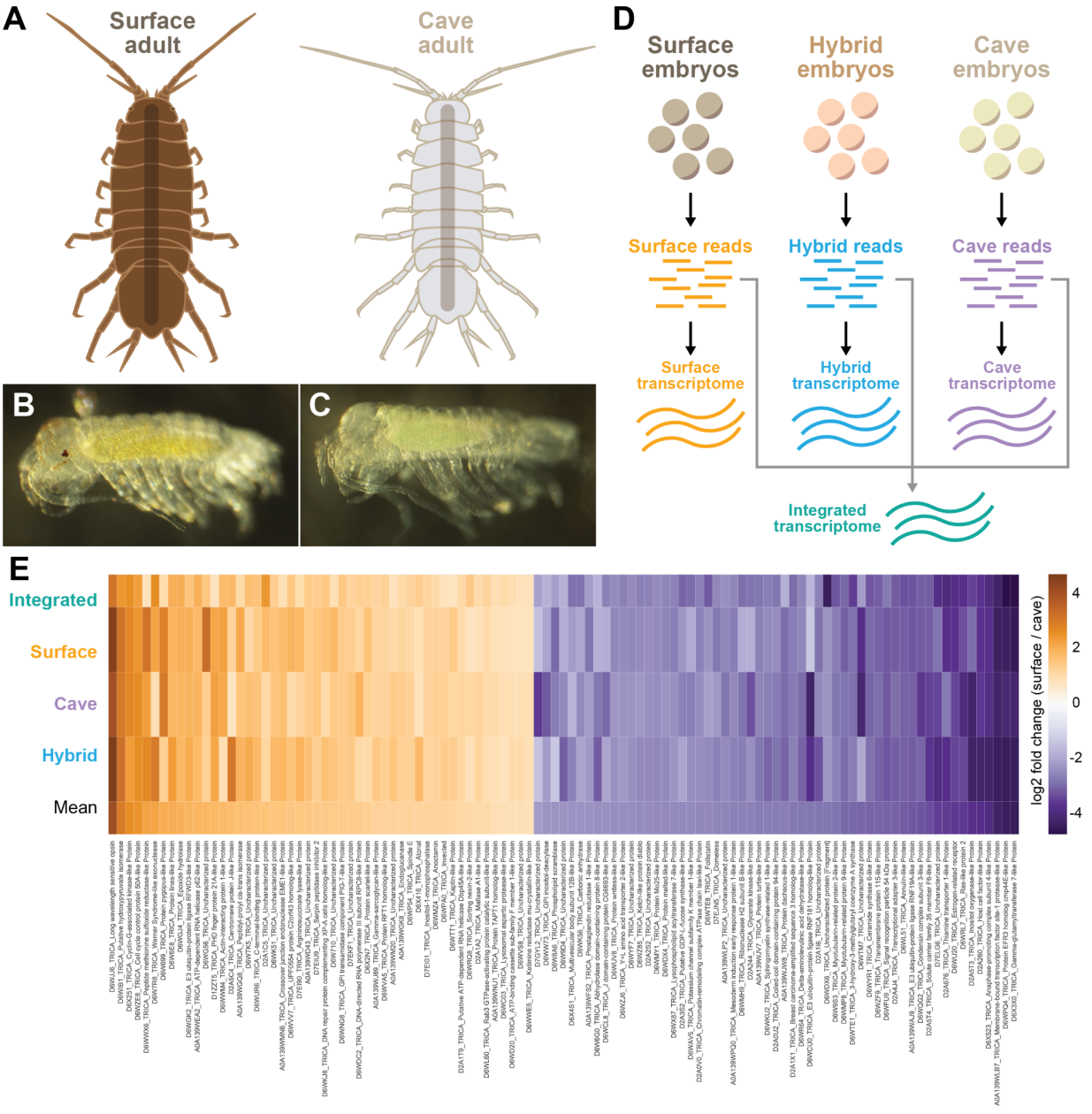


Fig. 4.1: Top 50 overexpressed genes and top 50 underexpressed genes in the cave samples as compared to the surface samples.

A) Illustrations of a surface adult and a cave adult. Cave adults lack pigmentation and eyes, and have longer limbs.

B) Surface embryo that has gone through 90% of embryonic development.

C) Cave embryo that has gone through 90% of embryonic development.

D) Four different transcriptomes were generated, one from the cave embryonic samples, one from surface embryonic samples, one from hybrid embryonic samples, and one from all embryonic samples (referred to as the integrated transcriptome).

E) Heatmap showing the top 50 downregulated genes in the cave form (various shades of orange) or top 50 upregulated genes in the cave form (various shades of purple). All genes shown had the same direction of fold change and a standard deviation of less than 8 across all four analyses. Uniprot ID and gene name from the Tribolium castaneum genome is shown.

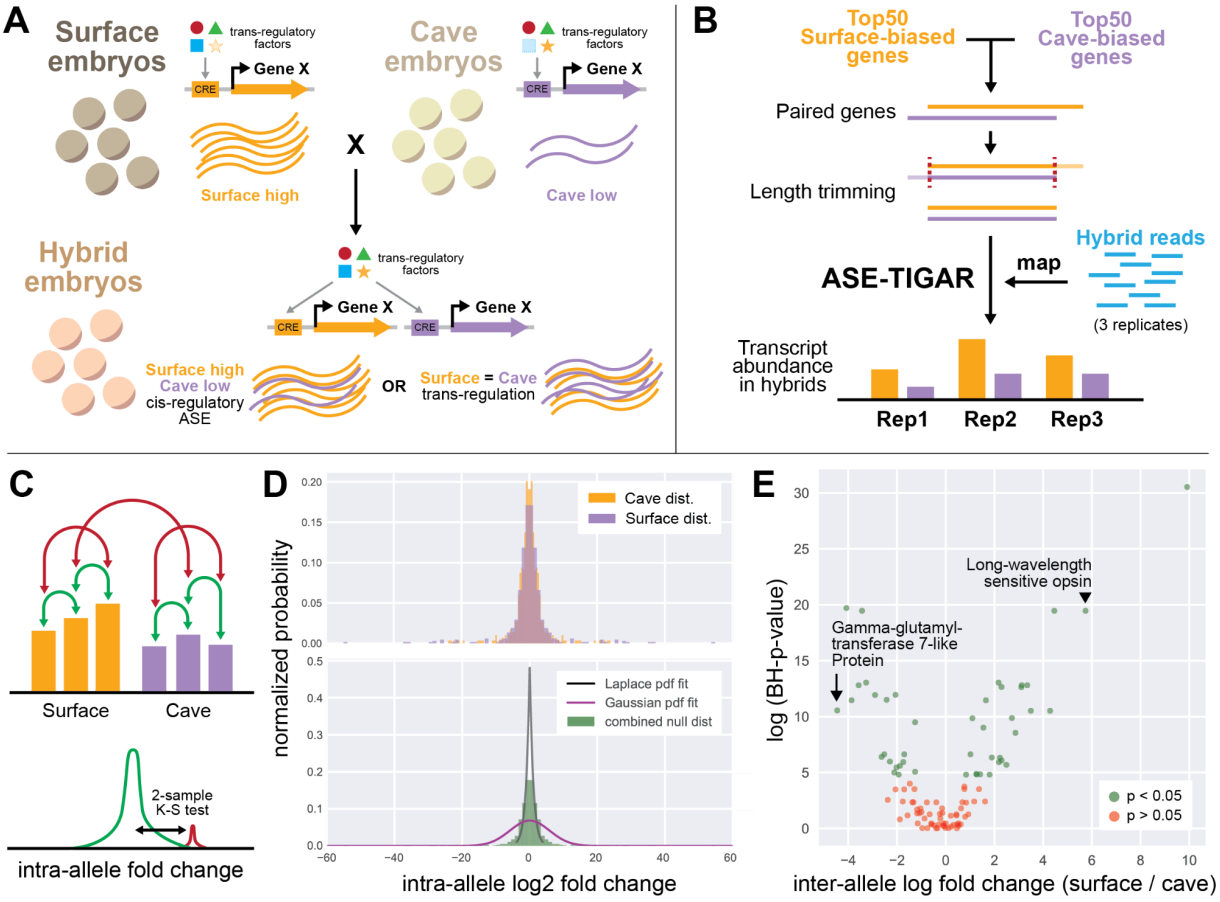


Fig. 4.2: Experimental design of allele-specific expression analysis.

A) A hypothetical example of gene X, which is differentially expressed between cave versus surface individuals and also shows allele-specific expression with lower expression of the cave allele. In this case, the underlying mechanism may be a result of differences between trans-regulatory factors between populations, or in the cis-regulatory sequence. If cis-regulatory mechanisms dominate, then we expect to see similar surface-biased expression in hybrids. However, if trans-regulatory mechanisms dominate, then the equalized trans-regulatory environment in hybrids will result in no allele-specific expression.

B) Pipeline of allele-specific expression (ASE) analysis. The top 50 differentially expressed genes, in both directions, present in both the cave and surface transcriptomes were selected, paired genes were trimmed to the same length, and hybrid reads were mapped to the trimmed cave and surface versions of each gene.

C) Intra-allele log2 fold change was calculated by comparing transcript abundance between replicates of a given allele (green arrows). This null distribution (green curve) was then compared to distributions of inter-allele log2 fold change (red arrows, red curve) using a two-sample Kolmogorov–Smirnov test.

D) Intra-allele log2 fold change distributions for cave and surface populations. The top panel shows each allele separately, overlapped. These distributions were determined to be indistinguishable (K–S test, see Methods). The bottom panel shows the combined distribution (green), and a Laplace fit (black line) and Gaussian fit (magenta line) to the distribution. The combined null distribution does not neatly fit either a Laplace or Gaussian distribution, validating that the two-sample Kolmogorov–Smirnov test is appropriate, as it does not assume that either distribution is parametric.

E) The analysis identified 45 genes that had significant allele-specific expression (green) and 55 genes that did not have significant allele-specific expression.

Fig. 4.3: Allele-specific expression suggests cis-regulatory contribution to population difference.

A) For the subset of genes that showed significant allele-specific expression, mean log2 fold change comparing surface and cave differential expression.

B) Mean log2 fold change comparing surface allele and cave allele expression within the hybrid, in the same order as in (A).

C) A Spearman correlation test indicates that differentially expressed (DE) and ASE analyses are significantly correlated (correlation = 0.5241, p-value = 2.197 × 10−4). Note: The gene with the highest allele-specific expression, an uncharacterized protein, was omitted from the scatter plot for ease of visualization.

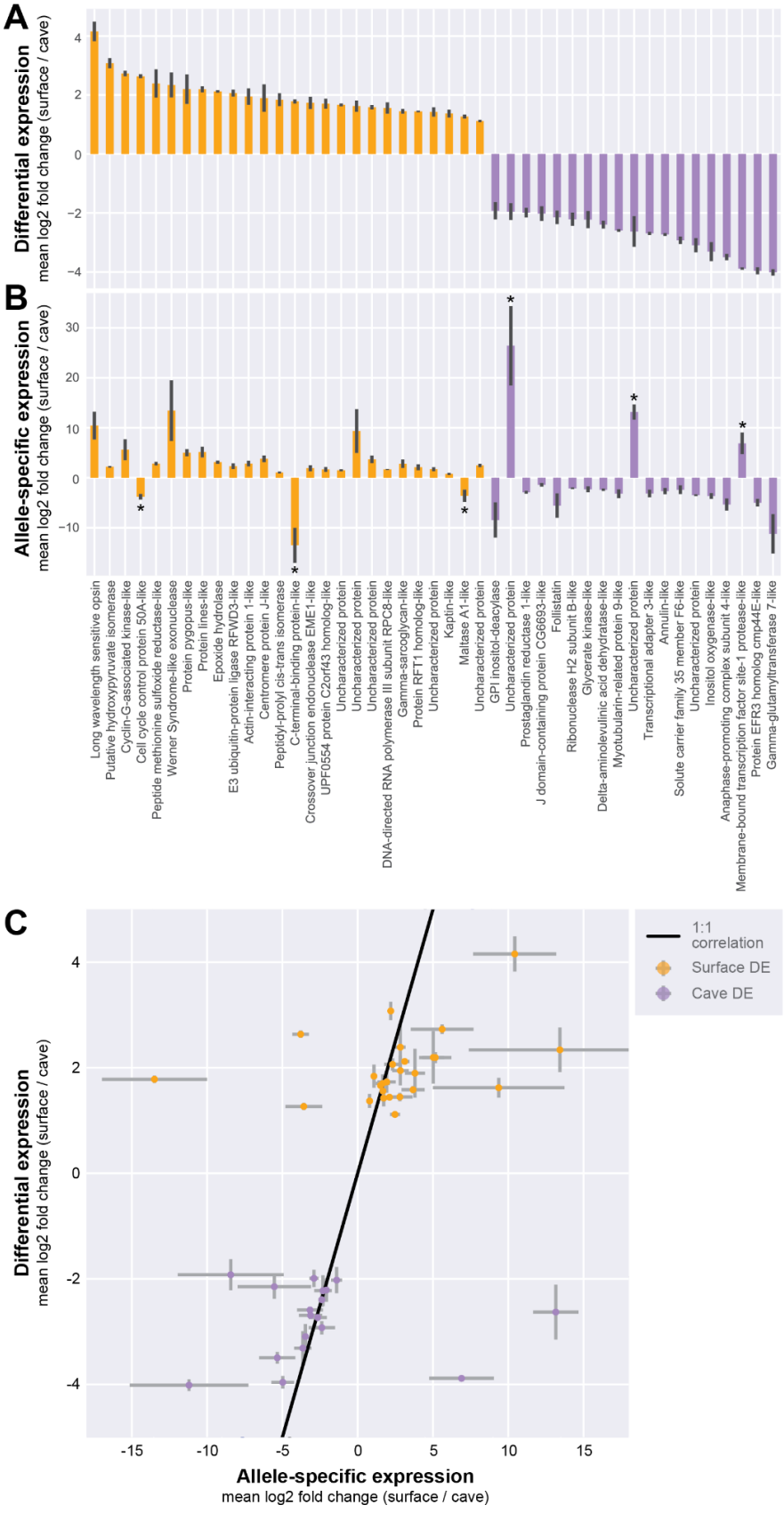


Table 4.1: Comparison of transcriptome assemblies of Asellus cave morphs, surface morphs and hybrid individuals.

Sequence Read Summary	**Surface Morphs**	**Cave Morphs**	**Hybrids**
Total Assembled Reads	155039720	164487662	132336702
Total Unassembled Reads	83386227	109047509	84947885
Total Reads Excluded by Sampling	126373422	87592389	176811792
Total Number of Reads	364799369	361127560	394096379
Transcript Summary			
Total number of Transcripts	113432	119569	143962
Average Length of Assembled Transcripts	1061	1069	952
Assembled Transcripts >1kb	49,233	51,822	52,390
Assembly Time	50.7 h	54.7 h	54.2 h

Table 4.2: Comparison of transcriptome assemblies of Asellus cave morphs, surface morphs and hybrid individuals.

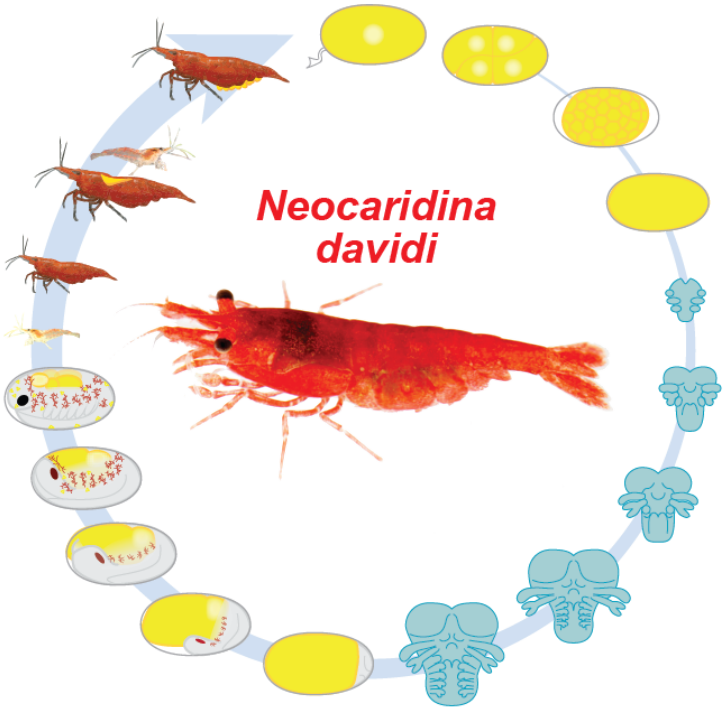
	Surface Morphs	**Cave Morphs**	**Hybrids**
Tribolium Genome Database			
Total Number of Transcripts	113,432	119,569	143,962
Assembled Transcripts >1kb	49,233	51,822	52,390
No BLAST hits	28,648	30,340	30,709
Ribosomal sequences	518	749	712
Mitochondrial sequences	880	734	973
Total number of annotated sequences	19,187	19,999	19,996
	Surface Morphs	**Cave Morphs**	**Hybrids**
SwissProt Database			
Total Number of Transcripts	113,432	119,569	143,962
Assembled Transcripts >1kb	49,233	51,822	52,390
No BLAST hits	29,918	31,928	32,157
Ribosomal sequences	603	624	839
Mitochondrial sequences	986	992	1066
Total number of annotated sequences	17,726	18,278	18,328

5 – Developmental staging and embryological techniques for the decapod crustacean *Neocaridina davidi*

Dennis A Sun[1], Taewoo Jean Kim[2], Priyanka Pradhan[3], and Nipam H Patel[4,5]

1. Department of Molecular and Cell Biology, University of California, Berkeley, CA
2. Department of Integrative Biology, University of California, Berkeley, CA
3. Feinberg School of Medicine, Northwestern University, Chicago, IL
4. Marine Biological Laboratory, Woods Hole, MA
5. Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL

Abstract



Decapods are a diverse group of crustaceans with a diverse variety of body plans, life histories, and geographic distributions. The crustaceans most commonly used as food sources – crabs, shrimp, and lobsters – are members of the decapod clade. Despite their commercial importance, the embryonic development of decapods is relatively less thoroughly characterized than for isopods such as *Porcellio scaber* or amphipods such as *Parhyale hawaiensis*. As of yet, a tractable and widely available laboratory model for decapod development has also not been established. Here, we characterize the development of the cherry shrimp *Neocaridina davidi*, previously known as *Neocaridina heteropoda* or *Neocaridina denticulata sinensis* using brightfield microscopy, nuclear staining, and immunofluorescence, with the goal of establishing this organism as a decapod model system. Moreover, we describe animal husbandry techniques for collecting early stage embryos, as well as early attempts and current challenges at establishing microinjection in this organism. This work will serve as a platform for future studies that will establish *Neocaridina* as a decapod genetic model organism.

5.0 – Introduction

Decapod crustaceans are a highly speciose group of organisms and contain the most commercially important crustaceans: crabs, shrimp, and lobsters. Their name – "deca": ten, "pod": leg – refers to their stereotypical collection of 10 visible legs, although in reality these organisms have an appendage on every segment of the body, as is true for other crustaceans (Wolfe et al., 2019).

The development of decapod crustaceans has been characterized to varying degrees of thoroughness. Some examples of developmental studies include gross morphological characterization of development in the crabs *Goniopsis pulchra* and *Aratus pisonii* (García-Guerrero and Hendrickx, 2004), scanning electron microscopy of the crayfish *Pacifastacus leniusculus* (Celada et al., 1987), brightfield microscopy of the ornamental shrimp *Lysmata ankeri* and the red-striped shrimp *Lysmata vittata* (Alves et al., 2019; Costa et al., 2021), cell lineage analysis using nuclear staining of the freshwater Amano shrimp *Caridina multidentata* (Klann and Scholtz, 2014), and nuclear developmental staining of the peppermint shrimp *Lysmata boggessi* (Romero-Carvajal et al., 2018). Thorough staging guides have been developed for far fewer organisms – one being the crayfish *Cherax destructor* (Sandeman and Sandeman, 1991). Embryonic gene expression has been examined in a few species, such as *Procambarus clarkii* (Abzhanov and Kaufman, 2000a; Martin et al., 2015). Despite their commercial importance, a tractable and widely-available genetic model system for decapods has not yet been established.

Many decapods, such as crabs and lobsters, undergo larval post-embryonic development where the body plan of the larva is not the same as that of the final adult (Jirikowski et al., 2013, 2015). A few groups of decapods undergo more direct development, wherein the first post-embryonic stage more closely resembles the final adult body plan. To establish a decapod model system for investigating *Hox* function, we selected the Caridean red cherry shrimp, *Neocaridina davidi* (previously known as *Neocaridina heteropoda* or *Neocaridina denticulata sinensis*), as a candidate for the development of embryological approaches.

Neocaridina davidi is a favorable system for establishment as a decapod model system, as it is a popular aquarium pet with a closed lifecycle, has a well-established animal husbandry system, and has some available genomic sequence for the design of CRISPR-Cas9 guide RNAs (Kenny et al., 2014; Mykles and Hui, 2015). Moreover, its early post-embryonic stages appear to mirror the adult body plan (Pantaleao et al., 2015). As such, it is well-positioned as a candidate for the establishment of genetic manipulation for embryological studies.

In this chapter, we describe the developmental stages of *Neocaridina davidi*, with comparisons to the development observed in previous studies of decapod development. We use DAPI staining and immunofluorescence to characterize the developmental stages and reveal gene expression patterns of interest. We also demonstrate early attempts at establishing injection techniques in this organism and illustrate the challenges of working with cherry shrimp embryos, which have high intraembryonic pressure. We provide a number of potential solutions for troubleshooting microinjections in this system. While our work was unable to establish microinjection in this organism, other researchers have shown that microinjection and CRISPR-Cas9 mutagenesis are possible (Li et al., 2022). In conjunction with other work, this staging guide will prove useful to researchers aiming to establish *Neocaridina davidi* as a decapod genetic model system.

5.1 – *Neocaridina* animal husbandry and reproductive biology

Adult *Neocaridina* are sexually dimorphic (Fig. 5.1). Males have a triangular body shape and usually have lighter or more transparent coloration, whereas females are larger and have rounded abdominal tergites, as well as more intense or opaque red coloration. Coloration of *Neocaridina davidi* varies by breed, but the animals we used were of the "Fire Red" variety, and had intense, opaque red pigmentation. Fertile females have a noticeable opaque yellow or lime green ovary located at the dorsal juncture between the thorax and abdomen. Upon fertilization, females attach embryos to the pleopods, as is observed across most decapods. Aquarists commonly refer to fertile females with a prominent ovary as "saddled," and gravid females as "berried."

Mating in *Neocaridina* does not begin with amplexus, as is true for isopod and amphipod crustaceans. As such, it is more difficult to predict when fertilization will occur. We developed a protocol to improve the frequency at which embryos could be obtained for embryological experiments (Fig. 5.1). In our approach, fertile females were isolated in female-only tanks and fed aggressively to stimulate ovary maturation. Once animals appeared to have very large, prominent yellow ovaries, we re-introduced a handful of males into the tanks. This resulted in more regular and predictable mating. Recent work has suggested that ovary development in cherry shrimp may require male touch (Tropea et al., 2018). We did not examine the effects of restriction from males in ovary development, but we were nonetheless able to regularly obtain fertilized embryos after re-introduction of males.

5.2 – Stages of *Neocaridina davidi* development

We characterized *Neocaridina* development through a series of partially-overlapping timecourses of embryogenesis, which we combined to generate a relatively complete picture of ontogeny from the 1-cell stage until hatching. We characterized development using brightfield microscopy and live imaging, as well as fluorescent microscopy of DAPI and antibody stains of fixed embryos.

Early cell divisions (0-24hpf)

We characterized the very earliest stages of *Neocaridina* development using brightfield microscopy. As has been observed for other Caridean shrimp such as *Caridina multidentata* and *Lysmata boggessi*, early embryos undergo holoblastic cleavage, where cell divisions fully partition the yolk into very large cells (Fig 5.2)

We used live microscopy to trace the timing of the early cell divisions at 21°C (Fig. 5.3). During a timecourse imaging of 1-cell embryos, we observed that the first cell division occurred after ~8 hours, and the second occurred at ~12 hours post embryo collection. This suggests that the 1-cell stage lasts for at least 8 hours, providing plenty of time for early embryonic injections such as CRISPR mutagenesis, which would then affect all descendant cells of the embryo.

Morula to blastoderm (1-4dpf)

Following the first rounds of cell division, the embryo arrives at a "morula"-like stage, where large round cells are visible as protrusions from the surface of the embryo, which is retracted inward from the chorion (Fig. 5.4). These protrusions flatten out, and the embryo gradually becomes a

relatively uniform and opaque field of yolk. DAPI staining reveals that embryonic cells appear to form a blastoderm-like surface that uniformly coats the yolk by around 3dpf. Cells begin to polarize and condense on one side of the embryo, which likely results in the formation of embryonic tissue around 4dpf. These data are consistent with the morphology of the embryo at similar stages in the more thoroughly characterized *Caridina multidentata* and *Lysmata boggessi*.

Naupliar limb bud stage (6dpf)

By 6dpf, the embryo develops into a recognizable morphology, with obvious limb buds for the antennae and the mandible (Fig. 5.4). This morphology appears identical to similar stages in *Caridina multidentata* and *Lysmata boggessi*. By this stage, the future telson is visible as an indentation in the presumptive posterior of the embryo, referred to as the caudal papillae. Relative to the final position of the embryonic anterior within the egg, the embryo is "upside down" at this stage, with the anterior positioned towards the final ventral side of the egg, and the posterior pointed towards the final dorsal side of the egg. As in other previously-studied decapods, embryonic tissue will gradually expand first ventrally, then anteriorly, and close dorsally around the yolk over the course of embryogenesis.

Head development and maxilliped bud stages (7-10dpf)

At 7dpf, the optic lobes of the embryo appear to develop, as evidenced by a more obvious bilobed morphology (Fig. 5.5). In addition, staining with a cross-reactive antibody for *engrailed* (4F11) shows that the optic lobe stripe of *engrailed* begins to be visible as a handful of nuclei at this stage.

This emergence of optic lobe *engrailed* staining is very early relative to that observed in the amphipod *Parhyale hawaiensis*, for which the optic lobe stripe only becomes visible at around S19, or about 39% development (Browne et al., 2005). At that stage in *Parhyale*, all of the head and thoracic limb buds have formed, and most of the abdominal segments have developed. In contrast, at this stage in *Neocaridina*, very few segmental stripes of *engrailed* expression are visible, and only the antennal and mandibular limb buds are visible, suggesting that segmentation has not completed. This may be reflective of the "embryonic nauplius" mode of development in *Neocaridina*, in whose clade naupliar development is the ancestral state.

The caudal papilla of 7dpf embryos is more obviously bilobed, and the future abdomen of the embryo is folded over the rest of the developing thorax. This folding, as observed in other decapods, persists throughout development.

As development progresses, additional *engrailed* stripes become visible, and the maxilliped limb buds become obvious by 9dpf (Fig. 5.5). The optic lobes continue to enlarge, and the optic lobe *engrailed* stripe becomes vertically oriented.

In brightfield microscopy, the embryo appears as a very thin transparency on one end of the egg. A head-on view (not shown) strongly resembles the morphology of embryos from other decapods, such as *Cherax destructor*, where the developing embryonic limb buds are clearly visible without any staining (Sandeman and Sandeman, 1991). The embryo continues to develop as a "cap" at the end of the egg until around 11dpf.

Anterior expansion stages (11-12dpf)

Beginning at 11dpf, the transparent embryonic tissue begins to gradually extend towards the future anterior of the egg (Fig. 5.6). The embryo thus begins the process of encircling the yolk. By this stage, the developing thorax and abdomen have elongated and are easier to unfold from the embryo for dissection purposes. Counting of the *engrailed* stripes suggests that segmentation has not been completed by 12dpf. Moreover, at this stage, the T3 maxilliped is the most-posterior limb bud that is morphologically obvious.

Thoracic limb elongation stages (13-15dpf)

Beginning around 13-14dpf, the thoracic limb buds become more obviously visible (Fig. 5.7). By these stages, the segments of the abdomen are obvious, and the early stages of pleopod elongation are visible. However, the thoracic appendages remain highly compacted into a more condensed area of the body. At the lateral sides of the embryo, the carapace begins to develop as a pair of thin sheets of cells.

At 13-14dpf, the eyes of the embryo become morphologically obvious in brightfield microscopy as a pair of thin angled spots of pigmentation on the ventral side of the egg. In the dorsal posterior of the embryo, the presumptive gut becomes morphologically distinct as a pair of slightly pale regions of the yolk, each of which has a concentration of yellow coloration in the center. At around this stage, an embryonic heartbeat also begins to be noticeable in the region between the gut bubbles.

By 15dpf, the thoracic appendages become elongated and encompass a much larger portion of the embryonic anterior. The pleopod limb buds begin to extend from the clearly partitioned abdominal segments. The embryo can be clearly divided into cephalon, pereon, and pleon at this stage.

In brightfield microscopy, the embryo appears to slightly retract towards the posterior, while the presumptive gut forms two morphologically distinct “bubbles” in the dorsal posterior. This pair of gut bubbles is characterized by a pale coloration on the exterior and a yellow coloration in the center. Along the AP axis, bright red pigment cells, which have a complex branching morphology resembling astrocytes or other cells with elongated processes, become visible. These cells appear to form on the lateral margins of each developing body segment, and the boundaries of each morphological segment can be clearly seen in the gaps between these cells.

Cephalothorax stages (17-20dpf)

By around 17dpf, the previously distinct head (An1 to T3/Mxp3) and thoracic regions of the body become morphologically unified into a clear cephalothorax (Fig. 5.7). DAPI staining at this stage is partially impeded by the development of cuticle; as such, images contain some tissue where DAPI stain has not penetrated fully. The pleopods of the animal are clearly distinguishable as petal-shaped protrusions to the lateral sides of each morphological segment.

In brightfield microscopy at 17dpf, the eyes begin to darken from a bright red pigment towards a brownish coloration. The stellate red pigment cells begin to elongate in their morphology. The presumptive yolk bubbles begin to rotate and migrate further towards the dorsal surface of the embryo. The anterior-most regions of the head of the embryo reach fully towards the anterior-most end of the

egg. Through brightfield microscopy, the developing thoracic limbs are also morphologically obvious, and become much larger than the mouthparts (Mx1 to T3/Mxp3).

By 20dpf, the embryo appears to approach its final morphology. The eyes of the embryo show clear and regular rows of ommatidia, and the carapace fully encompasses the cephalothorax, with the elongating tips of the thoracic limbs folded and pointing towards the anterior.

In brightfield microscopy at 20dpf, the stellate red pigment cells become considerably more complex and large, and expand further across the surface of the embryo. In addition, small opaque cells with a whitish-yellow color emerge throughout the body. These cells are considerably less stellate and elongated in their branching morphology, and instead have a more amoeboid morphology. The eyes of the embryo begin to expand into a more rounded, rather than narrow shape. The developing gut develops a more complex structure resembling a walnut, with a variety of symmetrical internal folds.

Late stages to hatching (23-26dpf)

At the latest stages of development, the embryonic yolk becomes progressively reduced into a series of discrete bubbles, which are rapidly consumed by the developing embryo (Fig. 5.7). The eyes of the embryo rotate and become round and very dark. The embryonic tissue is highly transparent at these stages, but the gross morphology of the hatchling is obvious. The embryo begins to twitch frequently within the chorion in the days leading up to hatching. The stellate red pigment cells and opaque yellow cells continue their expansion and multiplication across the surface of the embryo, although at hatching, the embryo is still largely translucent throughout.

Early hatchling morphology for *Neocaridina* had been thoroughly described in previous work, and the morphology of the hatchlings we observed is consistent with previous data. Hatchlings raised *in vitro* died very quickly after hatching in the absence of food, potentially as a result of stress or starvation. To improve survival of hatchlings, transferring late stage embryos into larger tanks and providing a food source is essential. At these late stages, keeping the embryos shaking does not appear to be essential for further development, although to maintain aeration of the freshwater, we used an airstone to bubble the water.

5.3 – Antibody staining reveals unexpected *Ubx* + *abd-A* expression

To further characterize the development of *Neocaridina davidi*, we performed antibody stains using cross-reactive antibodies that have previously worked well across different arthropods. In our timecourse analysis, we used the anti-*engrailed* antibody 4F11 (Fig. 5.8). We additionally used the mouse DP312 antibody, which stains Pax3/7; the mouse FP6.87 antibody, which stains both Ubx and abd-A; and the rat PhUbx antibody, which was built to target *Parhyale* Ubx, but which also appears to work in *Drosophila* and other crustaceans.

We observed that DP312 (Fig. 5.8 I-K) stains parasegments in a similar but broader pattern than 4F11. This antibody would also be appropriate for staging embryogenesis, as the meristic divisions of the body are clearly marked, and the staining is more robust than with 4F11. Additionally, DP312 appears to stain a population of large cells in the embryonic posterior, which may represent the proliferation zone that gives rise to individual body segments.

When we analyzed the staining pattern of the FP6.87 antibody, we observed a surprising pattern that was not expected based on previous data from other decapod crustaceans. FP6.87 stains both Ubx and abd-A protein. As such, it is not possible to know which gene is being stained along the FP6.87 pattern. However, our knowledge of the function and pattern of *Hox* expression in crustaceans such as *Porcellio scaber*, *Parhyale hawaiensis*, *Mysidium columbiae*, and *Procambarus fallax* would suggest that the anterior boundary of FP6.87 staining represents the anterior boundary of *Ubx* expression, whereas the posterior boundary of FP6.87 staining represents the posterior boundary of *abd-A* expression (Abzhanov and Kaufman, 2000b, 2000a; Martin et al., 2015). Using this interpretation, we observed some surprising results for the expected *Ubx* and *abd-A* expression profiles.

Across diverse crustaceans, *Ubx* expression is associated with the boundary between the thoracic appendages and the maxillipeds, which are appendages that are thoracic in origin (deriving from the post-naupliar body segments, e.g. those after Mx2), but which have morphology more similar to the mouthparts, or Mn / Mx1/ Mx2 segments. The number of maxillipeds varies between different groups of crustaceans, but *Ubx* has been shown to usually be expressed at the posterior boundary of the maxilliped identity. *Scr*, a more anterior *Hox* gene, is expressed in the maxillipeds. Overexpression of *Ubx* results in loss of T1 maxilliped identity in *Parhyale*, whereas loss of *Ubx* or other more-posterior *Hox* genes results in homeotic transformations towards maxilliped, likely caused by the expansion of the *Scr* expression domain.

In Caridean shrimp, such as *Periclimenes pedersoni* or *Neocaridina davidi*, there are three pairs of maxillipeds on the T1 / T2 / T3 segments (Averof and Patel, 1997). In previous work in *Periclimenes*, *Ubx* expression was shown to have an anterior boundary at the first segment after the maxilliped segments. Our FP6.87 data in *Neocaridina* show that the anterior boundary of *Ubx* expression is not the posterior of the T3 segment, as would be expected, but instead that *Ubx* has stepwise decreases in expression in T2 and T3, both of which eventually develop into maxilliped identity.

This expression pattern is unexpected, as in nearly all previously-examined crustaceans, loss of *Ubx* expression in a thoracic segment is associated with expansion of the maxilliped identity. A recent paper in the Mystacocarid crustacean *Derochellocharis remanei* argues that a new method of maxilliped identity might be found in that crustacean, as the maxilliped segment also contains *Ubx* expression, also detected by FP6.87 (Fritsch and Richter, 2017). However, in that crustacean, *Ubx* expression is largely restricted to the posterior half of the Mxp limb, indicating that a shift in *Ubx* is still associated with the Mxp identity.

In contrast, in *Neocaridina*, it appears that the T2 and T3 maxilliped express *Ubx* at varying levels in 13dpf embryos. The morphology of these appendages in the adult is unique, in that both T2 and T3 have elongated biramous structures. The T2 appendage has a cheliped-like endopod, and an elongated feather-like exopod, whereas the T3 appendage has an elongated walking-leg like endopod, and a feather-like exopod. The endopod morphology of T2 and T3 have a thoracic-like character, which may be reflective of the *Ubx* expression in these segments.

We also stained the limbs using the *Parhyale* Ubx antibody (Fig. 5.8). However, our staining was inconclusive. The thoracic appendage segments, as well as the T3 segment, appeared to show

distinct nuclear staining. However, potential T2 staining was not distinguishable from background fluorescence at 12dpf. Optimization of the antibody staining protocol may provide better results.

In the posterior of the *Neocaridina* embryo, we also observed an unexpected expression pattern: a gradient at the predicted posterior boundary of *abd-A* expression. In other decapods (*Procambarus fallax*), mysids (*Mysidium columbiae*), and isopods (*Porcellio scaber*), *abd-A* is expressed uniformly across the abdomen, with a posterior boundary at the posterior of the A5 segment (Abzhanov and Kaufman, 2000b; Martin et al., 2015). No *abd-A* is observed in the A6 segment in these organisms. In contrast, in the amphipod *Parhyale hawaiensis*, *abd-A* is expressed as a stepwise gradient, with high expression in T6-8 and A1-3, and decreasing expression in A4 and A5, with no expression in A6 (Serano et al., 2016).

As a decapod crustacean, where the A1-A5 appendages develop into pleopods, we would expect that *abd-A* in *Neocaridina* would show a similar pattern to that of *Procambarus*, in that *abd-A* would have strong uniform expression in A1-5. In our *Neocaridina* FP6.87 staining, however, we instead observed a stepwise pattern of expression from A2-A5, which more strongly resembled the *Parhyale* expression pattern. In the hatchling, the pleopods show a slight gradient in their size, which may be a result of this gradient of *abd-A* expression.

A caveat of the results we observed is that we only profiled the embryo within a narrow developmental time window. Both the anterior *Ubx* expression pattern and the posterior *abd-A* expression pattern could be dynamic during development. *Scr* expression in isopods, for example, shifts at different times of development. As such, further experiments will be needed to determine what the ultimate *Ubx* expression pattern is at different developmental stages. Moreover, additional experiments that precisely stain *Ubx* separately from *abd-A* will be needed to resolve the true boundaries of these two genes. Regardless of these caveats, the expression patterns we observed in this organism at this developmental stage were surprising, and if maintained at later stages, would suggest that *Neocaridina* uses a slightly different "*Hox* code" that has been observed in other decapod crustaceans.

5.4 – Challenges of establishing microinjection in *Neocaridina*

To establish an organism as a new potential genetic model system, it is essential to have the ability to perform genetic manipulations. In crustaceans such as *Parhyale*, *Daphnia*, and *Artemia*, the development of techniques to introduce exogenous genetic or cellular material has been instrumental in demonstrating experimental tractability (Copf Tijana et al., 2004; Hiruta et al., 2013; Rehm et al., 2009). As in our attempts to establish injections in *Asellus aquaticus*, we encountered various challenges as we attempted to inject *Neocaridina davidi* embryos. In particular, we observed that *Neocaridina* embryos have a very hard chorion that is difficult to penetrate; embryos at early stages exhibit immense intraembryonic pressure; and that, once punctured, the chorion does not readily seal on its own.

A recent publication claims to demonstrate that placing *Neocaridina heteropoda* embryos on wetted filter paper is sufficient to enable microinjection (Li et al., 2022). This manuscript also shows data suggesting successful CRISPR-Cas9 mutagenesis of *Neocaridina heteropoda Scarlet* genes. As *Neocaridina heteropoda* and *Neocaridina davidi* are synonymous, if the manuscript's techniques are valid, this would represent a major advance in establishing a decapod crustacean genetic model system.

The result that using filter paper is sufficient to allow *Neocaridina* embryo injection is highly surprising to us, given the challenges we faced in establishing microinjections in our experiments.

Here, we describe the challenges we observed when attempting to inject *Neocaridina* embryos. We found that *Neocaridina* embryos are under tremendous intraembryonic pressure, and usually explode upon any weakening of the chorion. We attempted to resolve this problem through the design of a diverse variety of microinjection needles, adjusting the intraembryonic pressure using a variety of different osmolarity-modifying solutions, and developing techniques to generate synthetic membranes for these embryos to avoid leakage of the yolk after injection. We propose a variety of more complex solutions than the reported solution of injecting embryos on wetted filter paper.

We have yet to apply the techniques described in the more recent paper, but the solution of the microinjection challenges would be a great boon to decapod genetic research. Given that the recent publication demonstrated a low survival rate from their injections, we would suggest that integrating some of the approaches we attempted with the newly-reported result could arrive at more successful genetic perturbation in *Neocaridina*, and to begin to establish this organism as a genuine decapod genetic model system.

Glass capillary needle design

For *Parhyale* microinjections, we use long, thin pulled borosilicate glass capillaries. These capillaries are loaded with our injection mix, such as capped mRNA or CRISPR-Cas9 ribonucleoprotein complexes (Martin et al., 2015). When we tested these needles on early-stage *Neocaridina* embryos, we observed that the needles when broken to their normal bore size were too flexible to puncture the very rigid chorion. When broken to a thicker bore size, and thus having less flexibility, the puncture was very large, resulting in rupture of the embryo.

We generated a variety of different borosilicate glass needles by varying the pull strength and heating parameters. Decreasing the pull strength or heat parameter resulted in a needle with a more squat taper. We also designed needles with a rapid taper, but a wider bore size throughout the length of the needle by using two rounds of heating and pulling, modifying a protocol used for injecting *C. elegans* embryos. These needles were considerably more rigid than the *Parhyale* or squat needles, and could easily penetrate the chorion without having to break the bore size to a wider diameter. However, these needles still were not sufficient to be able to break the chorion without causing rupture.

Finally, to generate the needles with the most optimal combination of thin bore size, sharpness, and rigidity, we used beveled quartz glass needles. These needles require laser-heated pulling using specialized equipment. We beveled these needles to have a point with an approximately a 30° angle using a Sutter needle beveler. Quartz needles are considerably harder and more rigid than borosilicate glass needles, and a beveled needle allows for a much smaller hole to be created upon puncture of the chorion. When we tested these needles, we discovered that having a highly rigid, fine-tipped needle was still not sufficient to prevent embryos from rupturing.

Modulating intraembryonic pressure

After testing a variety of different needles, we determined that a sufficiently rigid and sharp needle alone was not sufficient to enable microinjection in *Neocaridina* embryos. The largest problem

with these embryos appeared to be the very high intraembryonic pressure present at early developmental stages. While this pressure is a boon for microdissection, especially at later stages when the embryonic tissue is visible, the high intraembryonic pressure represented a major barrier to successful injections in this organism.

We attempted to modify the intraembryonic pressure of *Neocaridina* embryos by immersing them in a variety of solutions that could affect osmolarity, including phosphate-buffered saline (PBS), artificial seawater (ASW), sucrose, glycerol, and poly-ethylene glycol (PEG). We selected these solutions based on the observation that many of these solutions are commonly used for preserving live cells, and thus might be compatible with embryonic survival. We performed most of these experiments using 1dpf embryos, as we conducted these experiments prior to developing a consistent protocol for obtaining 0dpf embryos. For these experiments, we tested whether the embryonic pressure varied by using the "stinger" style needles, which we were able to generate on a regular basis, unlike the quartz needles, for which specialized equipment was needed.

Our tests indicated that *Neocaridina* embryos are highly sensitive to varying levels of salt, as even 0.125X ASW (3.75% salinity) resulted in clear cellular abnormalities, including separation of oil droplets out from the yolk. Moreover, altering the salinity of the surrounding medium had a negligible impact on embryonic pressure. Embryos also did not tolerate concentrations of glycerol that were able to sufficiently decrease embryonic pressure.

Among the solutions we tested on 1dpf embryos, sucrose and PEG-400 were able to modulate intraembryonic pressure. 20% sucrose did not appear to have obvious effects on cellular morphology as observed with glycerol or salt solutions, and resulted in sufficient depressurization that embryos could be injected with a "stinger" style needle without any leakage of yolk. However, returning embryos to freshwater revealed that the chorion of the embryo does not re-seal after puncture, resulting in a continuous stream of yolk emanating from the pore.

When we were able to obtain 0dpf embryos, we attempted the 20% sucrose solution and observed that neither 20% or 40% sucrose was sufficiently able to reduce intraembryonic pressure to enable injections, suggesting that 0dpf embryos have an even higher intraembryonic pressure. We would propose that PEG-400, which was too depressurizing at 1dpf embryos, might be useful for 0dpf embryonic injections in the future.

Generating artificial membranes

We were able to effectively modulate embryonic pressure using some of the osmolarity-modifying solutions we tested. However, during these tests, we observed that the embryonic chorion is rigid, and does not easily seal after puncture occurs. As such, to ensure embryonic viability, we determined that it would be necessary to generate a synthetic chorion, so that embryonic yolk would not be ejected after injection.

We developed two approaches to generating synthetic membranes. The first was to embed embryos in droplets of low-melt agarose. To do so, we melted agarose in freshwater and brought the agarose to ~37°C. We gently mixed embryos into the still-liquid solution to form a crude suspension. We then used a plastic transfer pipette to create small droplets of agarose onto parafilm placed atop ice. This resulted in rapid gelling of the agarose, creating small capsules that surrounded the embryo.

Injecting embryos embedded in agarose – without modifying osmolarity – appeared to protect from total yolk loss. The presence of the plastic-like agarose gel impeded large volumes from yolk from being ejected from the embryo upon withdrawal of a "stinger" type needle, and embryos appeared to survive for at least a few days after being injected. Uninjected embryos were able to survive the agarose embedding until hatching, suggesting that the agarose did not inhibit normal development.

A second approach we developed was to use sodium alginate. This salt readily forms a hydrogel when exposed to calcium-containing solutions such as $CaCl_2$. Sodium alginate is commonly used as a food additive, and in modernist "molecular gastronomy" cuisine, is used for a process called "spherification", in which sodium alginate added to a liquid containing some flavorant is able to be rapidly encapsulated by a gel membrane upon exposure to $CaCl_2$.

We suspended embryos in a solution of 10% sodium alginate, and then used a plastic transfer pipette to drop bubbles of embryos into $CaCl_2$ solution. A hard, rigid gel rapidly formed around the embryos, suggesting that this approach might be used to quickly generate a synthetic chorion. Unfortunately, embryos encapsulated in sodium alginate did not appear to continue to develop. This could have been a result of extended exposure to $CaCl_2$ solution (we allowed gel formation to occur over ~10 minutes), or a result of the thickness of the gel impeding oxygen from entering the embryo. Further experiments will need to be performed to optimize sodium alginate spherification and a technique for generating synthetic membranes.

Improved injection may require combining multiple approaches

While we were unable to demonstrate effective embryonic injection and CRISPR-Cas9 mutagenesis using our approaches, Li et al. adapted an injection approach designed for pea aphid injection to enable apparently successful embryonic injection in *Neocaridina* (Li et al., 2022). This approach places embryos on an air-water interface while maintaining moisture by touching embryos to wetted filter paper. This result, if true, is very exciting. However, there appear to be ways in which the described technique could be improved. Li et al. report a low rate of embryonic survival after microinjection (2% for injection with dextran sulfate, 23% after injection with phenol red), as well as a low efficiency of apparent CRISPR-Cas9 mutagenesis (~10%, or 3/28 surviving embryos).

It is important to consider some caveats of the reported results. First, we note that some phenotypes that might suggest successful CRISPR-Cas9 mutagenesis, such as asymmetric embryonic development or apparently aberrant structures, can be observed simply as a result of injection trauma. For example, in *Parhyale hawaiensis*, injection of water alone can result in gross morphological defects, such as wrinkled limbs, gut malformation, or improper cuticle development (Schmid, 2011). Moreover, embryonic injection in *Parhyale* can result in developmental delay, irrespective of mutagenesis.

The authors of this study targeted the *Neocaridina scarlet* gene, which is involved in eye pigmentation, and observed apparent disappearance or reduction of the eyes in 3/28 surviving embryos. The images they display show embryos with slight unilateral reductions to embryonic eye morphology. However, some of the phenotypes observed could be a result of injection-induced trauma, or developmental delay. For example, the anterior-posterior position of the eyes in Fig. 5D, the lack of folds lateral to the eyes, and the slit-shaped morphology of the eyes suggest that that embryo is

at an earlier embryonic stage than in the panels in Fig. 5A-C. We demonstrate in our developmental time course that eye shape and pigmentation vary during the course of development. The apparent "loss of eye" phenotype could be a consequence of the pigmentation of the embryo being less mature in these embryos, as pigmentation increases over developmental time. The number of embryos in which these phenotypes are observed is also consistent with the frequency of injection trauma-induced phenotypes in *Parhyale*.

A more reliable candidate for CRISPR-Cas9 mutagenesis would be a gene that produces a highly conserved phenotype, such as *distal-less*, which has been used consistently across taxa with a more interpretable result, and which could not be easily explained by developmental delay phenotypes. It would be helpful to test the Li et al. techniques with guides against such targets to confirm that the wetted filter paper method is able to enable embryonic injection.

The authors also show in Supplementary Fig. S1B that embryonic injection results in yolk leakage, similarly to what we observed in our injection experiments. We are surprised that these embryos continued to survive and develop, as in our experiments, embryos with this degree of yolk leakage and trauma died within a few days after injection. It is difficult to assess the morphology of the embryos in the images provided, due to apparent degradation of image quality.

If the results of the Li et al. manuscript are reproducible, we would be excited to test a variety of potential improvements to their protocol. One simple approach that could integrate some of the methods we tested with the methods used in the other study would be to wet the filter paper with a sodium alginate solution. This would allow a thin coating of sodium alginate to form around the surface of the embryo. Following injection, one could then apply a brief wash of CaCl2 solution to quickly form an artificial membrane around the eggs, potentially generating a thin artificial chorion that would reduce the potential for yolk loss, and improve embryonic survival.

Other approaches to embryonic injection might use combinations of embryonic osmolarity modulation, artificial membranes, and other techniques such as electroporation to improve efficiency of the process and embryonic survival. Yet other approaches might leverage injection into the adult ovary, such as ReMOT, to obviate the need to solve the challenges of embryonic injection (Chaverra-Rodriguez et al., 2018; Sharma et al., 2022). The development of a robust and reproducible protocol for injecting embryos will help establish *Neocaridina* as a decapod genetic model system.

5.5 – Discussion

The development of a decapod crustacean genetic model system would enable researchers to better understand the biology of this economically important clade of organisms. In this chapter, we provide a staging guide for the cherry shrimp, *Neocaridina davidi*, with the aim of building developmental resources for this organism. Along with other work that demonstrates the potential for CRISPR-Cas9 mutagenesis in this system, these data will serve as a useful resource for future developmental studies in *Neocaridina*.

In addition to examining the stages of development in *Neocaridina*, we also generated some preliminary information about *Hox* expression during development. Notably, we observed unexpected expression patterns for *Ubx* + *abd-A* using the cross-reactive FP6.87 antibody.

Ubx expression in *Neocaridina* appeared to be contrary to the expected expression pattern in this organism based on its hatchling morphology. In most malacostracan crustaceans, the anterior boundary of *Ubx* expression sits at the posterior of the final maxilliped segment (Averof and Patel, 1997). The number of maxillipeds varies between different groups of crustaceans, and the anterior boundary of *Ubx* varies in a pattern that correlates with maxilliped identity shifts. In *Neocaridina*, at the 13dpf stage, *Ubx* expression appears to occur in both the T2 and T3 segments, which develop into maxillipeds later in development, contrary to expectations.

In this study, we only examined expression at this time point; it is possible that *Ubx* expression shifts dynamically during embryogenesis, as is observed for other *Hox* genes in other arthropods. Nonetheless, the expression of *Ubx* in the T2 and T3 segments may be important for patterning the segmental identity of these maxilliped-bearing segments. T2 *Ubx* + *abd-A* expression appears to be lower than T3 expression at 13dpf, suggesting that these two segments have begun to diverge in their identity by this stage. Each of the T1, T2, and T3 segments, which eventually bears a maxilliped, has a different morphology in hatchlings. The variability in levels of *Ubx* expression among these segments may play a role in establishing different morphologies for the limbs.

abd-A expression in *Neocaridina* was also different than expected, as it showed a gradient of expression in the abdomen. *abd-A* expression is associated with the development of pleopod identity in a variety of crustaceans. Knockout of *abd-A* also results in loss of pleopod identity in the abdomen in *Parhyale*. Interestingly, the low level of *abd-A* expression in the A4 and A5 segments of *Parhyale* is not sufficient to generate a pleopod identity, and these segments instead default to a uropod identity, which is specified by *Abd-B*. *abd-A* expression appears to have the potential for quantitative effects, such as in *Abd-B* knockouts in *Parhyale*. In those experiments, *abd-A* retains its stepwise expression pattern, but *Ubx* expression expands to the posterior, resulting in transformations of swimming legs to jumping legs and anchoring legs to forward walking legs. In this unusual scenario, limbs with low levels of *abd-A* and high levels of *Ubx* are formed. These limbs develop a morphology that is intermediate between *Ubx*-only forward walking legs and *Ubx*+*abd-A*-specified reverse jumping legs. This phenotype does not appear to require *Ubx*. In *Ubx*+*abd-A* double knockout experiments, the obvious jumping leg transformations in high-*abd-A*-expressing A1-3 are retained alongside the intermediate phenotypes of the A4 and A5 transformations. This suggests that *abd-A* acts dominantly with respect to *Ubx* expression in *Parhyale*.

In *Neocaridina* at 13dpf, a gradated *abd-A* posterior expression boundary is also observed, analogous to the pattern observed in *Parhyale*. However, the A3-A6 appendages develop into pleopods. It is possible that this gradation is specific to this developmental time point, and more thorough investigations could reveal a uniform expression pattern. If the gradated expression pattern is maintained, however, this would suggest that the dosage needed for *abd-A* to pattern pleopod identity is different than in *Parhyale*. In *Parhyale*, only high *abd-A* is sufficient to generate a combinatorial effect with *Abd-B* to generate pleopods; in *Neocaridina*, it is possible that lower levels of *abd-A* would be sufficient to generate pleopod identities.

Future experiments in *Neocaridina* will benefit from tools such as *in situ* hybridization chain reaction, which will allow for the precise determination of the expression domains of each of the *Hox* genes. Moreover, more thorough time-course experiments will reveal whether the expression patterns

of these genes are dynamic in their segmental boundaries over time. The reported development of CRISPR-Cas9 mutagenesis will also enable future researchers to directly test the function of these genes by generating knockouts. The resultant homeotic transformations should shed light into the function of these genes.

Our embryological data provide a useful developmental staging guide for *Neocaridina* and our preliminary expression studies suggest that further investigation will be needed to understand the relationship between *Hox* expression and the body plan in this system. By combining our data with recent work, we aim to establish *Neocaridina davidi* as a novel decapod genetic model organism.

5.6 – Materials and Methods

Animals

Neocaridina davidi were obtained from Pan Ocean Aquarium, Inc. in Hayward, CA. Around 100 adults were raised in large tanks of dechlorinated tap water with Fluval Stratum Volcanic Sediment. Tanks were supplemented with freshwater plants such as Java moss, Java fern (*Microsorium pteropterus*), *Anubias nana*, and *Agaria densa*. Animals were fed a varied diet including Shirakura Ebi Dama pellets, Shirakura Chi Ebi powder, and New Life Spectrum Algaemax Wafers.

Embryo culture

Embryos were collected from females by anesthetizing them in 1mL Kava kava root extract/ 40 mL dechlorinated tap water (first suggested to us by Jonathon Stillman). Animals usually decreased in movement after 30-45 minutes in this solution. We also attempted anesthetizing animals in various concentrations of clove oil, as is used for *Parhyale* or *Asellus*, but even small volumes of clove oil (2.5μL/50mL dechlorinated tap water) were lethal to *Neocaridina*. After animals reduced movement, we flipped them ventrally and gently held them with a pair of forceps. Using a pulled glass Pasteur pipette with a loop or hook at the end, we were able to scrape embryos off of the pleopods. This technique requires some care, as pulling too hard on the tethers that attach the embryos to the pleopods can result in rupturing of the embryo or damage to the female's pleopods.

Embryos were raised at room temperature (~21°C) in aquarium tank water on a rocking shaker, to simulate the frequent movement of the maternal pleopods during normal brooding conditions. Water was changed every 2-3 days with fresh water from the aquarium tank. Embryos took around 25-26 days to hatch at room temperature. We observed that hatchlings die very shortly after hatching in the absence of food; as such, at very late stages of development, it is important to move embryos into larger tanks, and provide some food source, such as fish flakes, to allow the hatchlings to survive.

Dissections and antibody staining

Embryos were dissected using forceps by grasping the regions of the egg that were primarily yolk, and then pulling away the membranes. Antibody staining was performed using previously described protocols for *Parhyale* (Liubicich et al., 2009; Martin et al., 2015; Serano et al., 2016). Stained embryos were imaged on a Zeiss LSM 780 or LSM 700 confocal microscope.

Microinjection
Needles were generated using a Sutter P-80 pipette puller. A lower heat and a stronger pull resulted in more tapered needles. Embryos were injected in tank water atop Sylgard dishes. Embryos were held using fine forceps prior to injection.

Osmolarity-modulating solutions
Osmolarity-modulating solutions were made using dechlorinated tap water with the appropriate concentration or volume of additional material added.

Agarose membranes
A microwave was used to heat dechlorinated tap water in order to dissolve 0.1-0.2% low-melting agarose. A flask containing heated agarose was held at 37°C in a standard incubator. Embryos in minimal tank water were mixed with the warm, liquid agarose. Embryos were pipetted into tiny bubbles onto parafilm placed atop crushed ice using a plastic Pasteur pipette. Agarose quickly hardened, and encapsulated embryos were placed back into dishes with tank water.

Sodium alginate membranes
Sodium alginate was dissolved in room-temperature water at a concentration of 10mg/mL. For best results, thoroughly mixing the suspended sodium alginate particles and allowing them to dissolve overnight at 4°C results in a more uniform solution. Embryos in minimal tank water were mixed with sodium alginate, and small bubbles of sodium alginate, each containing a single embryo, were dropped into a medicine cup containing 10mg/mL $CaCl_2$. Sodium alginate bubbles became very hard after ~10 minutes. A brief incubation or a smaller volume of sodium alginate around the embryos may improve embryo survival.

5.7 – References, Chapter 5

Abzhanov, A., and Kaufman, T. (2000a). Embryonic expression patterns of the Hox genes of the crayfish Procambarus clarkii (Crustacea, Decapoda). Evol. Dev. https://doi.org/10.1046/j.1525-142x.2000.00066.x.

Abzhanov, A., and Kaufman, T. (2000b). Crustacean (malacostracan) Hox genes and the evolution of the arthropod trunk. Development.

Alves, D.F.R., López Greco, L.S., Barros-Alves, S. de P., and Hirose, G.L. (2019). Sexual system, reproductive cycle and embryonic development of the red-striped shrimp Lysmata vittata, an invader in the western Atlantic Ocean. PLOS ONE *14*, e0210723. https://doi.org/10.1371/journal.pone.0210723.

Averof, M., and Patel, N.H. (1997). Crustacean appendage evolution associated with changes in Hox gene expression. Nature *388*, 682. .

Browne, W.E., Price, A.L., Gerberding, M., and Patel, N.H. (2005). Stages of embryonic development in the amphipod crustacean, Parhyale hawaiensis. Genesis *42*, 124–149. https://doi.org/10.1002/gene.20145.

Celada, J.D., de Paz, P., Gaudioso, V.R., and Fernández, R. (1987). Embryonic development of the

freshwater crayfish (Pacifastacus leniusculus Dana): A scanning electron microscopic study. Anat. Rec. *219*, 304–310. https://doi.org/10.1002/ar.1092190311.

Chaverra-Rodriguez, D., Macias, V.M., Hughes, G.L., Pujhari, S., Suzuki, Y., Peterson, D.R., Kim, D., McKeand, S., and Rasgon, J.L. (2018). Targeted delivery of CRISPR-Cas9 ribonucleoprotein into arthropod ovaries for heritable germline gene editing. Nat. Commun. *9*, 1–11. .

Copf Tijana, Schröder Reinhard, and Averof Michalis (2004). Ancestral role of caudal genes in axis elongation and segmentation. Proc. Natl. Acad. Sci. *101*, 17711–17715. https://doi.org/10.1073/pnas.0407327102.

Costa, J.W.P., Brito, T.Y. da S., Marques Neto, J.C., Abrunhosa, F.A., Maciel, C.M.T., and Maciel, C.R. (2021). Reproductive cycle and embryonic development of the ornamental shrimp Lysmata ankeri Rhyne and Lin, 2006. Aquaculture *543*, 736994. https://doi.org/10.1016/j.aquaculture.2021.736994.

Fritsch, M., and Richter, S. (2017). Unexpected UBX expression in the maxilliped of the mystacocarid crustacean Derocheilocharis remanei—evidence for a different way of making a maxilliped? Dev. Genes Evol. *227*, 289–296. https://doi.org/10.1007/s00427-017-0586-3.

García-Guerrero, M., and Hendrickx, M.E. (2004). Embryology of Decapod Crustaceans I. Embryonic Development of the Mangrove Crabs Goniopsis Pulchra and Aratus Pisonii (Decapoda: Brachyura). J. Crustac. Biol. *24*, 666–672. https://doi.org/10.1651/C-2497.

Hiruta, C., Toyota, K., Miyakawa, H., Ogino, Y., Miyagawa, S., Tatarazako, N., Shaw, J.R., and Iguchi, T. (2013). Development of a microinjection system for RNA interference in the water flea Daphnia pulex. BMC Biotechnol. *13*, 96. https://doi.org/10.1186/1472-6750-13-96.

Jirikowski, G.J., Richter, S., and Wolff, C. (2013). Myogenesis of Malacostraca – the "egg-nauplius" concept revisited. Front. Zool. *10*, 76. https://doi.org/10.1186/1742-9994-10-76.

Jirikowski, G.J., Wolff, C., and Richter, S. (2015). Evolution of eumalacostracan development—new insights into loss and reacquisition of larval stages revealed by heterochrony analysis. EvoDevo *6*, 4. https://doi.org/10.1186/2041-9139-6-4.

Kenny, N.J., Sin, Y.W., Shen, X., Zhe, Q., Wang, W., Chan, T.F., Tobe, S.S., Shimeld, S.M., Chu, K.H., and Hui, J.H.L. (2014). Genomic Sequence and Experimental Tractability of a New Decapod Shrimp Model, Neocaridina denticulata. Mar. Drugs *12*. https://doi.org/10.3390/md12031419.

Klann, M., and Scholtz, G. (2014). Early embryonic development of the freshwater shrimp Caridina multidentata (Crustacea, Decapoda, Atyidae). Zoomorphology.

Li, R., Meng, Q., Qi, J., Hu, L., Huang, J., Zhang, Y., Yang, J., and Sun, J. (2022). Microinjection-based CRISPR/Cas9 mutagenesis in the decapoda crustaceans Neocaridina heteropoda and Eriocheir sinensis. J. Exp. Biol. *225*, jeb243702. https://doi.org/10.1242/jeb.243702.

Liubicich, D.M., Serano, J.M., Pavlopoulos, A., Kontarakis, Z., Protas, M.E., Kwan, E., Chatterjee, S., Tran, K.D., Averof, M., and Patel, N.H. (2009). Knockdown of Parhyale Ultrabithorax recapitulates evolutionary changes in crustacean appendage morphology. Proc Natl Acad Sci USA *106*, 13892–13896. https://doi.org/10.1073/pnas.0903105106.

Martin, A., Serano, J.M., Jarvis, E., Bruce, H.S., Wang, J., Ray, S., Barker, C.A., C, O., Liam, and Patel, N.H. (2015). CRISPR/Cas9 Mutagenesis Reveals Versatile Roles of Hox Genes in Crustacean Limb Specification and Evolution. Curr Biol *26*, 14–26. https://doi.org/10.1016/j.cub.2015.11.021.

Mykles, D.L., and Hui, J.H.L. (2015). Neocaridina denticulata: A Decapod Crustacean Model for Functional Genomics. Integr. Comp. Biol. *55*, 891–897. https://doi.org/10.1093/icb/icv050.

Pantaleao, J., Gregati, R., and Costa, R. (2015). Post-hatching development of the ornamental 'Red Cherry Shrimp'Neocaridina davidi (Bouvier, 1904)(Crustacea, Caridea, Atyidae) under laboratorial conditions. Aquac. ....

Rehm, E.J., Hannibal, R.L., Chaw, R.C., Vargas-Vila, M.A., and Patel, N.H. (2009). Injection of Parhyale hawaiensis Blastomeres with Fluorescently Labeled Tracers. Cold Spring Harb. Protoc. *2009*, pdb.prot5128. https://doi.org/10.1101/pdb.prot5128.

Romero-Carvajal, A., Turnbull, M.W., and Baeza, J.A. (2018). Embryonic Development in the Peppermint Shrimp, Lysmata boggessi (Caridea: Lysmatidae). Biol. Bull. *234*, 165–179. https://doi.org/10.1086/698468.

Sandeman, R., and Sandeman, D. (1991). Stages in the development of the embryo of the fresh-water crayfishCherax destructor. Rouxs Arch. Dev. Biol. *200*, 27–37. https://doi.org/10.1007/BF02457638.

Schmid, B. (2011). Molecular Studies on Head Development of the Amphipod Crustacean Parhyale hawaiensis (Georg-August-University Göttingen).

Serano, J.M., Martin, A., Liubicich, D.M., Jarvis, E., Bruce, H.S., La, K., Browne, W.E., Grimwood, J., and Patel, N.H. (2016). Comprehensive analysis of Hox gene expression in the amphipod crustacean Parhyale hawaiensis. Dev Biol *409*, 297–309. https://doi.org/10.1016/j.ydbio.2015.10.029.

Sharma, A., Pham, M.N., Reyes, J.B., Chana, R., Yim, W.C., Heu, C.C., Kim, D., Chaverra-Rodriguez, D., Rasgon, J.L., Harrell, R.A., et al. (2022). Cas9-mediated gene editing in the black-legged tick, Ixodes scapularis, by embryo injection and ReMOT Control. IScience *25*, 103781. https://doi.org/10.1016/j.isci.2022.103781.

Tropea, C., Lavarías, S.M.L., and López Greco, L.S. (2018). Getting ready for mating: The importance of male touching as an accelerator of ovarian growth in a caridean shrimp. Zoology *130*, 57–66. https://doi.org/10.1016/j.zool.2018.08.003.

Wolfe, J.M., Breinholt, J.W., Crandall, K.A., Lemmon, A.R., Lemmon, E.M., Timm, L.E., Siddall, M.E., and Bracken-Grissom, H.D. (2019). A phylogenomic framework, evolutionary timeline and genomic resources for comparative studies of decapod crustaceans. Proc. R. Soc. B Biol. Sci. *286*, 20190079. https://doi.org/10.1098/rspb.2019.0079.

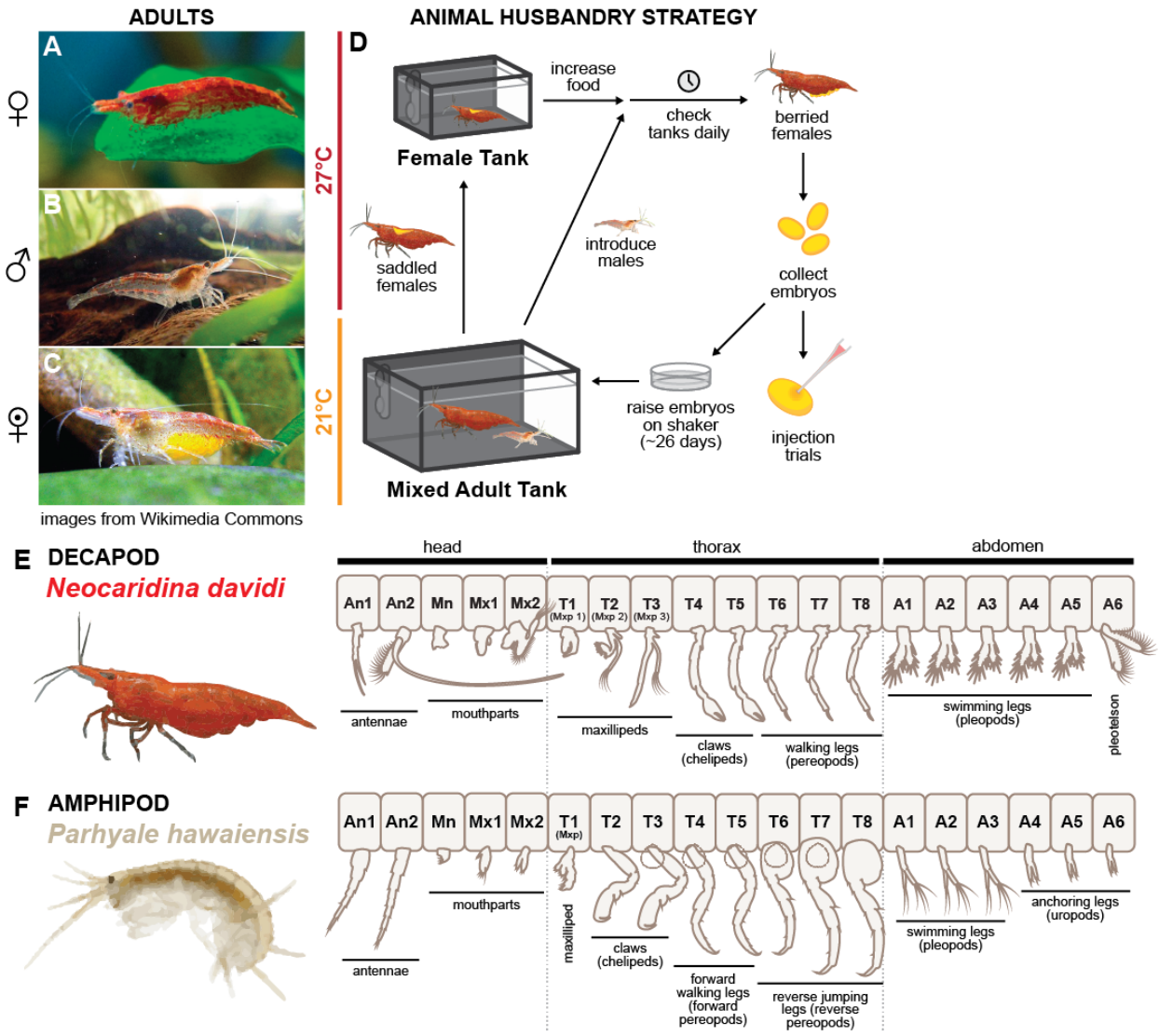


Figure 5.1: Adult morphology and animal husbandry strategy for *Neocaridina davidi*

A) Adult female *Neocaridinia*. Females have rounded abdominal tergites. The pictured female has a yellow-pigmented ovary (referred to as a "saddle") covering the dorsal side of the embryo at the junction between the cephalothorax and abdomen.

B) Adult male *Neocaridina*. Males tend to be more transparent and less well-pigmented than females. Males have a more triangular silhouette to the abdominal tergites. Males tend to spend more time swimming freely through the water.

C) Gravid female, referred to as "berried." The brightly visible yellow embryos are attached to the pleopods, as is common for many decapod crustaceans. While "berried," females tend to spend more time hiding in darkened parts of the aquarium tank, hiding beneath or behind the leaves of aquarium plants.

D) Animal husbandry strategy for obtaining early embryos. Saddled females are takeen from a mixed adult tank and fed regularly until the saddle becomes very prominent for most of the animals in the tank. Males are introduced prior to desired experimental timing, and tanks are checked regularly throughout the day for embryos. Embryos are collected and used for injection trials, or raised on a shaking plate at room temperature (~21°C) until hatching.

E) Adult body plan of *Neocaridina*. Like *Parhyale*, *Neocaridina* has multiple appendage types in the thorax. *Neocaridina* has more maxillipeds than *Parhyale* (three vs. one) and more pleopods in the abdomen.

F) Adult body plan of *Parhyale*.

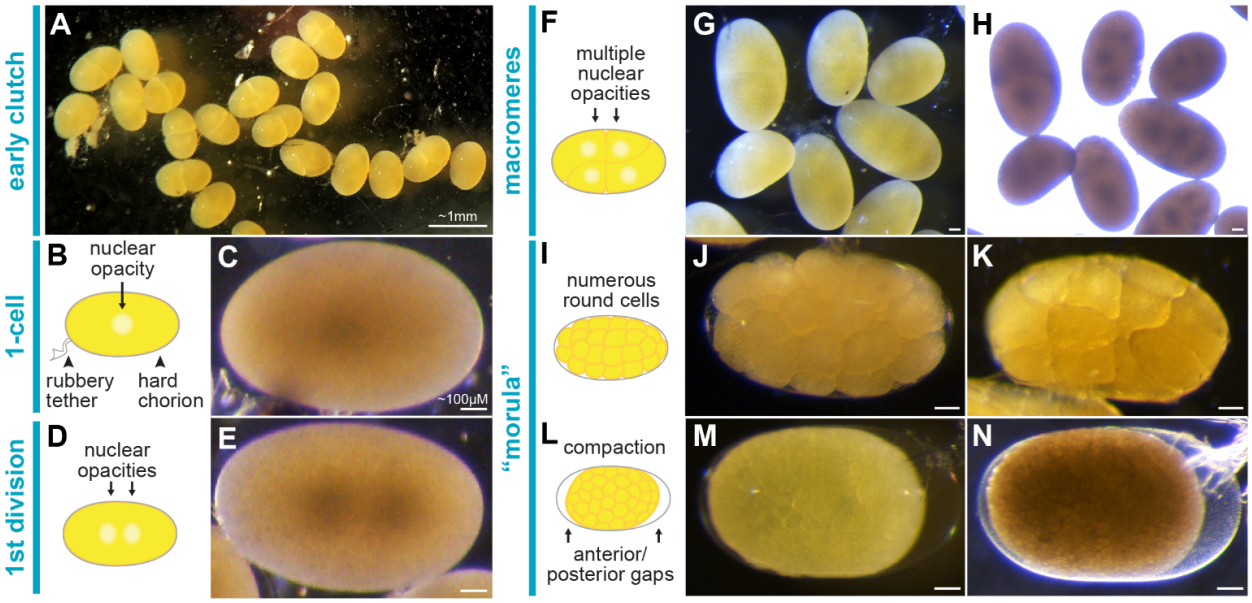


Figure 5.2: Early cell divisions of *Neocaridina davidi* embryos

Scale bars are approximate. Unless otherwise indicated, scale bars are approximately 100μM.

A) Field of early stage embryos all from a single clutch, showing variability in developmental progression for sibling embryos. Embryos at the 1-cell, 2-cell, and 4-cell stages are all observed.

B-C) Schematic (B) and angled light (C) images of 1-cell stage embryo. The most prominent feature is a pale, opaque focus within the yolk at the position of the zygotic nucleus. In transmitted or angled light microsopy, the opacity is visible as a nuclear shadow.

D-E) Schematic (D) and angled light (E) images of an early 2-cell stage embryo. Two nuclear shadows are observed, suggesting that the first nuclear cleavage has occurred, but the process of cytokinesis has not yet completed.

F-H) Schematic (F), brightfield (G), and transmitted light (H) images of macromere-stage embryos. Pale opacities are visible within the macromeres of the embryo, whose nuclei also are visible as shadows through transmitted light.

I-K) Schematic (I) image and two brightfield images (J, K) of early "morula" stage embryos. The embryonic cells are distinguishable as obviously separate cells with well-defined membranes. During cell division waves, the membranes become rounded (J) before compacting into a uniform surface again (K).

L-M) Schematic (L), brightfield (M), and transmitted light (N) images of later morula embryos. At this stage, the embryo contracts from the chorion, leaving obvious gaps in the anterior and posterior poles. The cell membranes become less obvious, although the embryo still retains a slightly rough and bumpy-looking shape, which distinguishes it from 2-3dpf embryos, which have a mostly uniform yolk with no noticeable features.

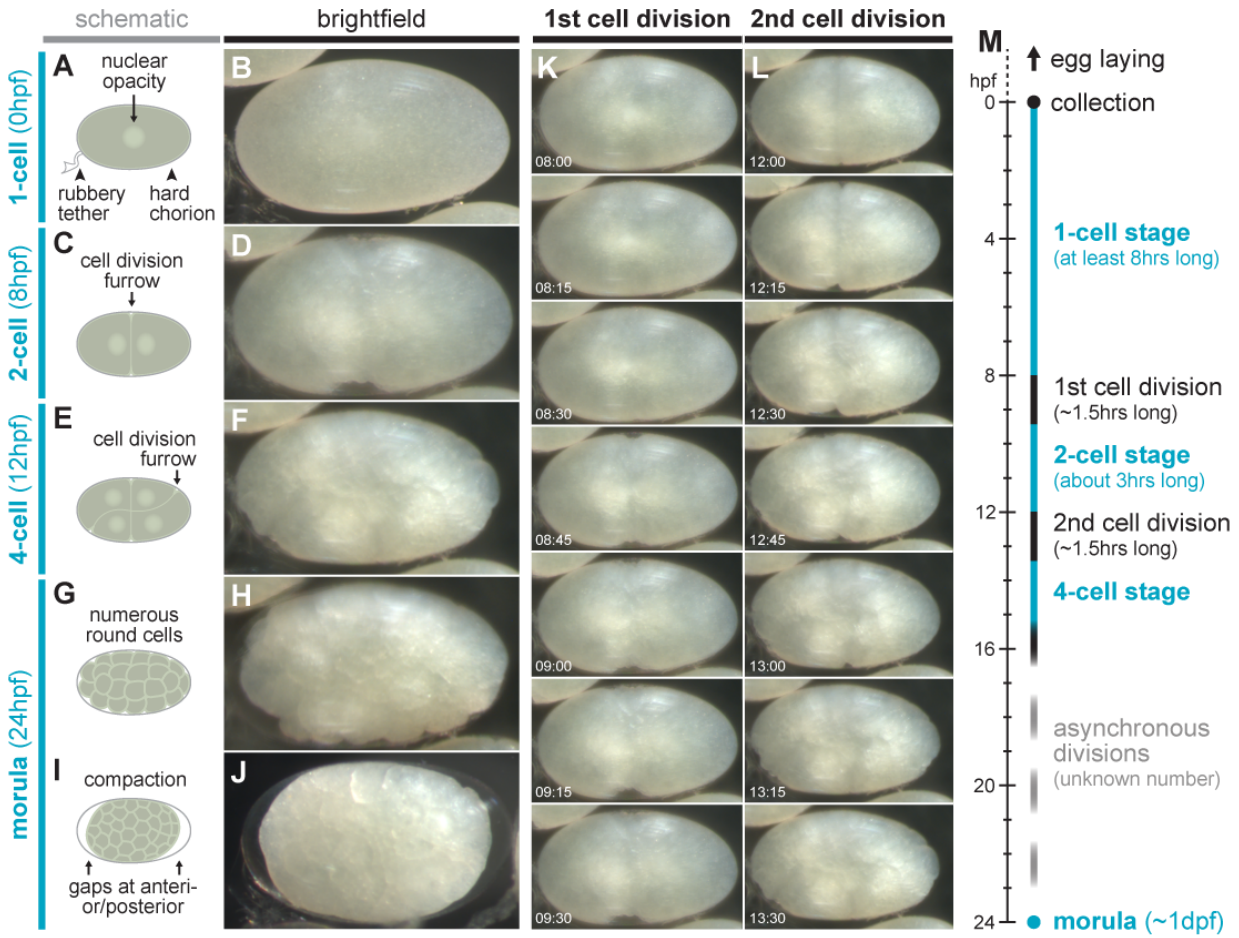


Figure 5.3: Live imaging reveals timing of early *Neocaridina* developmental stages

A-B) Schematic (A) and brightfield (B) images of 1-cell embryo.

C-D) Schematic (C) and brightfield (D) images of 2-cell embryo, taken about 8-9 hrs after the image in B.

E-F) Schematic (E) and brightfield (F) images of 4-cell embryo, taken abouut 12-13 hrs after the image in B.

G-J) Schematic (G, I) and brightfield (H, J) images of morula stage embryos. By J, around 24 hours after B, the embryo appears to have a compacted morphology with gaps at the anterioir and posterior.

K) Time series of images taken around the first cell division. Images are taken at 15 minute intervals.

L) Time series of images taken around the second cell division. Images are taken at 15 minute intervals.

M) Summary of embryonic stage timings inferred from time series live imaging. After the 4-cell stage, the image quality and angle of embryonic divisions made it difficult to determine the precise timing of additional cell divisions.

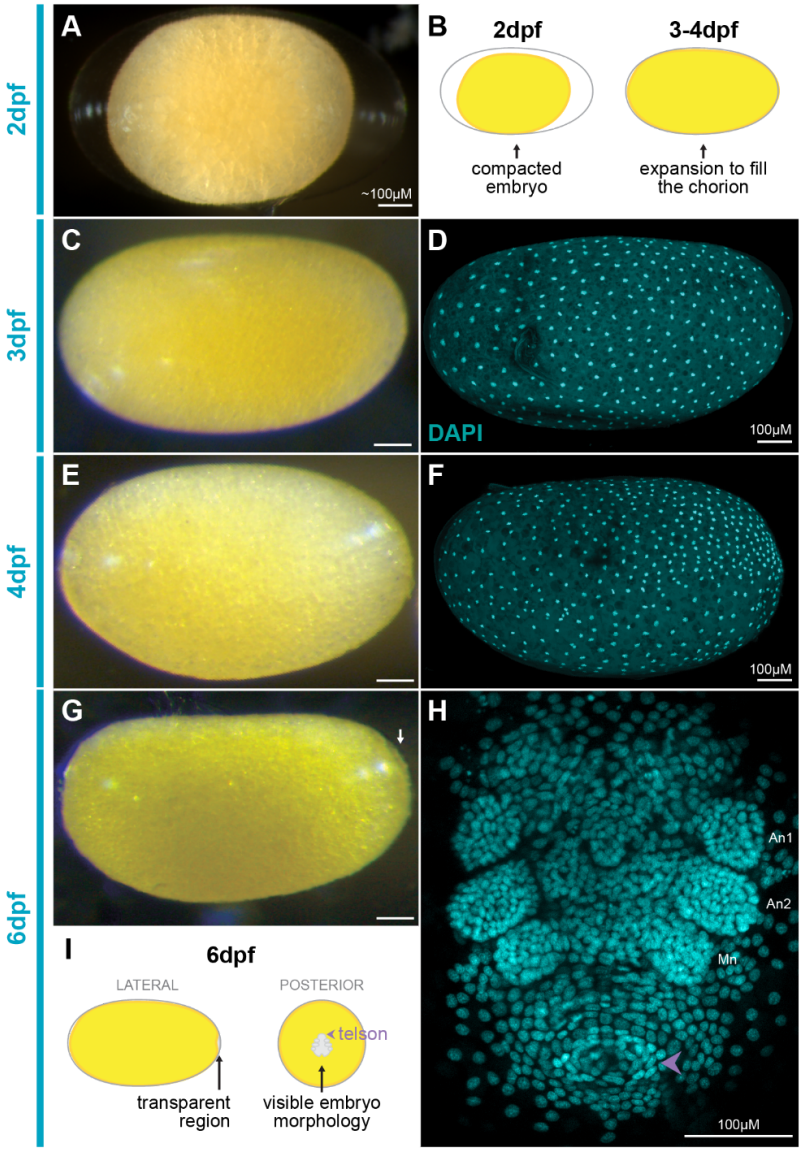


Figure 5.4: Soccerball to early germband stages (2-6dpf)

Brightfield scale bars are approximate. DAPI scale bars are exact.

A) 2dpf embryo showing gaps at the anterior and posterior ends of the embryo and compacted embryonic tissue and yolk.

B) Schematic of 2 and 3-4dpf embryos.

C-D) Brightfield (C) and DAPI (D) images of 3dpf embryo. Embryonic nuclei are uniformly distributed across the yolk surface.

E-F) Brightfield (E) and DAPI (F) images of 4dpf embryo. Embryonic cells appear to be condensing on one end of the embryo, which will likely give rise to the embryonic tissue.

G-I) Brightfield (G), DAPI (H), and schematic (I) images of 6dpf embryo. By this stage, the embryonic tissue is visible as a tiny semi-transparent pale structure in the posterior of the egg. From a head-on view (not shown) the limb buds of the naupliar segments (An1, An2, Mn) are visible. In the posterior, the caudal papillae are beginning to form (purple arrowhead).

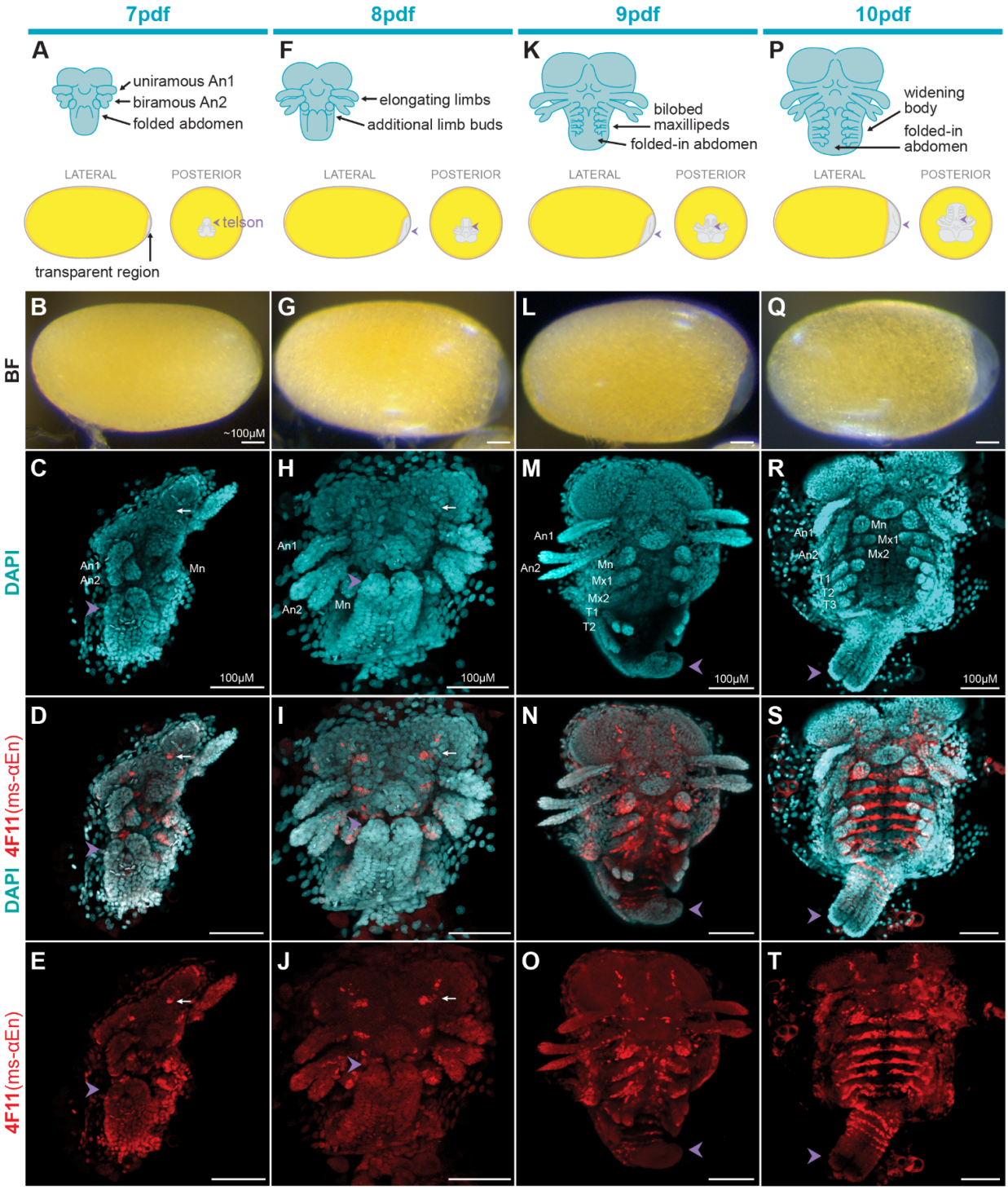


Figure 5.5: Head development and posterior segmentation stages (7-10dpf)

Brightfield scale bars are approximate. Fluorescence scale bars are exact.

Key: white arrow – optic lobe *engrailed* stripe; purple arrowhead – caudal papillae.

A-E) Schematic and images of 7dpf embryo. The embryo increases in size and becomes a more obvious transparent structure in the posterior of the egg. The optic lobe *engrailed* stripes become visible as a small number of nuclei in the anterior, and the developing eye segment becomes more obvious. The caudal papillae begin to take on a bilobed morphology.

F-J) Schematic and images of 8dpf embryo. The second antennae become obviously biramous, and the caudal papillae become more obviously bilobed. The thoracic and abdominal segments are obscured by the fold of the embryo, with the developing abdomen covering the developing thorax.

K-O) Schematic and images of 9dpf embryo. Additional leg segments develop, including the bilobed Mx1, Mx, T1, and T2. The eye lobes become larger and more prominent.

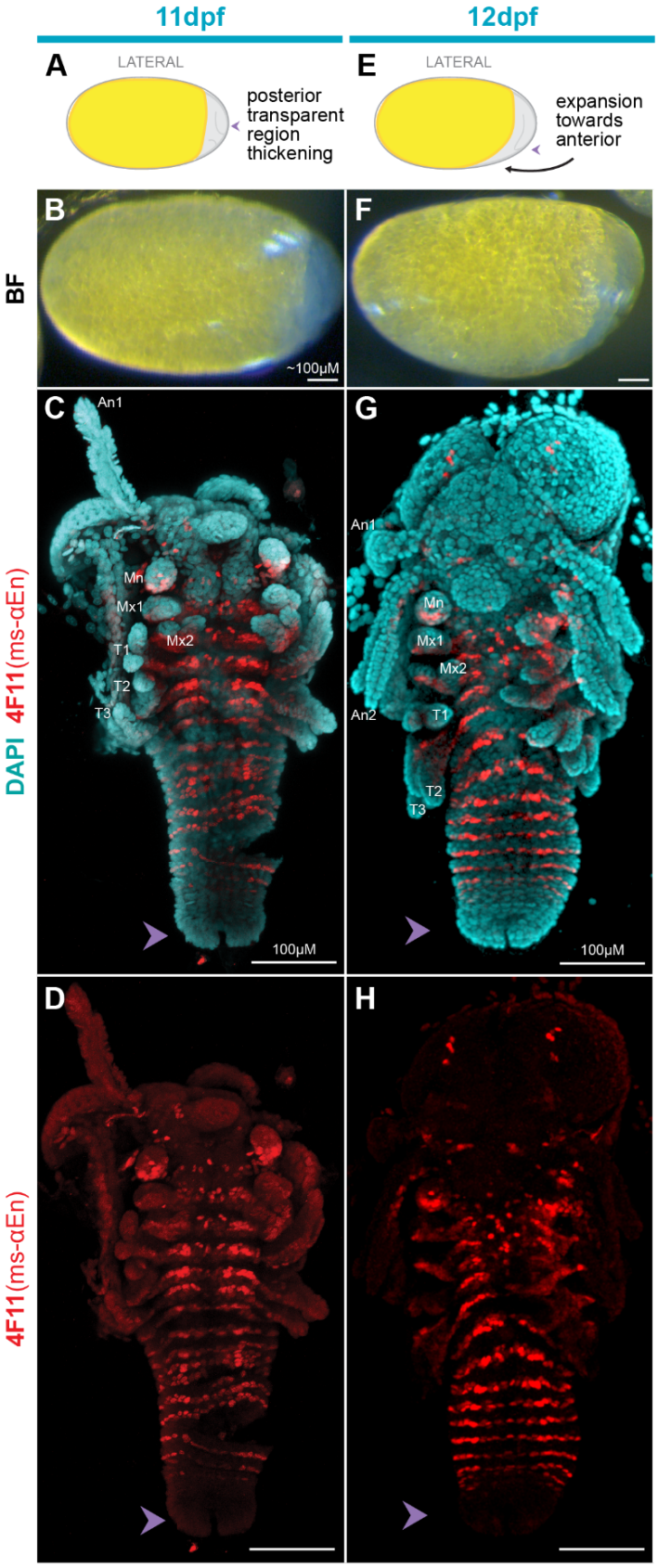
P-T) Schematic and images of 10dpf embryo. Naupliar segment and maxilliped limb buds are visible. The cephalothorax segments become wider, while the abdomen remains narrow and nested between the developing limb buds.

Figure 5.6: Anterior expansion stages (11-12dpf)

Brightfield scale bars are approximate. Fluorescence scale bars are exact.

A-D) Schematic and images of 11dpf embryos. The embryo has developed obvious morphology for the An1 to T3 (Mxp3) segments. In brightfield microscopy, the embryo is thickening in the posterior of the egg.

E-H) Schematic and images of 12dpf embryos. The anterior of the embryo begins expanding towards the anterior of the egg. The *engrailed* stripes of the thorax begin to curve as the limb-bearing segments expand in their development.



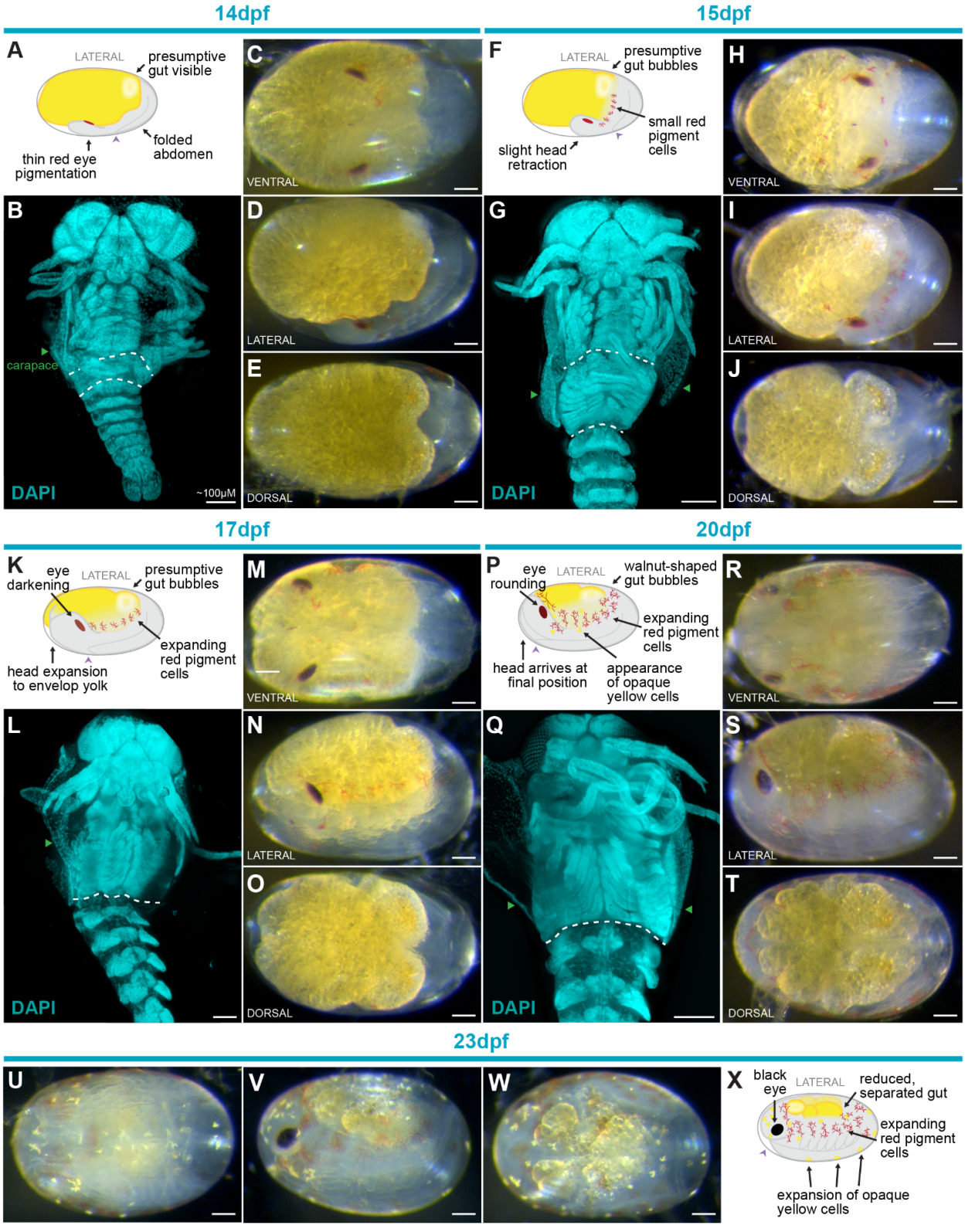

14dpf
15dpf
17dpf
20dpf
23dpf
A
LATERAL
presumptive gut visible
folded abdomen
thin red eye pigmentation
B
carapace
DAPI
~100μM
C
VENTRAL
D
LATERAL
E
DORSAL
F
LATERAL
presumptive gut bubbles
small red pigment cells
slight head retraction
G
DAPI
H
VENTRAL
I
LATERAL
J
DORSAL
K
eye darkening
LATERAL
presumptive gut bubbles
expanding red pigment cells
head expansion to envelop yolk
L
DAPI
M
VENTRAL
N
LATERAL
O
DORSAL
P
eye rounding
LATERAL
walnut-shaped gut bubbles
expanding red pigment cells
head arrives at final position
appearance of opaque yellow cells
Q
DAPI
R
VENTRAL
S
LATERAL
T
DORSAL
U
V
W
X
black eye
LATERAL
reduced, separated gut
expanding red pigment cells
expansion of opaque yellow cells

Figure 5.7: Late developmental stages (14dpf to hatching)

Scale bars represent approximately 100μM.

Key: green triangles – carapace; dotted lines in B and G mark the boundaries of the head (An1 to T3 (Mxp3)), thorax (T4-8), and abdominal (A1-6) segments; dotted lines in L and Q mark the boundary between the cephalothorax and the abdomen.

A-E) Schematic and images of 14dpf embryos. The abdominal segments are fully distinct, as well as the pleotelson. Regular rows of developing ommatidia are visible in the developing eye. The eyes also begin developing a red pigmentation. On the dorsal side of the embryo, a pair of pale regions develop in the yolk, each with a yellow region in the middle. These will become the gut bubbles. At the lateral edges of the embryo, the carapace begins to develop.

F-J) Schematic and images of 15dpf embryos. The thoracic segments develop substantially, revealing the elongated thoracic limb buds of the T4-8 segments. The gut bubbles become morphologically distinct from the embryonic yolk. Along the embryonic axis, small stellate red pigment cells begin to emerge.

K-O) Schematic and images of 17dpf embryos. The thoracic limbs have developed considerably at this point to be larger than the maxillipeds, and the head segments and thoracic segments seem to coalesce into a cephalothorax. Cuticle has formed, impeding complete DAPI staining in panel L. The pleopod limb buds become morphologically distinct from the pleopod segments. The eyes begin to darken to a brown color. The stellate red pigment cells expand.

P-T) Schematic and images of 20dpf embryos. The eyes darken to a black coloration and begin to become more round in their morphology. The gut partitions the yolk into a walnut-like shape. Opaque yellow cells appear throughout the body with a different morphology than the stellate red pigment cells.

U-X) Schematic and images of 23dpf embryos. The embryo is nearly fully developed and will hatch within 2-3 days after this point. The eyes are fully black and rounded, and have rotated to become longer along the anterior-posterior than dorsal-ventral axes. Opaque yellow cells expand in their distribution across the embryo.

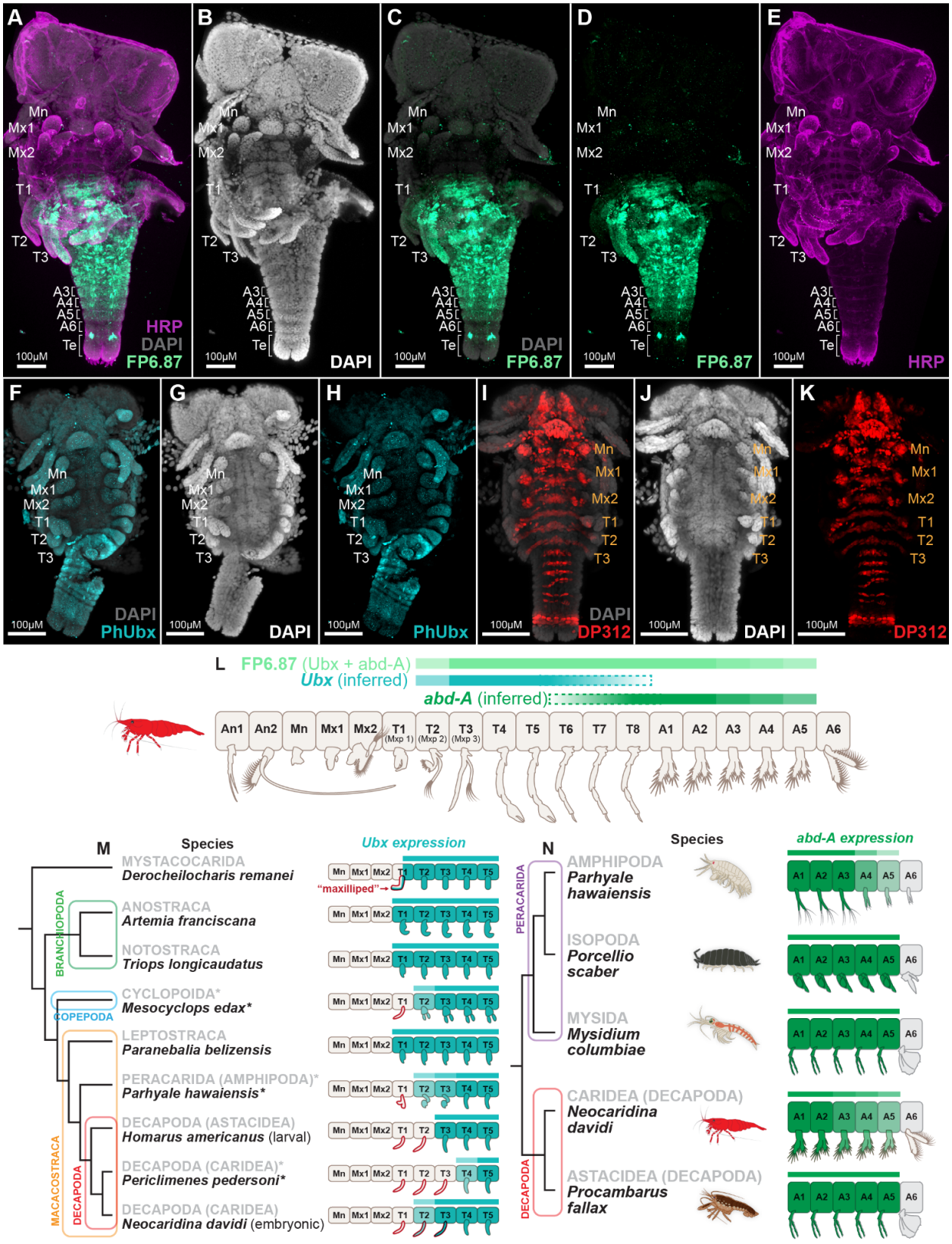

A
B
C
D
E
Mn
Mx1
Mx2
T1
T2
T3
A3
A4
A5
A6
Te
HRP
DAPI
FP6.87
100μM
F
G
H
I
J
K
PhUbx
DP312
L
FP6.87 (Ubx + abd-A)
Ubx (inferred)
abd-A (inferred)
An1
An2
T4
T5
T6
T7
T8
(Mxp 1)
(Mxp 2)
(Mxp 3)
M
Species
Ubx expression
MYSTACOCARIDA
Derocheilocharis remanei
"maxilliped"
ANOSTRACA
Artemia franciscana
NOTOSTRACA
Triops longicaudatus
CYCLOPOIDA*
Mesocyclops edax*
LEPTOSTRACA
Paranebalia belizensis
PERACARIDA (AMPHIPODA)*
Parhyale hawaiensis*
DECAPODA (ASTACIDEA)
Homarus americanus (larval)
DECAPODA (CARIDEA)*
Periclimenes pedersoni*
DECAPODA (CARIDEA)
Neocaridina davidi (embryonic)
BRANCHIOPODA
COPEPODA
MACACOSTRACA
DECAPODA
N
Species
abd-A expression
AMPHIPODA
Parhyale hawaiensis
ISOPODA
Porcellio scaber
MYSIDA
Mysidium columbiae
CARIDEA (DECAPODA)
Neocaridina davidi
ASTACIDEA (DECAPODA)
Procambarus fallax
PERACARIDA
DECAPODA

Figure 5.8: Unexpected *Ubx* + *abd-A* expression patterns in *Neocaridina* embryos

Scale bars are exactly 100μM.

A-E) Antibody stain of FB6.87 (cross-reactive *Ubx* + *abd-A* antibody, mint); HRP (purple), and DAPI (white/gray) in a 13dpf embryo. *Ubx* + *abd-A* expression appears to be bounded at the anterior in the T2 segment and in the posterior by the A5 segment.

F-H) Antibody stain of PhUbx with DAPI in a 12dpf embryo. Background staining is substantial, but nuclear morphology is visible in cells of the T3 segment onwards.

I-K) Antibody stain of DP312 (cross-reactive *Pax3/7* antibody, red) in a 12dpf embryo. In addition to stripes that mirror *engrailed* expression, there is a population of cells with large nuclei that are stained in the posterior portion of the abdomen anterior to the caudal papillae.

L) FP6.87 staining schematic and inferred *Ubx* and *abd-A* expression patterns. Contrary to the expectation from *Periclimenes*, in which the anterior boundary of *Ubx* expression occurs after the last maxilliped, in *Neocaridina* embryos at 13dpf, *Ubx* appears to be expressed the T2 and T3 appendages, each of which forms a different type of maxilliped. *abd-A* expression in *Neocaridina* appears to be gradated in the posterior, similar to *Parhyale*, which contrasts with other decapod crustaceans such as *Procambarus fallax*.

M) Schematic of *Ubx* expression patterns in diverse crustacean groups, summarizing work from (Averof and Patel, 1997; Fritsch and Richter, 2017). *Neocaridina davidi* appears to be a crustacean with a unique expression of *Ubx* in maxillipeds.

N) Schematic of *abd-A* expression patterns in diverse crustacean groups, summarizing work from (Abzhanov and Kaufman, 2000b; Martin et al., 2015).

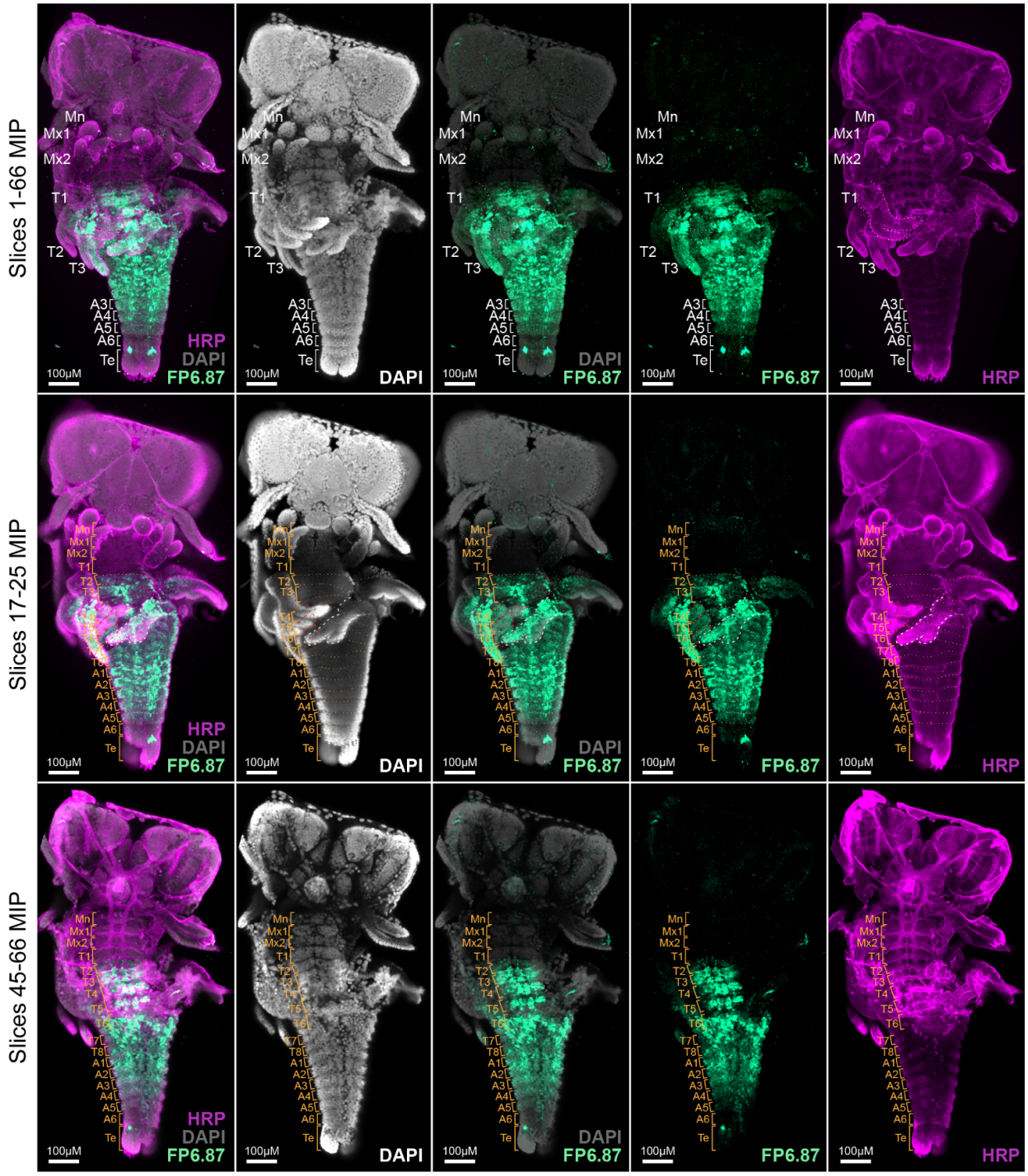


Figure 5.9: Z-slice analysis confirms unexpected register of *Ubx* + *abd-A* staining

Scale bars are exactly 100μM.

Top row: Maximum intensity projection (MIP) of slices 1-66 of embryo from top row of Fig. 4.8.

Middle row: MIP of slices 17-25. Segment boundaries are labeled in brackets.

Bottom row: MIP of slices 45-66. Segment boundaries are labeled in brackets.

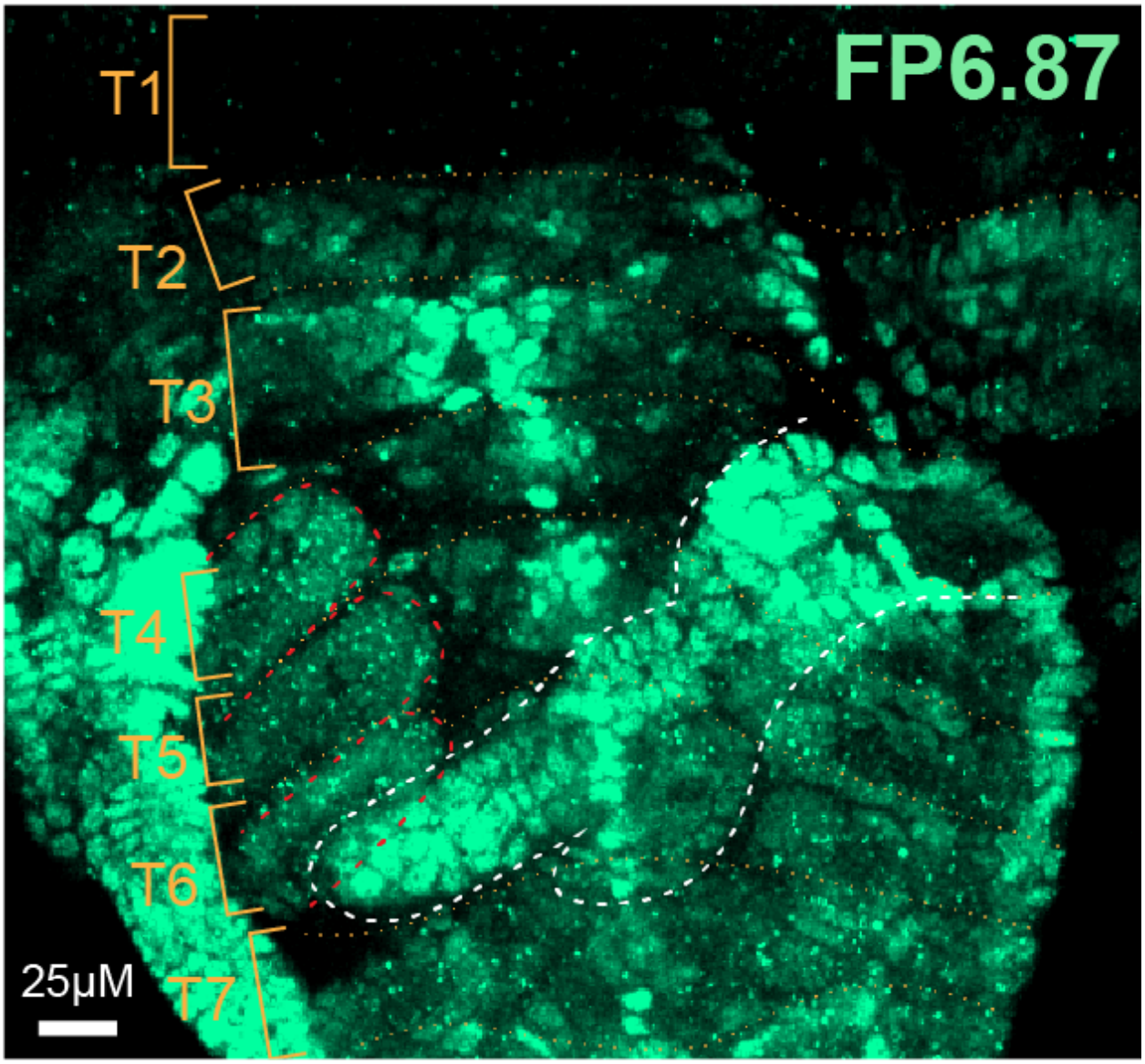


Figure 5.10: Thoracic limb buds show *Ubx* + *abd-A* staining

Scale bar is exactly 25μM.

Zoomed view of FP6.87 (*Ubx* + *abd-A*) expression in slices 17-25 from figure 4.9. Red dotted lines indicate outline of T4-6 thoracic limbs stained with FP6.87.

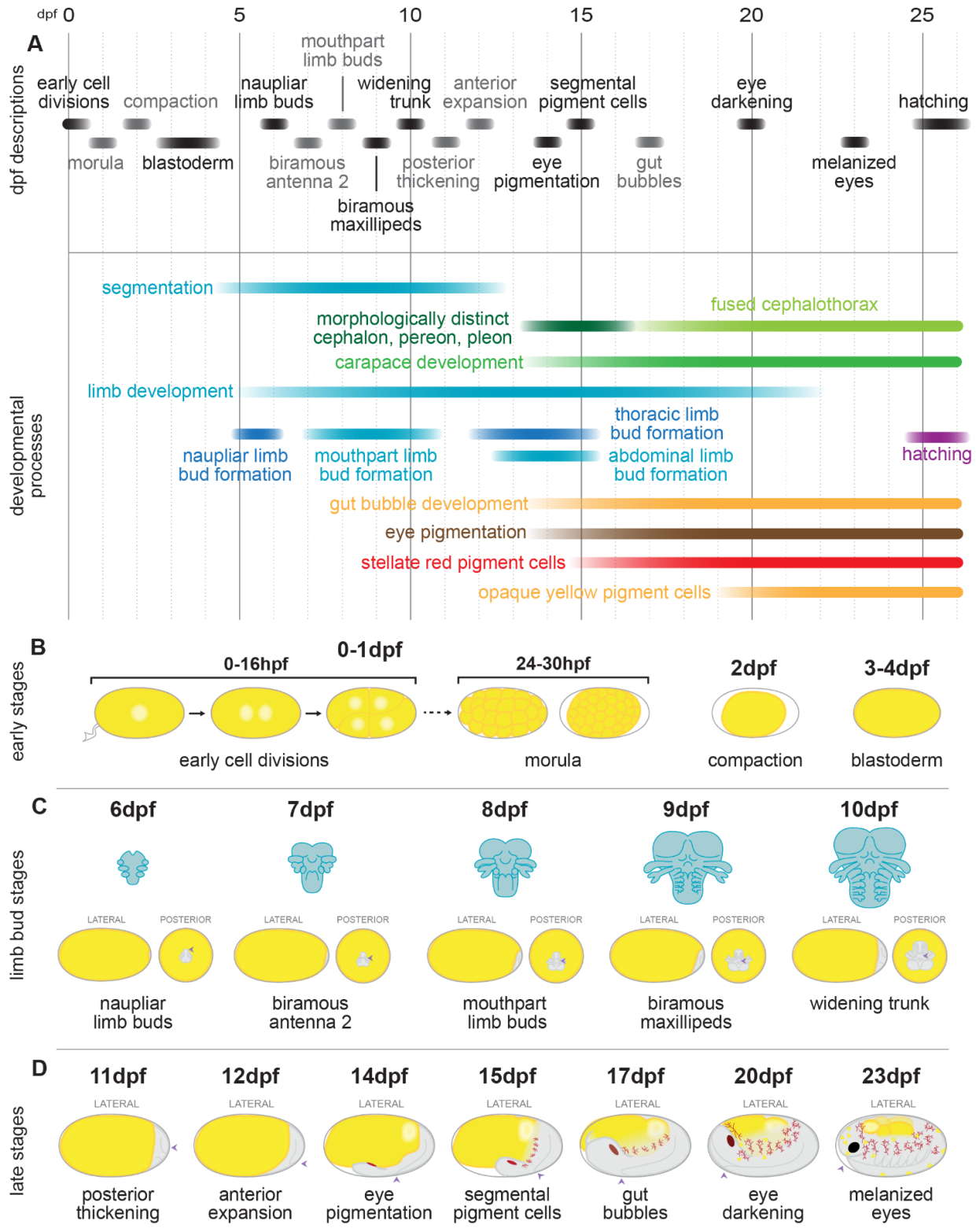

dpf 0
5
10
15
20
25
A
dpf descriptions
early cell divisions
compaction
naupliar limb buds
mouthpart limb buds
widening trunk
anterior expansion
segmental pigment cells
eye darkening
hatching
morula
blastoderm
biramous antenna 2
biramous maxillipeds
posterior thickening
eye pigmentation
gut bubbles
melanized eyes
developmental processes
segmentation
fused cephalothorax
morphologically distinct cephalon, pereon, pleon
carapace development
limb development
thoracic limb bud formation
naupliar limb bud formation
mouthpart limb bud formation
abdominal limb bud formation
hatching
gut bubble development
eye pigmentation
stellate red pigment cells
opaque yellow pigment cells
B
early stages
0-16hpf
0-1dpf
24-30hpf
2dpf
3-4dpf
early cell divisions
morula
compaction
blastoderm
C
limb bud stages
6dpf
7dpf
8dpf
9dpf
10dpf
LATERAL
POSTERIOR
naupliar limb buds
biramous antenna 2
mouthpart limb buds
biramous maxillipeds
widening trunk
D
late stages
11dpf
12dpf
14dpf
15dpf
17dpf
20dpf
23dpf
LATERAL
posterior thickening
anterior expansion
eye pigmentation
segmental pigment cells
gut bubbles
eye darkening
melanized eyes

Figure 5.11: Schematic summary of *Neocaridina* development

A) Schematic timeline of stages and developmental events during *Neocaridina* development.
B) Schematic of early phases of *Neocaridina* development from zygote to blastula-like embryonic stages, 0-4dpf.
C) Schematic of limb bud stages, including representations of embryonic morphology, from 6-10dpf.
D) Schematic lateral view of embryos from 11-23dpf, illustrating gross morphological movements of the embryo within the egg, stages of eye development, and emergence of pigment cell types.

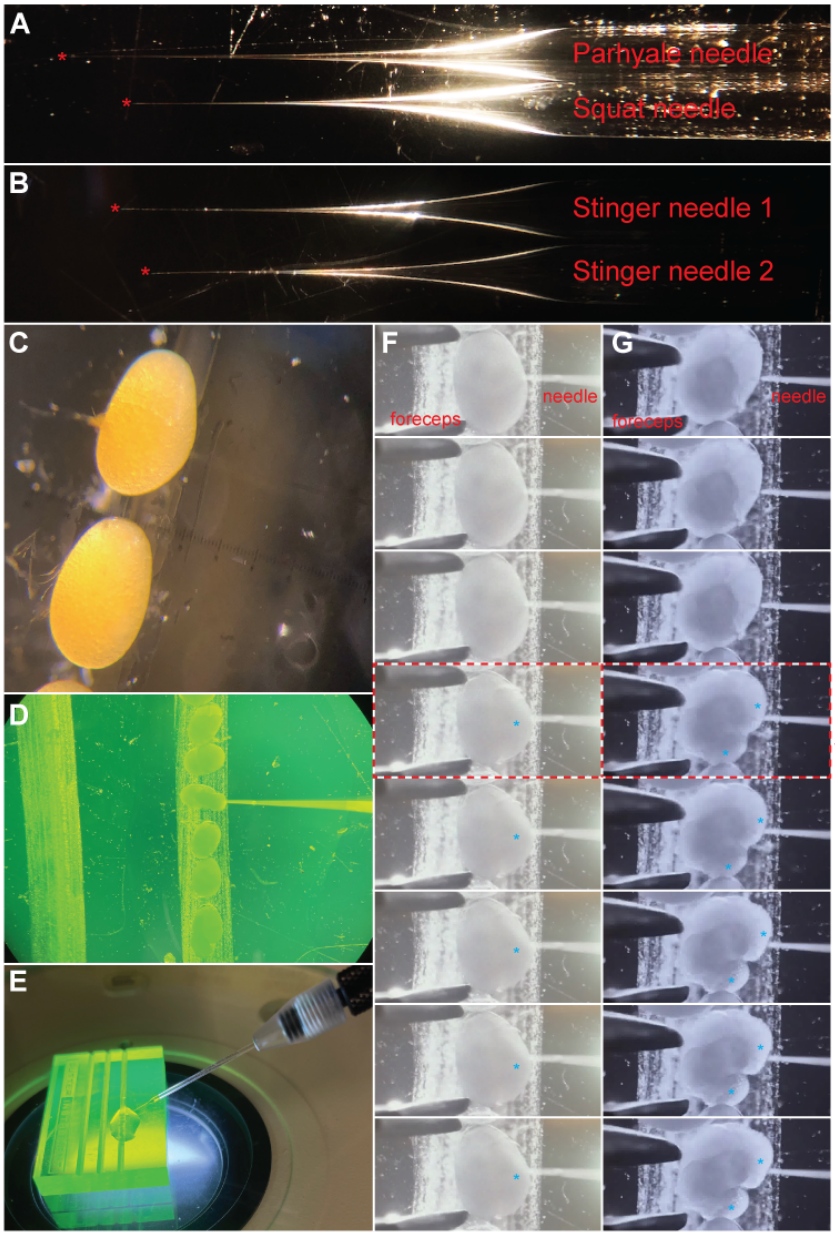


Figure 5.12: High intraembryonic pressure as a barrier to successful *Neocaridina* injections

A) *Parhyale* injection needle and "squat" style needle.

B) Two different "stinger" type needles generated with multiple rounds of heating and pulling.

C) 1- and 2-cell embryos being prepared for microinjection in a Sylgard dish.

D) Early blastomere embryos lined up for injection in a plastic needle holder.

E) Injection angle and setup for early stage embryos using a quartz needle.

F) Time-series of embryonic injection trials showing exploding embryo. Red dotted line marks frame in which the chorion is punctured. Blue asterisks mark escaping yolk or cells.

G) Time-series of embryonic injection trial in a different embryo than F.

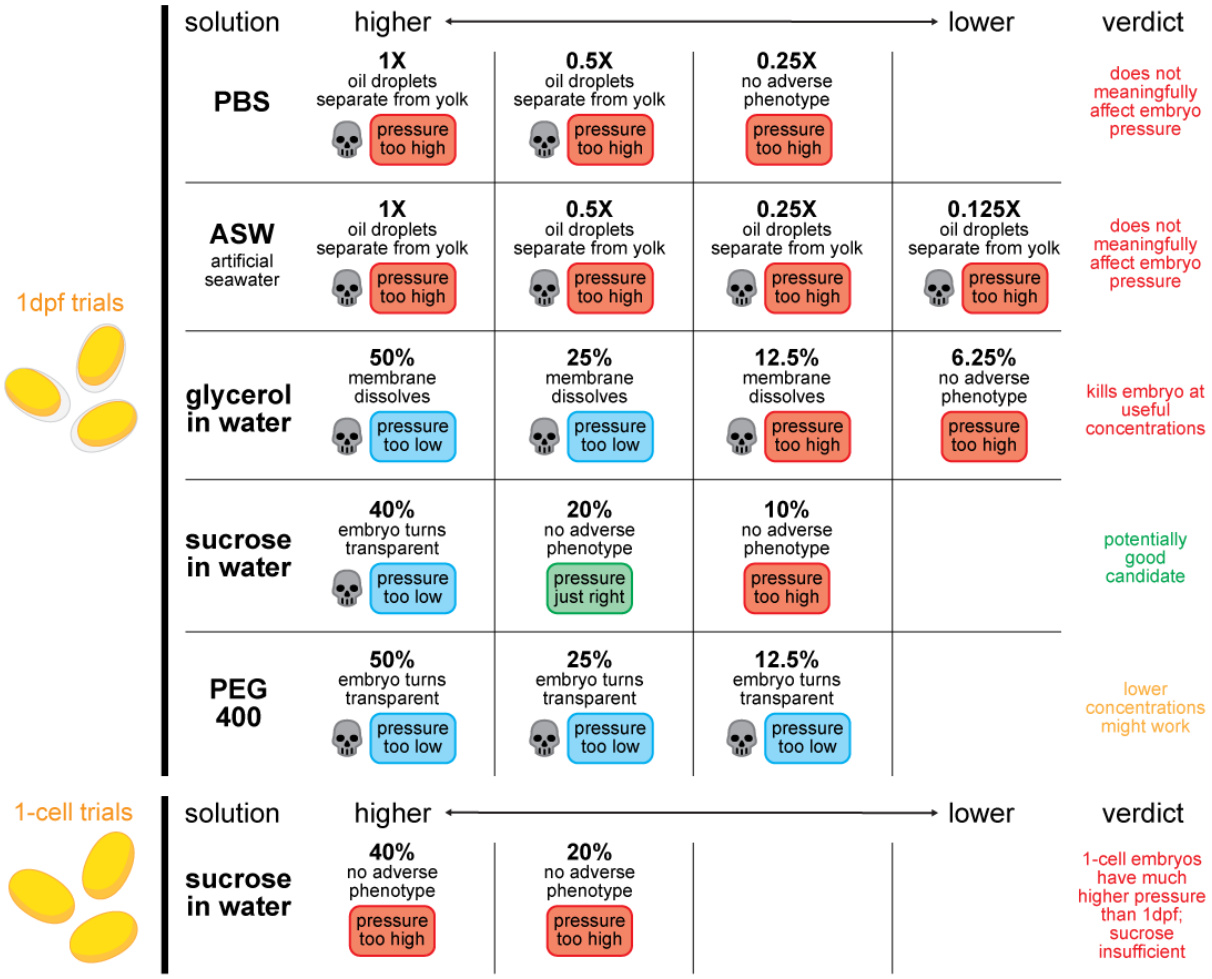


Figure 5.13: Effects of various osmotic solutions on early embryonic pressure and survival

Table showing effects of a variety of different solutions on embryonic survival and intraembryonic pressure.

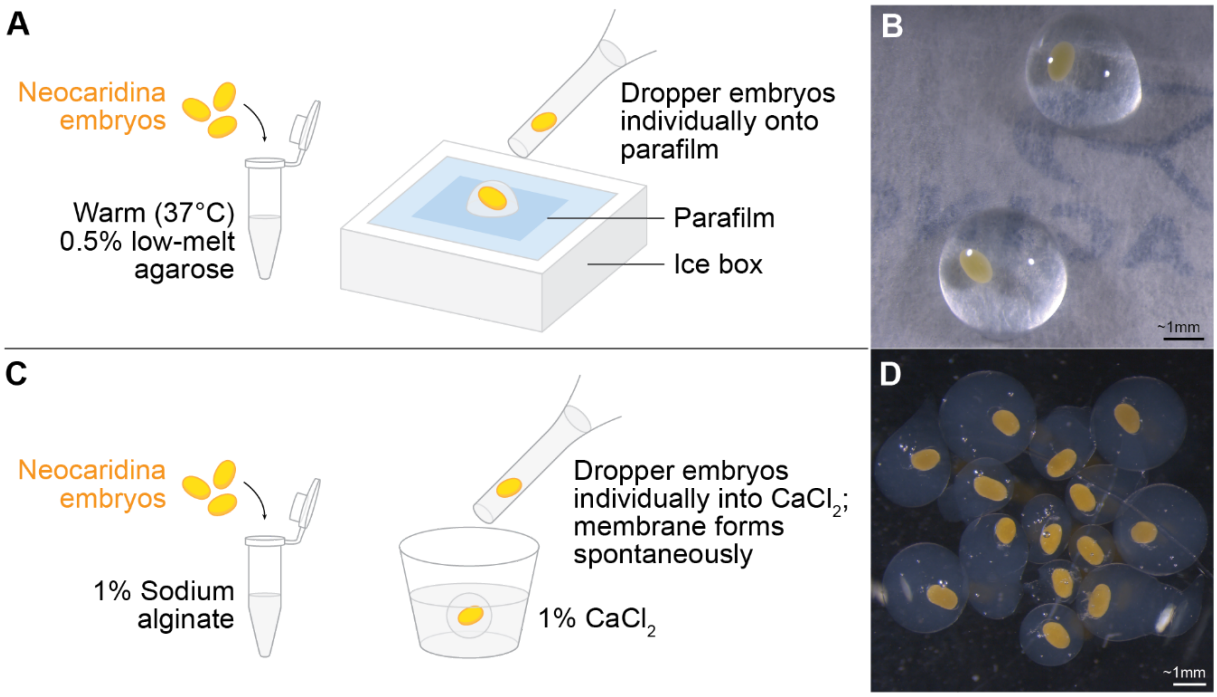


Figure 5.14: Methods to generate synthetic "membranes" for *Neocaridina* injection

Scale bars represent approximately 100μM.

A-B) Method to generate artificial membrane for embryos using agarose embedding.

C-D) Method to generate artificial membrane for embryos using sodium alginate.

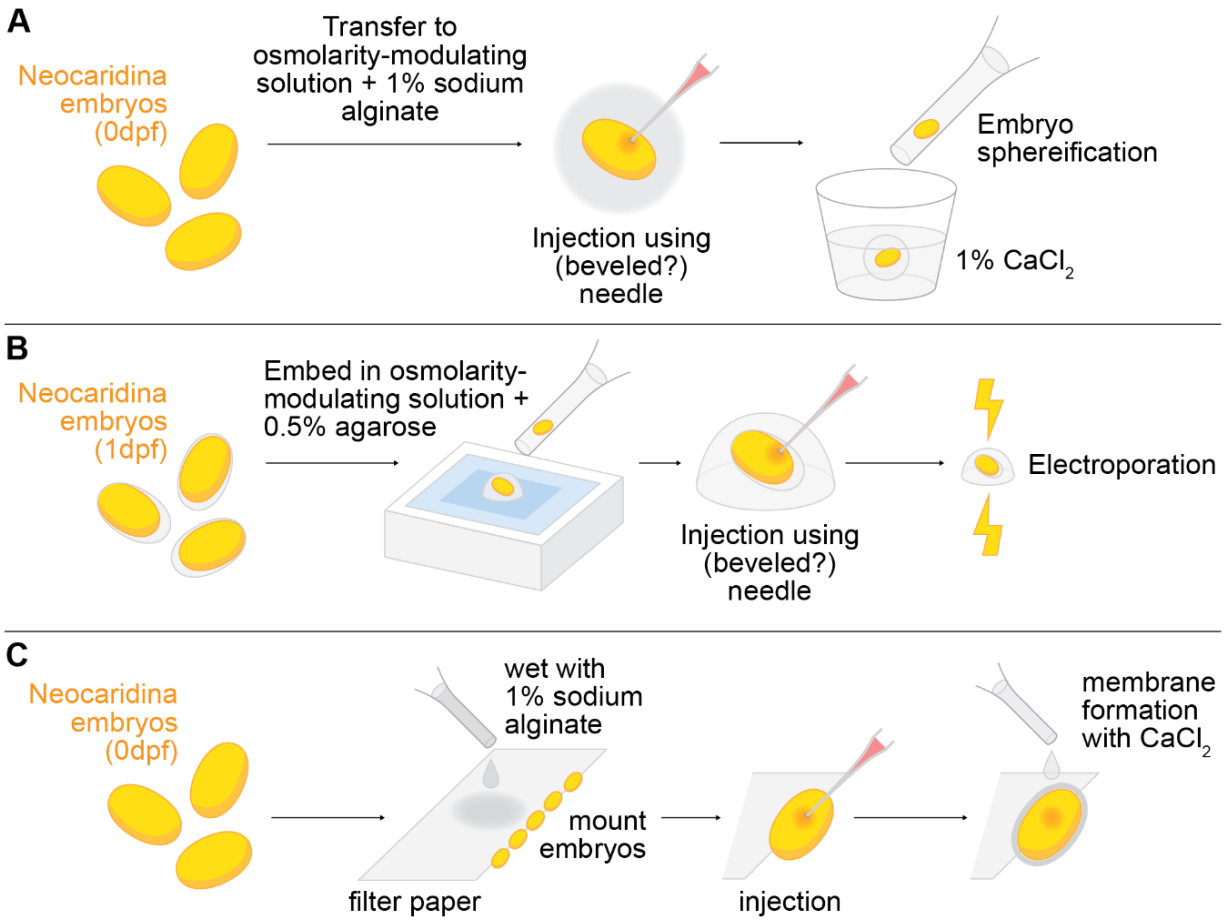


Figure 5.15: Proposed future strategies for embryonic injection in *Neocaridina*

A) Proposed approach combining sodium alginate with embryonic depressurization to improve injection success.

B) Proposed approach combining agarose embedding with electroporation to improve injection success.

C) Proposed approach combining sodium alginate methods described in current manuscript with Li et al. 2022 filter paper approaches to generate synthetic membranes rapidly after injection. This approach may improve embryo survival by decreasing leakage.