FORTH
INSTITUTE OF MOLECULAR BIOLOGY AND BIOTECHNOLOGY

Characterization of *Distal-less cis*-regulatory sequences across pancrustaceans

Graduate program in Molecular Biology and Biomedicine

Master Thesis

Ioannis (John) Rallis

Supervisor:

Anastasios (Tassos) Pavlopoulos

Developmental Morphogenesis Lab IMBB-FORTH

Heraklion, 2020

Contents

Acknowledgements

I would like to express my sincere gratitude towards:

Tassos Pavlopoulos, my MSc Thesis advisor who accepted me in his lab for my Master's Thesis and stood by my side as a true mentor throughout the whole year. Tassos introduced me to the world of *Parhyale* and *Evo-Devo* and helped me develop as a future scientist. He permitted me to take initiative, motivation to study the literature, and actively assisted me with grant applications and oral presentations. Beyond that, Tassos was always there to support me in situations inside and outside the lab, and for that, I am thankful.

Michalis Averof, who inspired me to follow the research field of *evo-devo*, supported my research by sharing the ATAC-seq genomic datasets and providing meaningful comments on my project.

Valia Stamataki, the post-doc in our lab, who helped me acquire all the technical skills needed for *Parhyale* research, helped me with some of my experiments, and gave me fruitful advice regarding my projects and helpful comments on my oral presentations. Apart from that, together with Tassos, she supported me in situations outside the lab, and I thank her for that.

Marina Ioannou, our lab manager who supported me with experimental, technical, and organizational matters.

Ioannis Livadaras, for conducting the *Drosophila* injections.

Gentjan Kapaj, Themis Archontidis, Eirini Daskalaki, Myrto Ziogas, and Georgia Chalkia for the moments we spent inside and outside the lab.

Christina Efraimoglou, my partner in life, for tolerating me for the past years, supporting my decisions, and commenting and editing my thesis, oral presentations, and grant proposals.

IMBB-FORTH and Bodossaki Foundation for their fellowships.

The other members of my committee Christoforos Nikolaou and Ioannis Iliopoulos

Summary

Macroevolutionary comparisons between species which are separated by large evolutionary timescales have focused on addressing and highlighting similarities in developmental genetic programs. One such striking case is the finding that representatives of all Bilaterian clades, use a common genetic toolbox for patterning their outgrowing limbs along the three body axes [anterio-posterior (AP), dorso-ventral (DV), proximo-distal (PD)] (Pueyo and Couso, 2005; Tarazona *et al.*, 2019). So far, comparative developmental studies in species that are not established as model systems were limited to expression analyses of several candidate genes whose role was extensively investigated in the major developmental genetic model organisms. Only rarely, unbiased functional studies have been conducted in a small number of emerging non-model species that have proved to be genetically tractable and amenable to an expanding palette of experimental manipulations. One of them is the crustacean amphipod *Parhyale hawaiensis*, which we use in the lab to get insights into the decades-old problem of the conservation and divergence of limb patterning mechanisms across the arthropods (Wolff *et al.*, 2018; Pavlopoulos and Wolff, 2020) .

During my Master's thesis, I investigated the genetic basis of radical changes in the developmental timing of limb patterning mechanisms between direct and indirect developing pancrustacean limbs. In particular, I made a comparative study of the *cis*-regulatory sequences of a very important and highly conserved limb patterning gene, *Distal-less* (*Dll*), between the insect *Drosophila melanogaster* and the crustacean *Parhyale hawaiensis*. *Parhyale* has a direct embryonic limb development, whereas *Drosophila* limb development is indirect over embryonic, larval and pupal development. In this study, I identified putative *Dll* cis-regulatory elements in *Parhyale* using ATAC-seq data from *Parhyale* embryos and phylogenetic foot-printing between the *Parhyale* and the few other available amphipod genomes. Functional analysis of the identified *cis*-regulatory elements using reporter constructs revealed an enhancer mediating the early expression of *Dll* in the *Parhyale* limb buds. Bioinformatics analyses for overrepresented motifs in the *Parhyale Dll* enhancer led to the discovery of two putative binding sites for known *Dll* regulators in *Drosophila*. In a

complementary approach, I tested in *Parhyale* previously characterized *Drosophila Dll* enhancers with reporter constructs to investigate their activity pattern and timing of expression. These functional cross-species comparisons of *Dll cis*-regulatory sequences and their upstream regulators provided some first insights into the conservation and divergence of patterning mechanisms between direct and indirect developing pancrustacean limbs.

Περίληψη

Οι μακροεξελικτικές συγκρίσεις μεταξύ ειδών τα οποία χωρίζονται από μεγάλης κλίμακας εξελικτικούς χρόνους επικεντρώνονται κατά κύριο λόγο στις ομοιότητες των αναπτυξιακών γενετικών προγραμμάτων. Ένα χαρακτηριστικό παράδειγμα είναι ότι αντιπρόσωποι όλων των κλάδων των αμφίπλευρα συμμετρικών ζώων χρησιμοποιούν κοινά γονίδια προκειμένου να καθορίσουν το πρότυπο ανάπτυξής τους κατά τους τρεις σωματικούς άξονες (εμπροσθο-οπίσθιο, ραχο-κοιλιακό, εγγύ-απομακρυσμένο). Μέχρι στιγμής οι συγκριτικές αναπτυξιακές μελέτες σε είδη τα οποία δεν αποτελούν οργανισμούς μοντέλα περιορίζονται σε αναλύσεις της έκφρασης υποψηφίων γονιδίων των οποίων ο ρόλος έχει μελετηθεί εκτενώς στα καθιερωμένα μοντέλα αναπτυξιακής γενετικής. Σε πολύ λίγες περιπτώσεις, έχουν διεξαχθεί λειτουργικές μελέτες σε αναδυόμενους οργανισμούς μοντέλα που είναι γενετικά προσβάσιμοι χρησιμοποιώντας ένα ευρύ φάσμα πειραματικών τεχνικών. Ένα από αυτά τα αναδυόμενα μοντέλα είναι το αμφίποδο καρκινοειδές *Parhyale hawaiensis*. Στο εργαστήριο χρησιμοποιούμε το *Parhyale* μεταξύ άλλων για τη μελέτη της συντήρηση ή απόκλισης των μηχανισμών καθορισμού του αναπτυξιακού προτύπου των άκρων κατά την εξέλιξη των αρθροπόδων.

Κατά τη διάρκεια της μεταπτυχιακής μου εργασίας, μελέτησα τη γενετική βάση των χρονικών αλλαγών που παρατηρούνται στους μηχανισμούς καθορισμού του προτύπου ανάπτυξης άκρων μεταξύ πανκαρκινοειδών με άμεση και έμμεση ανάπτυξη άκρων. Συγκεκριμένα, έκανα μία συγκριτική μελέτη των *cis*-ρυθμιστικών αλληλουχιών ενός σημαντικού και εξελικτικά συντηρημένου γονιδίου που συμμετέχει στον καθορισμό των άκρων, του *Distal-less (Dll)*, μεταξύ του εντόμου *Drosophila melanogaster* και του καρκινοειδούς *Parhyale hawaiensis.* Το *Parhyale* εμφανίζει άμεση ανάπτυξη άκρων κατά την

εμβρυογένεση, ενώ στη *Drosophila* η ανάπτυξη των άκρων γίνεται έμμεσα κατά τη διάρκεια του εμβρύου, της προνύμφης και της νύμφης. Σε αυτή τη μελέτη ταυτοποίησα πιθανά *cis*-ρυθμιστικά στοιχεία του *Dll* στο *Parhyale* χρησιμοποιώντας δεδομένα από ATAC-seq πειράματα σε έμβρυα και συγκρίνοντας το φυλογενετικό αποτύπωμα μεταξύ του γονιδιώματος του *Parhyale* και άλλων διαθέσιμων γονιδιωμάτων αμφιπόδων. Η ανάλυση της λειτουργικότητας αυτών των *cis*-ρυθμιστικών στοιχείων του *Dll* με τη χρήση γονιδίων αναφοράς αποκάλυψε έναν ενισχυτή που κατευθύνει την έκφραση του γονιδίου *Dll* στα πρώιμα άκρα του *Parhyale*. Βιοπληροφορικές αναλύσεις για υπερεκπροσώπηση μοτίβων DNA στους *Parhyale Dll* ενισχυτές οδήγησαν στην ανακάλυψη δύο πιθανών θέσεων πρόσδεσης μεταγραφικών παραγόντων για τους οποίους γνωρίζουμε ότι ρυθμίζουν το *Dll* στη *Drosophila.* Συμπληρωματικά, έλεγξα με τη χρήση γονιδίων αναφοράς στο *Parhyale* τους ήδη χαρακτηρισμένους ενισχυτές του *Dll* της *Drosophila*, προκειμένου να διερευνήσω τη λειτουργικότητά τους και το χωροχρονικό πρότυπο έκφρασής τους. Αυτές οι λειτουργικές συγκρίσεις των *Dll cis*-ρυθμιστικών αλληλουχιών μεταξύ ειδών προσέφεραν τις πρώτες πληροφορίες σχετικά με την εξέλιξη των μηχανισμών καθορισμού προτύπου μεταξύ πανκαρκινοειδών με άμεση και έμμεση ανάπτυξη άκρων.

Introduction

Parhyale hawaiensis as a model system

The biology of *Parhyale hawaiensis*

Parhyale hawaiensis is a marine amphipod crustacean that was introduced in the lab by Browne and Patel in the late 1990s (Browne *et al.*, 2005) to enable broader comparative studies of arthropod development. Phylogenetically, *Parhyale hawaiensis* serves both as an outgroup for insects and as a representative member of crustaceans (Stamataki and Pavlopoulos, 2016). Crustaceans form a speciose, diverse and ancient arthropod group of organisms that originated during the Cambrian era. In addition, *Parhyale* is a malacostracan crustacean, which makes it a close relative to commercially important organisms such as shrimps, lobsters, and crabs (Sun and Patel, 2019).

Parhyale has a global distribution. It is a detritvorous animal that lives in tropical shallow marine ecosystems and has a high tolerance for environmental osmotic changes. This lifestyle makes *Parhyale* a robust experimental organism that thrives under standard laboratory conditions. *Parhyale* has a life cycle of 8 weeks at 26^{o}C. Embryogenesis lasts about 10 days and the juvenile that hatches from the egg looks like a miniature of the adult (Browne *et al.*, 2005). Thanks to this mode of direct development, the majority of the body patterning, growth and differentiation mechanisms (with the exception of a few traits associated with sexual maturation) can be studied during the 10 days of embryogenesis. Its also offers many advantages for experimentation, because *Parhyale* is genetically and optically tractable, it is supported by an ever increasing suite of embryological, genetic, genomic and imaging resources, and its embryogenesis is well-described and comprehensively staged (Browne *et al.*, 2005; Kao *et al.*, 2016).

Sexually mature adult females are distinct from males by their easily detectable gonads and their smaller grasping appendages in the thorax. Male *Parhyale* seize and retain hold of the female until copulation ends. During copulation, the

female gets released from the male by molting and oviposits the fertilized eggs in her ventral brood pouch where the male deposits its sperm (Browne *et al.*, 2005). In the lab, we take advantage of these ethological features to streamline the collection of embryos from gravid females (Kontarakis and Pavlopoulos, 2014). Thousands of animals can be raised routinely in small containers and, since adults breed year-round, hundreds of fertilized eggs can be obtained daily for experimental manipulations.

The *Parhyale* body plan

The body of *Parhyale*, consists of 19 appendage-bearing segments that are organized into the head, thorax (pereon) and abdomen (pleon), similar to the rest malacostracans (Stamataki and Pavlopoulos, 2016). The body is laterally

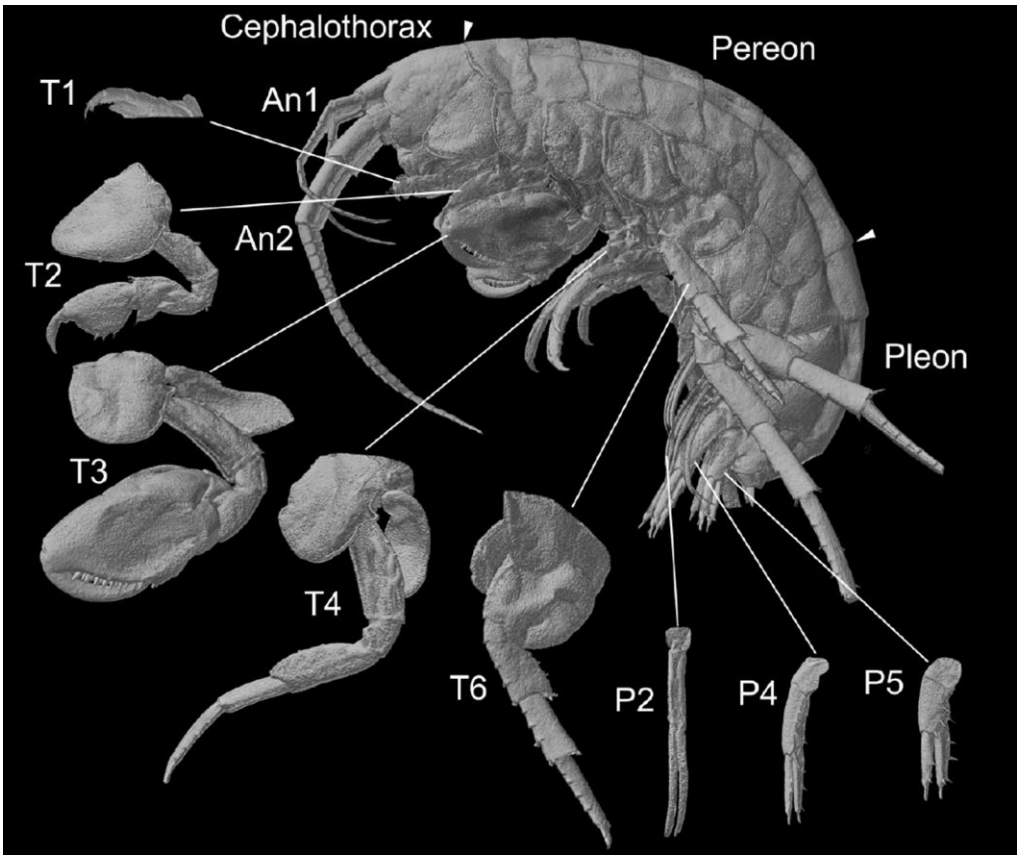

Figure 1: Parhyale body plan and appendage diversity. *Lateral view of adult Parhyale imaged with microcomputed tomography with anterior towards the left and dorsal towards the top. The cephalothorax consists of six segments and has two pairs of sensory antennae (An1 and An2) and three pairs of feeding appendages (not shown). The first thoracic segment (T1) is fused to the head and has a pair of feeding appendages, the maxillipeds. The thorax (pereon) consists of the T2 and T3 appendages, which are used for grasping and are called gnathopods, and the T4 to T8 walking appendages called pereopods. The abdomen (pleon) consists of six segments with three pairs of swimming appendages called pleopods (P1 to P3), followed by three pairs of thickened appendages, called uropods, used for anchoring and jumping (P4 to P6). Image reproduced from Pavlopoulos and Wolff, 2020).*

compressed with the anterior thoracic limbs oriented forward and the posterior thoracic limbs backwards, giving amphipods their characteristic appearance and name. Many malacostracan groups, including amphipods, exhibit a tremendous specialization in their appendages that have been adapted for different functions like sensation, feeding, locomotion and others (VanHook and Patel, 2008). The head is composed of the compound eyes, two pairs of sensory organs (antennae 1 and 2) and three pairs of feeding appendages (mandibles, maxillae 1, maxillae 2). The head is fused to the first thoracic segment (T1) that has a pair of feeding appendages, known as maxillipeds. The remaining seven thoracic segments (T2–T8) develop larger segmented appendages. T2 and T3 form subchelate grasping appendages, called gnathopods, and T4–T8 bear walking appendages called pereopods. The abdomen is divided into two parts with three segments each: the pleon bearing three pairs of biramous swimming appendages (A1–A3 pleopods) and the urosome with three small pairs of thickened appendages (A4–A6 uropods) (Fig. 1).

The early cleavages of fertilized eggs are holoblastic (Fig. 2). The first cell division is slightly unequal and divides at this early stage the left from the right side of the animal (for most of the ectodermal and mesodermal lineages). The second cell division is also slightly unequal and the third cleavage is highly unequal, resulting in a stereotypic arrangement of four macromeres and four micromeres that are uniquely identified by their relative position, contacts and size. Each of these blastomere is committed to a particular germ layer (Fig. 2, 8 hrs). The Mav macromere will give rise to the visceral and anterior mesoderm, the remaining El, Er and Ep macromeres to the ectoderm, the en micromere to the endoderm, the g micromere produces the germline, and the two mr and ml micromeres will give rise to the somatic mesoderm (Gerberding, Browne and Patel, 2002)(Browne *et al.*, 2005). The later asymmetric cell divisions separate the cells from the yolk (Fig. 2, 12 hrs) and cells come together to form the embryonic rudiment (Fig. 2, 2 days). Ectodermal cells organize themselves into the two head lobes anteriorly and into aligned rows of cells posteriorly, creating a grid-like structure which is typical for amphipods and other malacostracan

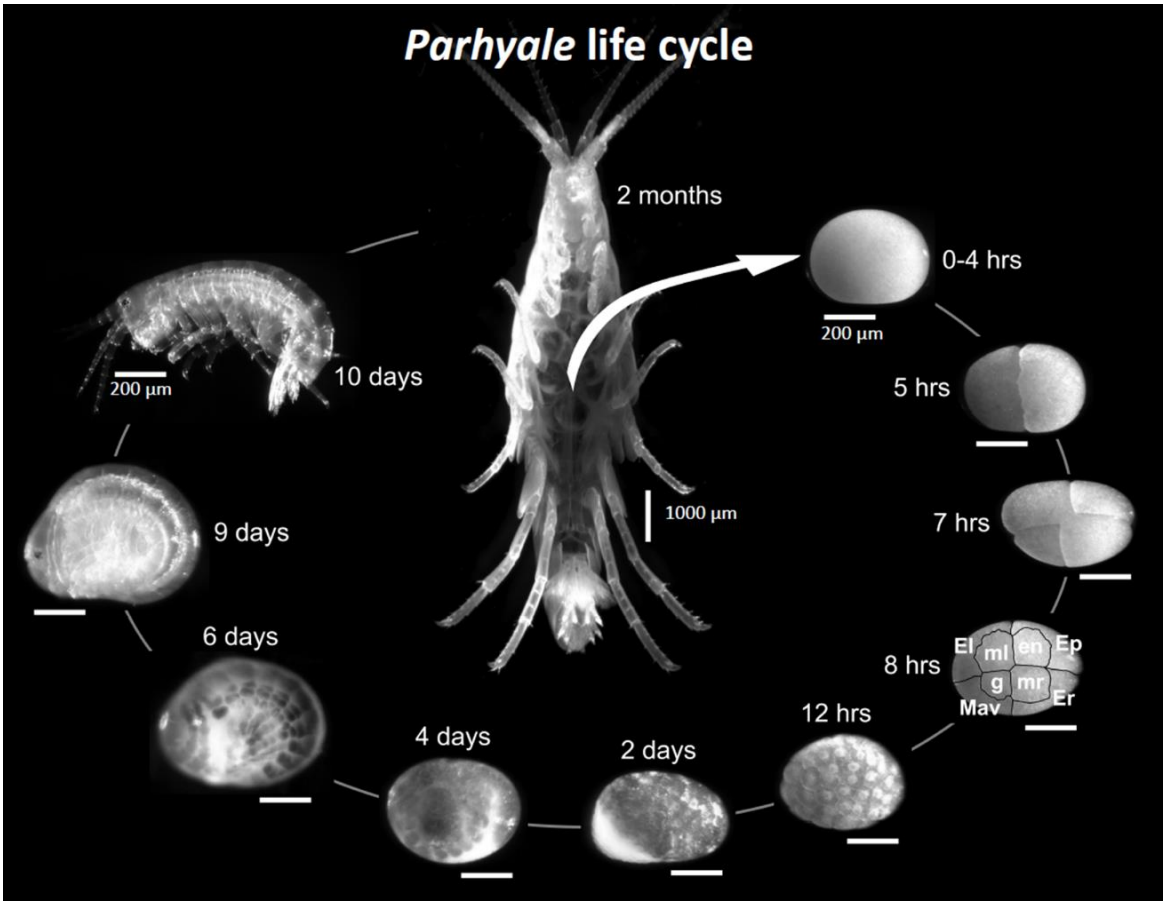

***Figure 2: Parhyale life cycle**. Parhyale eggs can be dissected from the female's ventral brood pouch. During the first 8 hours after egg lay, each egg devides three times producing a stereotyped arrangement of four macromeres and four micromeres with restricted cell fates. Later divisions produce cells that aggregate ventrally and anteriorly to form the embryo rudiment (2 days). Next, the embryo elongates posteriorly and the appendage buds develop in an anterior to posterior progression (4 days). Appendages continue to grow, the yolk gets sequestered in the developing gut and the head region separates from the rest of the body (6 days). All organs have completed thei development during the last days of embryogenesis and the compound eyes become pigmentated (9-10 days). The hatched juvenile looks like a miniature version of the adult (10 days). The hatchling grows via successive molts and becomes sexually maturate about 2 months after egg lay. All scale bars are 200 mm except in the adult female that is 1000 mm. Image reproduced from Stamataki and Pavlopoulos, 2016.*

embryos. These cell rows exhibit a parasegmental organization, like in *Drosophila* and other arthropods, with each row in the grid corresponding to one parasegment. *Parhyale* patterning and growth occur in an anterior-to-posterior progression with the more anterior structures forming earlier than the more posterior ones (similar to vertebrate somitogenesis).The parasegmental rows undergo two rounds of stereotypical longitudinal cell divisions, that together with the sequential addition of more rows posteriorly, lead to the elongation of the body axis of the embryo (Fig. 2, 2-6 days) (Dohle and Scholtz, 1988; Gerberding, Browne and Patel, 2002; Browne *et al.*, 2005). Later cell divisions have a more complex pattern, break the uniformity of the grid and lead to the formation of segmental units and appendage bud outgrowths. During the mid to

late stages of embryogenesis (Fig. 2, 4-9 days), the appendages extend and differentiate along their proximo-distal (PD) axis and the yolk gets sequestered into the developing gut. At the end of the embryogenesis (Fig. 2, 9-10 days), the compound eyes in the head form and get pigmented, the heart along the dorsal midline starts beating and the muscles start twitching before hatching (Browne *et al.*, 2005) (Fig. 1).

The evolution of malacostraca like *Parhyale* into living Swiss army knives with their tremendous specialization of appendages has no equal among metazoans, and offers excellent material to investigate the molecular, cellular and biophysical basis of organ morphogenesis (Browne *et al.*, 2005; Pavlopoulos *et al.*, 2009; Martin, Julia M. Serano, *et al.*, 2016; Stamataki and Pavlopoulos, 2016).

Experimental tools and resources for *Parhyale* research

Over the last two decades, the *Parhyale* community has developed a palette of experimental approaches and standardized resources that have advanced *Parhyale* into a powerful system to address fundamental questions in developmental biology (Fig. 3). *Parhyale* embryos are amenable to various embryological manipulations and gene expression studies using whole-mount *in-situ* hybridization and immunohistochemistry (Gerberding, Browne and Patel, 2002; Extavour, 2005; Rehm, Hannibal, C. R. Chaw, *et al.*, 2009; Rehm, Hannibal, R. C. Chaw, *et al.*, 2009). To facilitate functional genetic research, many genomic and transcriptomic resources were first generated by high-throughput sequencing of BAC clones and cDNA libraries (Parchem *et al.*, 2010). Recently, the *Parhyale* genome, that resembles the human genome in terms of size, chromosome count and heterogeneity, was sequenced, assembled *de-novo* and annotated (Kao *et al.*, 2016).

Transgenesis in *Parhyale* was first achieved using a member of the mariner Tc1 family of tranposable elements, namely the *Minos* transposon from *Drosophila hydei* (Franz and Savakis, 1991; Pavlopoulos and Averof, 2005) that is active in a large variety of animal systems. Engineered transposons consist of the two *Minos* terminal inverted repeats flanking a desired transgene and a marker gene

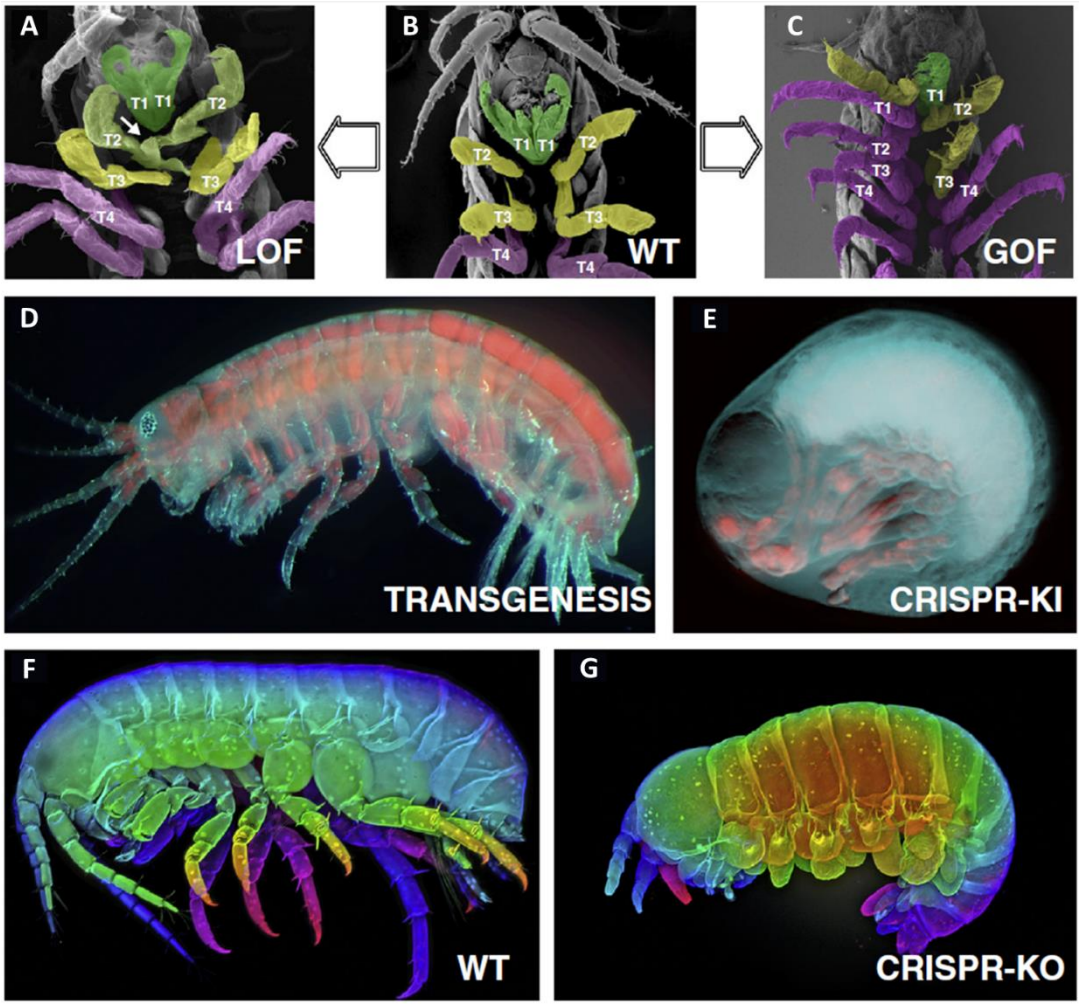

***Figure 3**: **Functional genetic tools in Parhyale. (A-C**) Loss-Of-Function and Gain-Of-Function studies of Hox genes. Scanning electron microscopy of juvenile Parhyale with transformed appendages after RNAi knocked-down (LOF in A) or heat-inducible misexpression (GOF in C). Transformed appendages are color-coded as in the wild-type in B. **(D)** Transgenic animal expressing a muscle-specific fluorescent reporter (in red) merged with the auto-fluorescence of the cuticle (in cyan). **(E)** CRISPR-mediated knock-in of a fluorescent reporter in the Distal-less locus marking developing appendages (shown in red) merged with the corresponding bright field image (in cyan). **(F)** Wild-type juvenile stained for cuticle and color-coded by depth, and **(G)** similarly stained mutant with truncated appendages after CRISPR-based knock-out of Distal-less. Image reproduced from Stamataki and Pavlopoulos, 2016.*

for detection of transgenic *Parhyale* (Fig. 3D). In transgenesis experiments, engineered transposons are injected together with mRNA encoding the *Minos* transposase into 1- or 2-cell *Parhyale* embryos. The mobilized transposons are randomly inserted into the genome of injected embryos producing with high efficiency G0 adults (founders) with transformed germlines (Pavlopoulos and Averof, 2005). In all transgenesis experiments, transgene expression is detectable not only in transgenic animals of G1 and G2 generations onwards, but also in a large fraction of injected G0 embryos that exhibit transgene expression with various levels of mosaicism. This aspect is very useful for experimentation in *Parhyale*, because it enables fast and reliable G0 genetic

approaches and, due to the early lineage restrictions, the comparison of the wild-type versus the genetically perturbed conditions in the same embryo. The transgenesis toolkit in *Parhyale* was further expanded with the ΦC31 site-specific integrase system (Kontarakis, Pavlopoulos, *et al.*, 2011). The establishment of transposon and integrase-based transformation systems has increased the sophistication and versatility of genetic manipulations in *Parhyale* with unbiased gene trapping screens and the redeployment of gene traps for various applications (Kontarakis, Konstantinides, *et al.*, 2011). The characterization of endogenous heat-inducible promoters allowed the development of conditional misexpression systems for gain-of-function genetic studies (Pavlopoulos *et al.*, 2009), while RNA interference and morpholino-mediated gene knock-down were employed for complementary loss-of-function approaches (Liubicich *et al.*, 2009) (Fig. 3A-C). Gene knock-down approaches in *Parhyale* have many limitations, like transient and incomplete reduction in gene function. Recently, the application of CRISPR/Cas system for targeted genome editing has helped to overcome those limitations (Gilles, Schinko and Averof, 2015). Specifically, the CRISPR/Cas system has been adapted to completely knock-out gene function in *Parhyale* embryos, as well as for knock-in approaches to generate live fluorescent reporters of gene expression (Kao *et al.*, 2016; Martin, Julia M. Serano, *et al.*, 2016; Serano *et al.*, 2016) (Fig 3E-G). Similar to all previously developed functional genetic manipulations, the complete cleavage mode and the slow tempo of cell divisions in early *Parhyale* embryos results in high mutagenesis efficiency and low levels of mosaicism in G0 embryos (Kao *et al.*, 2016; Martin, Julia M. Serano, *et al.*, 2016).

Finally, *Parhyale* has stood up to the challenge of making the link between the genetic and cellular basis of development. The advent of genetic tools for live imaging, in combination with the optical properties of *Parhyale* embryos, have allowed detailed microscopic inspections of cellular dynamics with exceptional spatiotemporal resolution. The eggshell is transparent and embryos are approximately 500 μm in length and exhibit low autofluorescence and light scattering. Early embryogenesis, including gastrulation and germband formation, can be live imaged with transmitted light and fluorescence time-lapse microscopy (Alwes, Hinchen and Extavour, 2011; Chaw and Patel, 2012;

Hannibal, Price and Patel, 2012). For longer time-lapse recordings of early, mid and late embryogenesis, transgenic embryos with fluorescently labeled nuclei can be imaged for several days using Light-sheet Fluorescence Microscopy (LSFM). LSFM has enabled to study appendage development in intact developing *Parhyale* embryos at very high spatial and temporal resolution, for several days or even a week, and with minimal photo-bleaching and photo-damage (Wolff *et al.*, 2018). In addition, *Parhyale* embryos can be optically sectioned from multiple angles (multi-view-LSFM) that are combined computationally to reconstruct the whole sample with fairly isotropic resolution (Fig. 4).

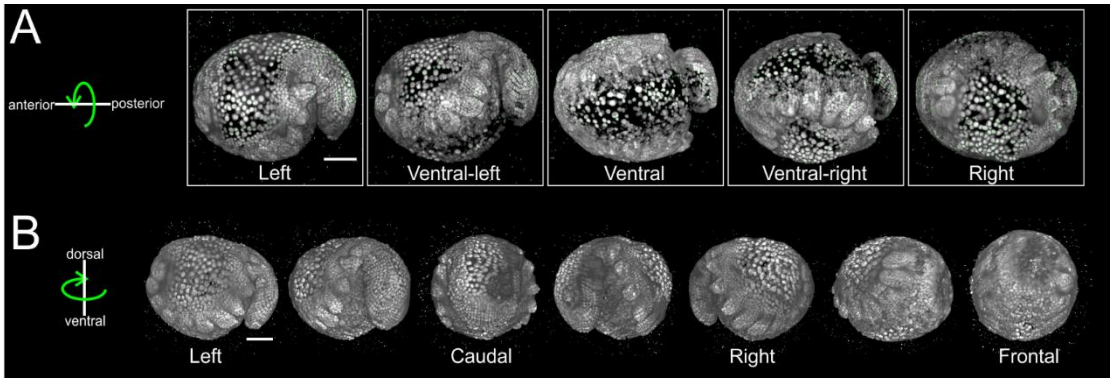

***Figure 4*: Reconstruction of Parhyale embryogenesis with multi-view Light-Sheet-Fluorescent-Microscopy. (A)** *Transgenic Parhyale embryo with fluorescent nuclei imaged from 5 views with 45˚ rotation around the AP axis between neighboring views. Panels show renderings of the acquired views with anterior to the left.* **(B)** *Raw views were registered and fused into an output image rendered here in different positions around the DV axis. Image reproduced from Wolff et al., 2018.*

Genetic basis of arthropod appendage development

The early mechanisms that specify the nascent limb primordia in the young embryo vary dramatically within arthropods (Angelini and Kaufman, 2005b). So far, *Drosophila* has been the dominant model used to understand arthropod appendage development. In *Drosophila*, limb specification, patterning, growth and differentiation processes take place at different developmental stages. Limb primordia are specified as small thoracic clusters of cells during early embryogenesis (Fig. 5A-B) (Cohen, Simcox and Cohen, 1993). These clusters form sac-like structures in the larva, called imaginal leg discs, that grow and undergo extensive patterning during larval stages establishing the adult limb and body wall fates (Fig. 5C-E). During pupal metamorphosis, the discs evert and telescope out (similar to a radio antenna) forming the cylindrical adult legs

protruding from the adult body wall (Fig. 5F) (von Kalm, Fristrom and Fristrom, 1995). Contrary to *Drosophila*, in most other insect and arthropod groups, appendages develop as direct three-dimensional outgrowths from the embryonic body wall, similar to the vertebrates. Embryos hatch with functional limbs that are in many cases miniature versions of the adult limbs.

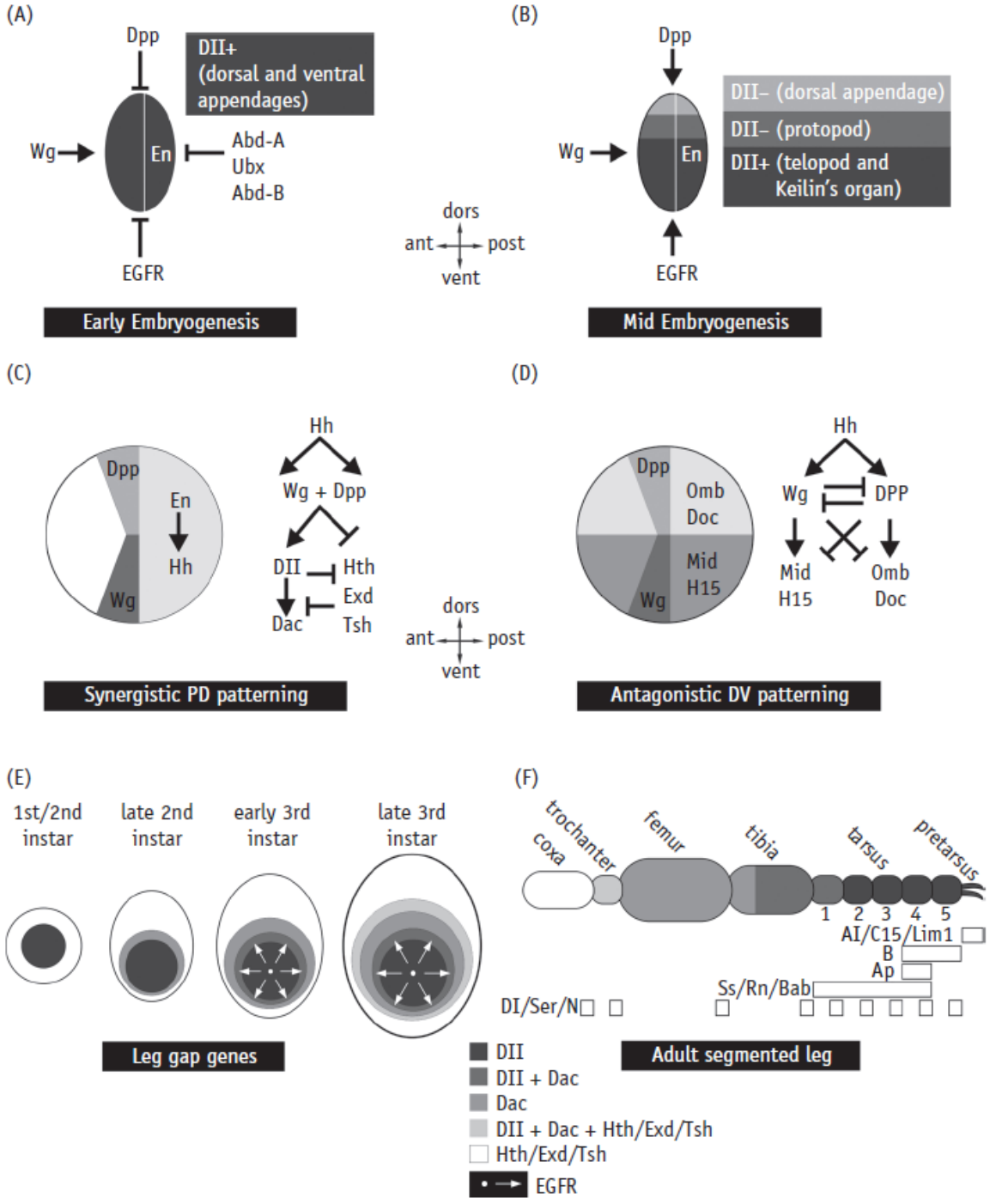

***Figure 5**: **The genetic basis of** **Drosophila** **leg development.** Schematic representations of **(A)** the early specification and **(B)** the proximo-distal subdivision of the embryonic limb primordium, **(C)** the proximodistal patterning, **(D)** dorso-ventral patterning and **(E)** leg gap gene expression in the leg imaginal disc, and corresponding subdivisions in **(F)** the adult segmented leg. It is likely that A and B are Drosophila innovations and are not applicable to crustacean limb development. The patterning mechanisms for early limb specification and subdivision in crustaceans are unclear. Image reproduced from Pavlopoulos and Wolff, 2020.*

Comparative studies have shown that many of the genes that orchestrate appendage development in *Drosophila* have similar roles in crustacean and arthropod appendages (Angelini and Kaufman, 2005a). One of those key limb patterning genes is *Distal-less (Dll)*, which encodes a homeo-domain transcription factor with a conserved role in limb specification and PD patterning. In all arthropods, *Dll* is first expressed in the specified limb primordia and later on in the distal parts of developing limbs (Panganiban *et al.*, 1995). Arthropods mutant for *Dll* develop truncated limbs that retain the proximal limb structures (considered an extension of the body wall called protopod) but miss the distal limb structures (considered the true limb called telopod) (Panganiban, 2000). Shortened appendages have been obtained after *Dll* knockdown by RNAi in *Parhyale* (Liubicich *et al.*, 2009), whereas more severe truncated appendages are observed after *Dll* CRISPR/Cas knock-out (Fig. 6) (Kao *et al.*, 2016). Besides its key role in limb specification and PD patterning, *Dll* has also another conserved, ancestral role in the development of the peripheral sensory organs in arthropods (Panganiban *et al.*, 1997).

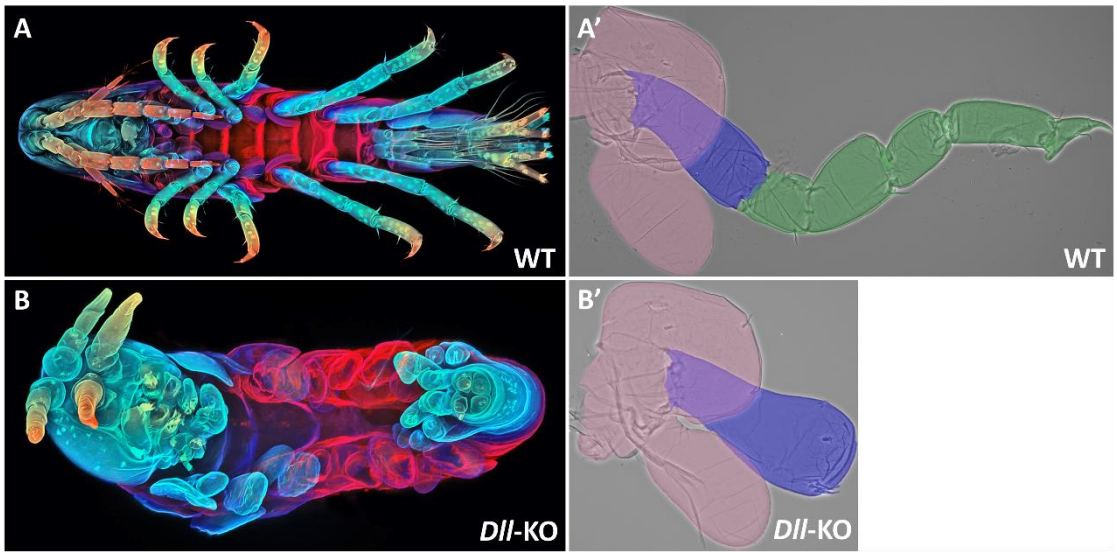

Figure 6: CRISPR/Cas9-based Dll knock-out in Parhyale. ***(A)*** *Ventral view of wild-type Parhyale and* ***(A')*** *full-sized thoracic limb.* ***(B)*** *Ventral view of mutant Parhyale and* ***(B')*** *truncated thoracic limb after CRISPR/Cas9-based Dll knock-out missing the distal segments (telopodite) shown in green in B'.*

Dll regulation in *Drosophila*

So far, the *cis*-regulatory sequences mediating *Dll* expression in developing limbs have been characterized only in *Drosophila* (Fig. 7A). The early *Dll304*

cis-regulatory element initiates *Dll* expression in the nascent limb primordia in the thoracic segments of *Drosophila* embryos (Vachon *et al.*, 1992). *Dll* is activated through signaling by the Wnt ligand Wingless (Wg) at the parasegmental boundaries determining the correct positioning of limb primordia

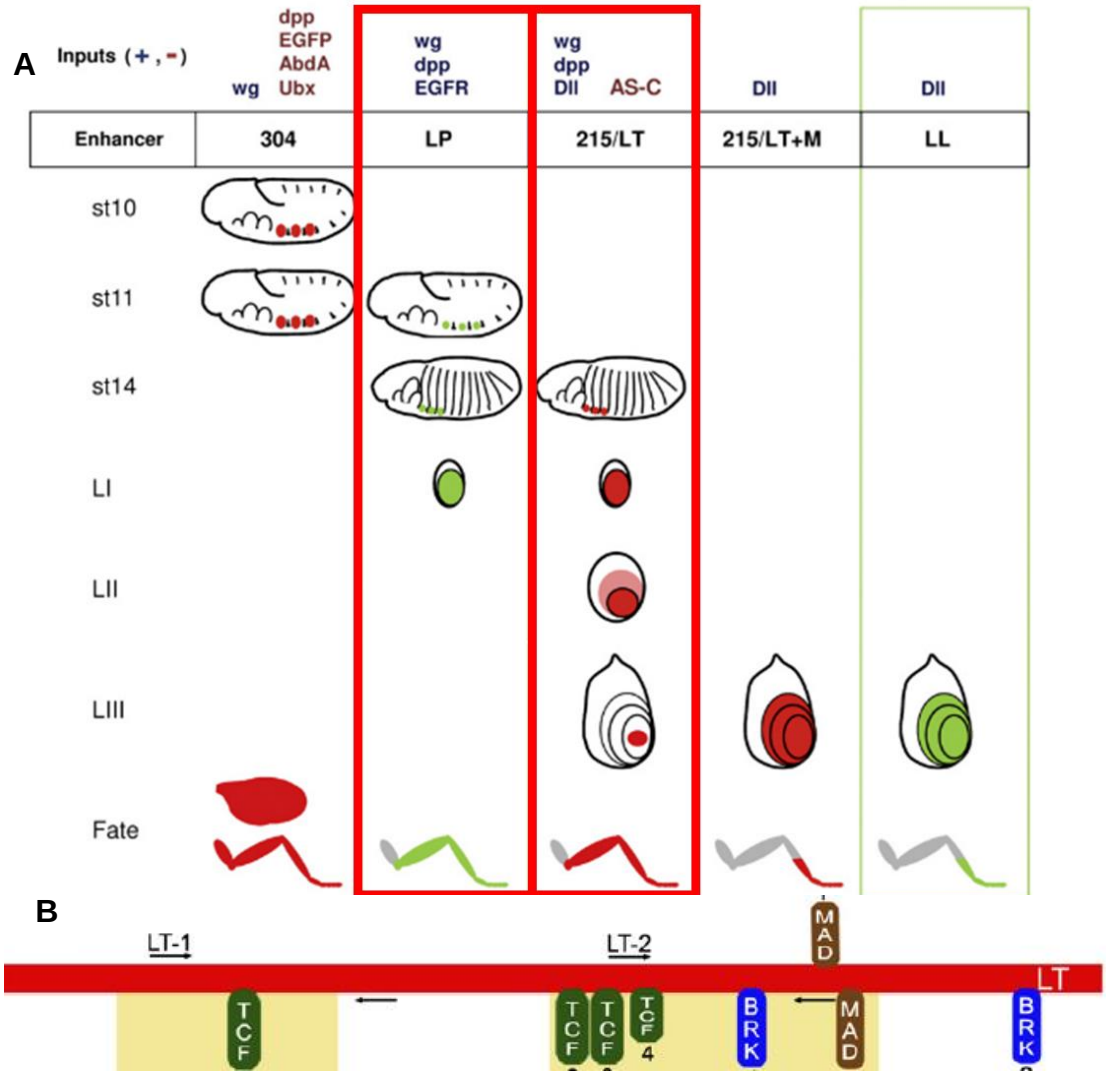

***Figure 7*: *Dll* regulation in *Drosophila* leg development**. *(A) The patterns of expression of Dll enhancers during leg development in embryonic and larval stages are depicted. The developmental stages are embryonic stages 10, 11 and 14; and LI, LII and LII larval stages. In the lowest row the adult fates derived from these domains of expression are shown. In the top row the known positive and negative regulatory inputs for each enhancer are shown. Image reproduced from Galindo et al., 2011 (B) Schematic representation of Dll LT enhancer with the transcription factor binding sites reproduced from Estella et al. 2008.*

along the anterior-posterior (AP) axis (Cohen, Simcox and Cohen, 1993). The correct positioning of the limb primordia along the dorsoventral (DV) axis is specified through repression of *Dll304* by the TGF-β ligand decapentaplegic (Dpp) dorsally and by the epidermal growth factor receptor (EGFR) signaling

pathway ventrally (Fig. 5A and 7A) (Goto and Hayashi, 1997; Kubota *et al.*, 2000).

The activity of the early *Dll304* decays after a few hours. *Dll* expression persists through the *Dll LT cis*-regulatory element in a ring of 15 cells cells that will give rise to the telopod (and rudimentary larval limbs called Keilin's organs) but not in the cells that will give rise to the protopod. This *Dll LT* CRE receives positive inputs from Wg, Dpp and EGFR pathways, as well as positive feedback from *Dll* itself (Fig. 5B and 7). The Wg and Dpp inputs into *Dll LT* are direct. Wg activates the mammalian TCF homolog, *pangolin (pan)*, a downstream transcriptional activator, whereas Dpp activates the transcription factor *Mothers against dpp* (*Mad*) and represses the transcriptional repressor *brinker (brk)*. *Dll LT* carries several binding sites for those transcription factors (Fig. 7B). (Estella and Mann, 2008; Estella, McKay and Mann, 2008; McKay, Estella and Mann, 2009)**.**

A more recent study reported additional *cis*-regulatory elements of *Dll* located 3' to the *Dll* transcription unit. The first element is called *Dll LP* and its activation is limited to a subdivision of *Dll* expressing cells in stage 10 embryos and persists until the first instar larval stage. *Dll LP* is also triggered by Wg, Dpp and EGFR pathways, but it is still unknown whether their input is direct or not. The second 3' *cis*-regulatory element is called *Dll LL.* It is activated only in the mid-third instar larva and is triggered by *Dll* activity (Fig. 7A) (Galindo *et al.*, 2011). Overall, the subdivision between proximal (Dll-negative) and distal (Dll-positive) fates, corresponding to a rudimentary PD axis, is established during mid-late *Drosophila* embryogenesis and carries on into early larval development.

In the early larva, the leg disc is divided into two distinct domains that do not overlap and antagonize each other: the distal domain defined by *Dll* expression that will give rise to the telopodite of the adult fly leg, and the proximal domain defined by the co-expression and nuclear localization of the Extradenticle (Exd) and Homothorax (Hth) homeodomain proteins that will give rise to the protopodite (Fig. 5E) (Rieckhof *et al.*, 1997)(González-Crespo and Morata, 1996; McKay, Estella and Mann, 2009).

Dll regulation in direct developing crustacean

Expression studies of *Dll* have been carried out in diverse arthropods, including crustaceans (Fig. 8A). The early clusters of *Dll*-expressing cells appear in an anterior-to-posterior progression as segmentation proceeds. *Dll* expression appears first in a few anterior cells adjacent to the posterior *engrailed* (*en*) stripes and then expands laterally and posteriorly (Fig. 8B) (Williams, Nulsen and Nagy, 2002; Hejnol and Scholtz, 2004; Prpic, 2008) (Browne *et al.*, 2005). Thus, crustacean appendage primordia are subdivided from early inception into posterior En-positive and anterior En-negative cells, and *Dll* expression appears to be regulated by positional cues emanating from compartment boundaries similar to *Drosophila* (Panganiban *et al.*, 1995; Wolff *et al.*, 2018; Pavlopoulos and Wolff, 2020). These expression studies, together with more recent lineaging experiments in crustaceans (Wolff *et al.*, 2018), have suggested that *Dll* is confined from early on to the cells that will contribute to the distal region of developing appendages. Thus, unlike in *Drosophila*, it appears that crustacean

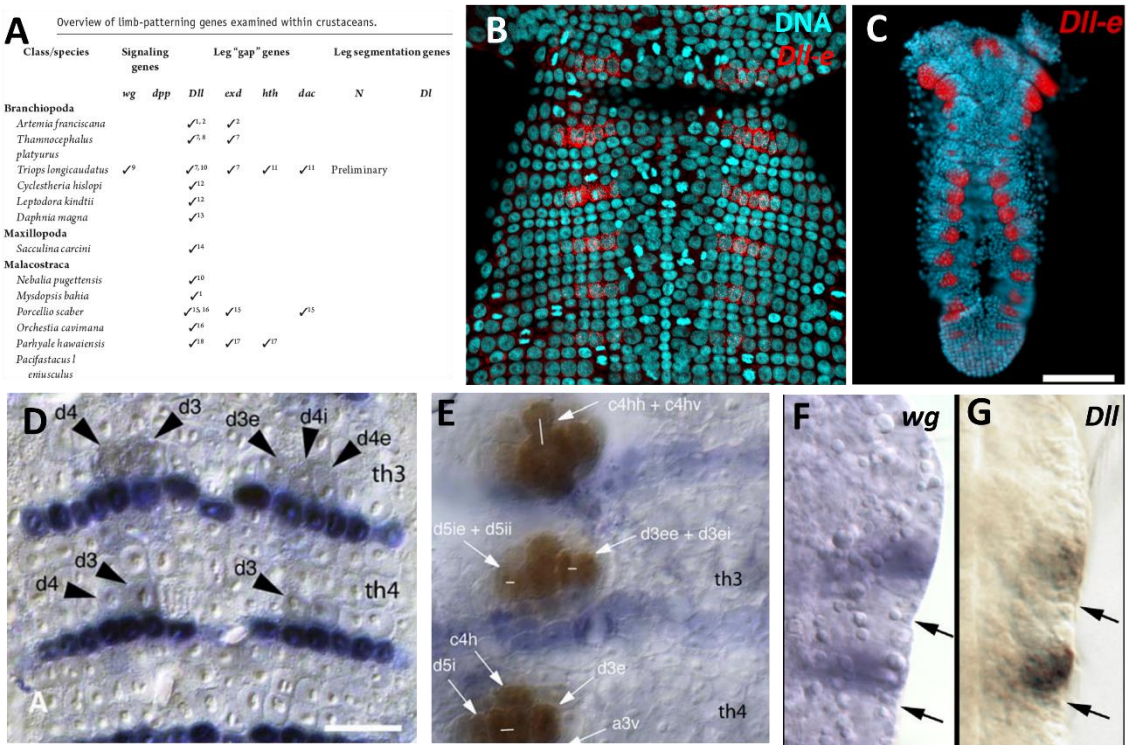

Figure 8: Dll expression in crustaceans. (A) *Overview of expression studies of limb patterning genes in crustaceans. Reproduced from Williams, 2013.* ***(B)*** *Parhyale Distal-less mRNA expression in early limb primordia and* ***(C)*** *in limb buds. Reproduced from Liubicich et al., 2009.* ***(D-E)*** *Dll (in brown) and en (in blue) protein expression at two consecutive stages of thoracic limb specification in the isopod Porcellio scaber. Reproduced from Hejnol and Scholtz, 2004.* ***(F)*** *wg and* ***(G)*** *Dll protein expression in early limb primordia in the branchiopod crustacean Artemia franciscana. Reproduced from Prpic, 2008.*

limb specification and early PD patterning happen at the same time. However, it is unclear how *Dll* expression initiates in arthropods other than *Drosophila*, which of the Wg, Dpp, or other pathways are required for limb specification and early PD subdivision, and which *Dll cis*-regulatory elements respond to these inputs. The timing and segmental expression of *wg* in some pancrustaceans are suggestive of its conserved role in *Dll* activation and early appendage development (Niwa *et al.*, 2000; Prpic *et al.*, 2003; Jockusch, Williams and Nagy, 2004; Prpic, 2004). This role has been functionally confirmed by RNAi knockdown of Wg pathway components in holometabolous but not hemimetabolous insects (Ober and Jockusch, 2006)(Angelini and Kaufman, 2005a). The generality of Wg signaling in appendage specification and early PD patterning is even more questionable outside insects (Angelini and Kaufman, 2005b). In crustaceans, the conserved segmental pattern of *wg* expression has been proven in two branchiopods (Fig 8F,G) (Williams, Nulsen and Nagy, 2002; Prpic, 2008; Constantinou *et al.*, 2016), but not in a malacostracan crustacean (Duman-Scheel, Pirkl and Patel, 2002).

Our knowledge about Dpp signaling is even less clear. It seems that *dpp* expression diverged between *Drosophila* and other pancrustaceans and arthropods. In *Drosophila, dpp* has a dynamic early expression pattern and opposite effects upon *Dll* regulation during appendage allocation and early patterning (Cohen, Simcox and Cohen, 1993; Goto and Hayashi, 1997). In other pancrustacean and arthropod species, *dpp* is expressed in segmentally transverse stripes and becomes increasingly restricted at the distal tip of the limb bud. Furthermore, *dpp* RNAi knockdown has no apparent effect on *Dll* expression and early appendage development in a more basal holometabolous insect than *Drosophila* (Ober and Jockusch, 2006). The expression pattern of *dpp* in crustaceans has been analyzed only in *Parhyale* (Wolff *et al.*, 2018). *dpp* expression is first detected in segmentally transverse domains corresponding to the dorsal-fated cells in the limb primordia. High levels of *dpp* expression persist in a row of anterior dorsal cells abutting the AP boundary of outgrowing *Parhyale* limbs, which reminds us of *dpp* expression in the *Drosophila* leg disc.

The diverging expression patterns and the lack of reliable functional data for pleiotropic genes like *wg* and *dpp* outside *Drosophila*, make it difficult to reach

safe conclusions about the pathways involved in appendage specification and early PD patterning in pancrustaceans (and arthropods). Views vary widely, ranging from (1) a low conservation in these early mechanisms (Angelini and Kaufman, 2005b), (2) the redundancy (or replacement) of the Wg and Dpp signaling activities in early appendage development with (or by) other Wnt and TGF-β family members in some lineages (Ober and Jockusch, 2006), to (3) the mechanistic conservation but diverging expression patterns to account for the topological differences between the flat *Drosophila* leg disc and the three-dimensional limb outgrowths in most other arthropods (Prpic *et al.*, 2003).

The available data suggest remarkable modifications and heterochronic shifts of these mechanisms during the evolutionary transformation of direct into indirect developing limbs. One plausible scenario is that the early embryonic phases of broad *Dll* activation and later restriction in *Drosophila* are evolutionary innovations not present in other pancrustacean lineages. In these lineages, embryos may use patterning mechanisms more comparable to those operating in the early *Drosophila* leg disc for coupling appendage specification with subdivision into proximal and distal cell fates. Additional evidence supporting this scenario comes from the inferred early compartmentalization of the *Parhyale* limb primordia along the DV axis, in addition to AP, which in *Drosophila* takes place only during the early second instar larval stage (Wolff *et al.*, 2018).

The aim of my Master's thesis reserach has been to compare the regulation of *Dll* during appendage specification and early PD patterning between the direct developing *Parhyale* limbs and the indirect developing *Drosophila* limbs.

Materials and Methods

Bioinformatic analysis

Quality control of ATAC-seq reads

The starting point for the ATAC-seq analysis pipeline was the control of the quality of reads using the FastQC tool that was downloaded for Linux from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. The overall read quality was satisfying with the exception of the “per base sequence content”. Analyzing the base content of reads recovered from a sequencing run, one would expect little to no differences between the A-T and G-C contents along the read length, as well as values resembling the global base content in the *Parhyale* genome. To remove the erroneous base content detected in the beginning of the reads (Fig. 9A,A’), we used Trimmomatic (Version 0.36) (Bolger, Lohse and Usadel, 2014) to trim the first 14 nt from all reads reducing read length to 61 nt (Fig. 9B).

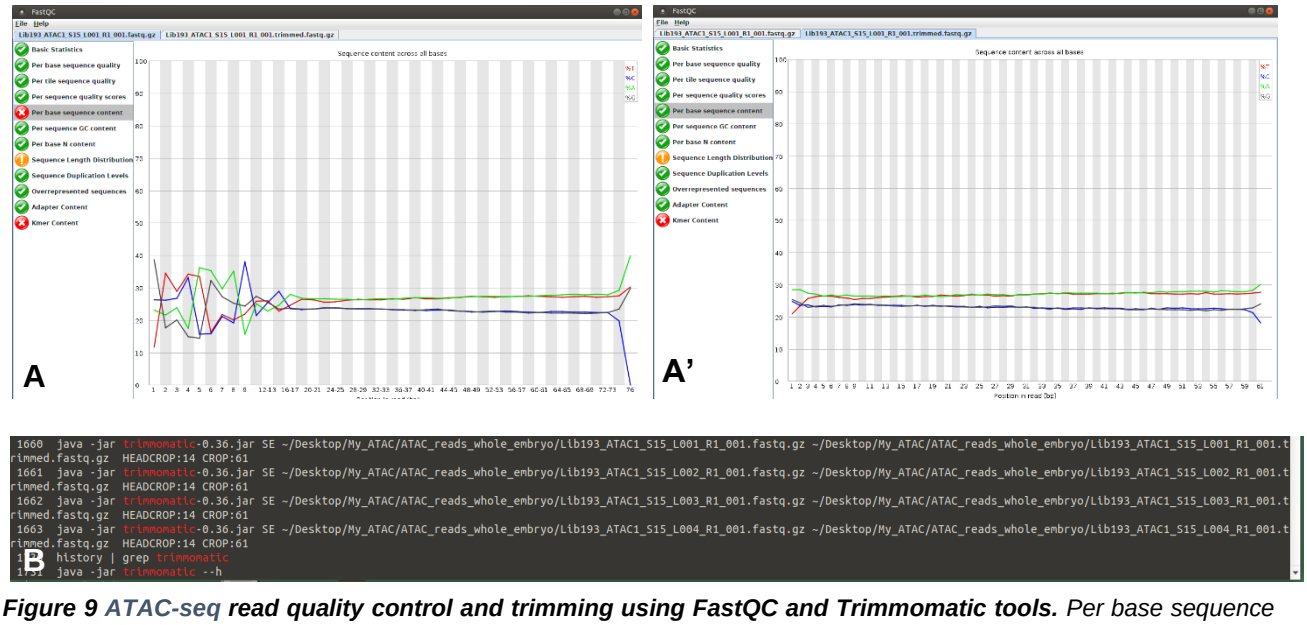

Figure 9 ATAC-seq read quality control and trimming using FastQC and Trimmomatic tools. *Per base sequence content* ***(A)*** *before read trimming and* ***(A’)*** *after read trimming.* ***(B)*** *Execution of Trimmomatic in commandline.*

Read alignment, genome coverage and Peak-calling

I used the Burrows-Wheeler Aligner (BWA-MEM) (Li and Durbin, 2009) to map the trimmed ATACseq reads onto the latest version of the *Parhyale* genome

Phaw_5.0 (https://www.ncbi.nlm.nih.gov/assembly/GCA_001587735.2/). The algorithm seeds alignments that have maximal exact matches (MEMs) and then extends the seeds with the affine-gap Smith-Waterman algorithm. The BWA-MEM algorithm generated the Sequence Alignment Map (SAM) files that were processed using the SAMtools suite (Version 1.9) (Li *et al.*, 2009). SAM files were converted into compressed Binary Alignment Map (BAM) representations using the SAMtools view command, BAM files were merged using the SAMtools merge command, and reads were sorted according to their coordinates using the SAMtools sort command. In order to get an estimate of the quality of the mapping, I used the SAMtools flagstat command to check the percentage of uniquely mapped reads (Fig. 10).

Figure 10: Mapping ATACseq reads onto the Parhyale genome. *SAMtools flagstat showing the percentage of uniquely mapped reads on the genome*

The histogram of read coverage against the genome was then computed with the BEDtools suite (Version 2.28) (Quinlan and Hall, 2010) and the bedtools genomecov command. The created bgd file was sorted by scaffold name and coordinates using the sortBed command and ATACseq peaks were called using the MACS2 algorithm (Zhang *et al.*, 2008) and the peakcall command. In order to run MACS2 smoothly, ideally a genome with chromosome contiguity is needed. Due to the highly fragmented genome of *Parhyale*, we could not run this algorithm with available computer-power. Therefore, the downstream analysis was performed with the bdg format coverage files. To speed up this analysis, we created a simplistic bash script with all the commands mentioned above and let it run overnight (Fig. 11).

*commands.sh

```
#Alignment
cd ~/Desktop/My_ATAC/ATAC_reads_whole_embryo
for i in 1 2 3 4
do
    bwa mem ~/Desktop/My_ATAC/genomeV5.fa Lib193_ATAC1_S15_L00$i"_R1_001.trimmed.fastq.gz" > WE_$i".sam"
    samtools view -h -b WE_$i".sam" > WE_$i".bam"
done

samtools merge WE_merged.bam WE_1.bam WE_2.bam WE_3.bam WE_4.bam

samtools sort WE_merged.bam > WE_merged_sort.bam

bedtools genomecov -ibam WE_merged_sort.bam -bga -g ~/Desktop/My_ATAC/genomeV5.fa > WE_merged_sort.bdg

sortBed -i WE_merged_sort.bdg > WE_merged_sorted.bdg
```

***Figure 11: ATACseq read coverage against the Parhyale genome**. Bash script created for read alignment and genome coverage acquisition*

Genomic DNA prep

High quality genomic DNA (gDNA) was isolated from 5 wild-type *Parhyale* adults and 10 y,w *Drosophila* adults using the DNAzol reagent (Thermo Fisher Scientific). Individuals were anesthetized with CO_2, transferred to nuclease-free tubes containing 500 µl of DNAzol and manually homogenized with pestles. Each lysate was centrifuged for 10 min at 13.000 rpm at room temperature. 500uL of supernatant that contained the gDNA was transferred to a new tube. For gDNA precipitation, 250 µl of ice cold 100% ethanol were added, mixed by shaking the tubes and kept at room temperature for 1-3 min. Samples were then centrifuged for 10 minutes at 13.000 rpm at 4ºC, the gDNA pellet was washed twice in 750 µl of 70% ethanol, carefully dried and resuspended in 50 µl of nuclease-free water. 1 µl of gDNA was analyzed by 1% agarose gel electrophoresis for quality control and the rest was stored at -20ºC.

PCR and Primers

Table 1 lists all primers used in this study that were diluted in nuclease-free water at a stock concentration of 100 µM. The Phusion High-Fidelity DNA Polymerase (New England Biolabs) was used for PCR amplification of all DNA fragments cloned in this study. For each PCR product, I performed a gradient PCR using three to five different annealing temperatures covering +/-5ºC around

the lower primer's Tm calculated with the NEB online tool (https://tmcalculator.neb.com/). Each 20 µl PCR reaction contained 4 uL of 5X Phusion HF Buffer, 0.4 uL of 10mM dNTPs, 1uL of 10uM Forward and Reverse primers, ~5ng of genomic DNA, 0.2uL of Phusion Polymerase (0.4units) and nuclease free water up to 20uL.The general PCR program used was: initial denaturation at 98°C for 30 seconds, 30 cycles of denaturation at 98°C for 15 seconds, annealing at 55-72°C for 20 seconds, extension at 72°C for 30 seconds per kb of estimated PCR product, a final extension step at 72°C for 5 minutes and hold at 4°C. 5 µl from each PCR reaction were analyzed by agarose gel electrophoresis and the reactions containing the amplicons of expected size were purified with the DNA Clean & Concentrator-5 kit (Zymo Research) and used in downstream cloning applications.

Primer name	primer bases (5' to 3')	Tm total (°C)	product (bp)
mCherry_F_(XhoI)	TTA CTCGAGATGGTGAGCAAGGGCGAG	68	731
mCherry_Stop_R_(NotI)	ATATGCGGCCGCTACTTGTACAGCTCGTCCAT	72	731
Dm_Dll_5UTR_F(NcoI)	TACCATGGCACACCACTGCTACAGAG	66	904
Dm_Dll_5UTR_R	GTGACTCCACTCCACTGATCT	58	904
Dmel_Dlle 5UTR_Nested F1	TGTTGTCCGCAACTGTTGAT	53	1895
Dmel_Dlle 5UTR_Nested R1	GATCGAAATAAGACTATTGGC	47	1895
Phaw_Dlle 5UTR F (PciI)	AATACATGTTTTGAATTAATTTTGCGA	56	753
Phaw_Dlle 5UTR R (NcoI)	TACCATGGATGAGTACAGAGAG	56	753
Phaw_Dlle 5UTR_Nested F1	TGCACGCTCGTATCGCTAC	55	1638
Phaw_Dlle 5UTR_Nested R1	TCTCAAGTGTGTAGTGCGAGG	54	1638
215/LT_F (PciI)	TTAACATGTAATTCGTTCTTGTTGGCT	60	939
215/LT_R (PciI)	TTAACATGTCTAAGGCTGCGTGAGATG	63	939
215/LT_Nested_F1	GCGAACTCCAACTTCAATT	56	1687
215/LT_Nested_R1	TTGTCCGTTCTCGCCTGTT	59	1687
LP_F (NcoI)	TACCATGGAACAATGTAATCATCAT	58	991
LP_R (PciI)	TTAACATGTAACTACAACTACAGAAGG	61	991
LP_Nested_F1	ATGATTTGTAGTGGGCTGT	54	1728
LP_Nested_R1	TGTACGAAGTAATGTGCTC	52	1728
F1_New_enh1_nested	TTAGCGATGTTAAAGGAG	50	2850
R1_New_enh1_nested	AGAACCTGTCAACCATTT	51	2850
F_New_enh1 (NcoI O/H)	TACCATGGCGGATCTTCGTTACATTA	64	2580
R_New_enh1(PciI O/H)	TTAACATGTATCTGTTTGTTCGTTCGT	61	2580
New_Enh2_F1_Nested	TTTTAGCTCTATGCCTGCC	57	3800
New_Enh2_R1_Nested	CTATCTAACCAGCCAGCCA	58	3800
New_Enh2_F_PciI	TTAACATGTATCTGTGGCCATCCGTAC	66	3022
New_Enh2_R_PciI	TTAACATGTCGAAGATCCGTTCAAACC	67	3022
F1_enh_3UTR_nested	GCTACTGGTGCAAATTAC	53	2655
R1_enh_3UTR_nested	TTCAGCTGCTACGAATAG	53	2655
F1_E3_new_BspHI(O/H)	TTATCATGAGGTAATACCCATTGGATC	63	3071
R1_E3_new_BspHI(O/H)	TTATCATGACCTGATGTTTGACAACCT	64	3071
F1_enh4_nested	AGCTGCTACTATCCCTGC	58	3480
R1_enh4_nested	GCTCAGAGTCTGTGTCAG	56	3480
F_enh.sp4 (F1_O/H)	TCGATACGCGTACGGTTAGTGCTAAGTAATTAG	67	3000
R_enh.sp4(pGTZ_linker O/H)	CGGCCTAGGCGCGCCATTATATGGCTGGAATTC	74	3000
F1_E5(3UTR)_new_PciI(O/H)	TTAACATGTACTTGTGCTGAAACACTG	62	2044
R1_E5(3UTR)_new_PciI(O/H)	TTAACATGTCGTTAACAGTGTGACGG	65	2044

Table 1: List of primers. *Sequence, Tm and amplicon size for all primers used in this study.*

Molecular constructs for the CRE expression analysis

Each desired DNA fragment was PCR amplified from genomic DNA using nested primer sets. Because of the size and the complexity of *Parhyale* genome, more than one product was often produced during the first amplification step. To increase the specificity, a small amount of the first PCR was used as template in a nested PCR reaction with internal primers that contained overhangs of the desired restriction enzyme sites. All PCR products were purified using the Zymo Research DNA Clean & Concentrator-5 kit before

cloning. mCherry was PCR amplified from p {NHEJ_KI_DS3_T2A-GGFF_UAS-mCherry} plasmid using primer pairs with XhoI and NotI overhangs and cloned in frame with PhH2B histone into the pGEMT Easy plasmid (Promega) pGTZ{PhH2B-Ruby2-SV40-attB} which was digested with XhoI and NotI replacing Ruby2 fluorescent protein. The resulting plasmid was named pGTZ {PhH2B-mCherry-SV40-attB}. The other components of this plasmid are an SV40 3'UTR that is known to increases the stability of the reporter mRNA leading to higher amounts of transcripts and an *attB* site that can be used for *AttP/AttB φC31* integrase mediated transformation. The PCR amplified *Parhyale* and *Drosophila Dll* promoter and 5'UTR fragments were flanking an NcoI restriction site at their 5' end and a PciI restriction site at their and 3' end were further cloned into the NcoI digested and dephosphorylated with Antarctic Phosphatase (NEB) pGTZ {PhH2B-mCherry-SV40-attB} vector upstream of H2B. PciI has compatible cohesive ends with NcoI resulting in the loss of the NcoI restriction site upstream of H2B, for strategic reasons. The outcoming vectors were named pGTZ {PhDlle_p_mCherry} and pGTZ {DmDlle_p_mCherry}. Each amplified by PCR putative CRE region was cloned into the NcoI digested pGTZ {PhDlle_p_mCherry} vector upstream of the *Parhyale Dlle* endogenous promoter region, giving rise to pGTZ {Ex_PhDlle_p_mCherry} plasmids (x =1,2,3,4,5). Similarly, the *Drosophila* LT and LP enhancers were PCR amplified and cloned into pGTZ {DmDlle_p_mCherry} which was digested by NcoI and dephosphorylated to create pGTZ {LT_DmDlle_p_mCherry} and pGTZ {LP_DmDlle_p_mCherry}. All constructs were verified by sequencings.

For the creation of the transgenesis vector, the reporter gene cassettes were extracted from pGTZ by AscI restriction digests and sub-cloned into pMi {PhOps1-EGFP}. The construct maps can be found in materials and methods (Fig. 12-15). For all restriction digests we used New England Biolabs enzymes. Dephosphorylation of linearized vector was conducted using NEB Antarctic Phosphatase before ligation. For ligating linearized DNA fragments, we used NEB T4 DNA Ligase and a vector to insert molar ratio ranging between 1:5 and 1:7. For every ligation we used 20-50 ng of linearized vector.

Bacterial transformation and colony selection & purification

DH5a chemo-competent bacterial cells were transformed with each molecular construct using half of the ligation reaction (100-300ng of DNA). Bacteria were incubated on ice for 30 minutes following a heat shock at 42ºC for 1 minute. After heat-shock bacteria were incubated on ice for 2-3 minutes, 900uL of LB medium were added and incubated at 37ºC for 40-55 minutes shaking at 700 rpm in an Eppendorf ThermoMixer® Comfort. Next 100uL of bacteria were plated a 90mm Petri-dish containing LB agar and the appropriate antibiotic. The remaining bacteria were centrifuged for 13 seconds at full speed the bacterial pellet was resuspended in 100uL of supernatant and plated on a second 90mm Petri-dish. The 90mm Petri-dishes were incubated at 37ºC for 16 hours. Next, six to twelve single bacterial colonies were selected for each construct, inoculated into sterile glass tubes that contained 2mL LB with antibiotic and incubated at 37ºC for 16 hours shaking at 200 rpm. Plasmid DNA was extracted from each colony using the standard alkaline lysis DNA extraction protocol and verified for the presence of the correct plasmid with several restriction digest reactions run on a 1% agarose gel. The DNA preparations that carried the desired plasmids were either used for sub-cloning or were kept at -20ºC. The small amounts of the liquid cultures of the desired colonies were subsequently streaked into LB-agar-antibiotic plates and incubated at 37ºC for 16h. Next, a single colony was inoculated into a sterile flask containing 50mL or 5mL of LB-antibiotic and incubated at 37ºC for 16h. Following 400uL of the grown culture were used for bacterial glycerol stabs in 30% glycerol, and the rest 49,6mL were used for Midi-preparations if the plasmid was destined for micro-injections or mini-preparations. For the Midi preparations we ZymoPURE II Plasmid Midiprep Kit. For the Mini-preparations we used the QIAprep Spin Miniprep Kit. All preparations were measured on nanodrop spectrophotometer for quality and quantity, and kept at -20ºC. To avoid multiple freezing and thawing of Midi-preparations used for injections we also kept them in aliquots of 2,5µg of DNA.

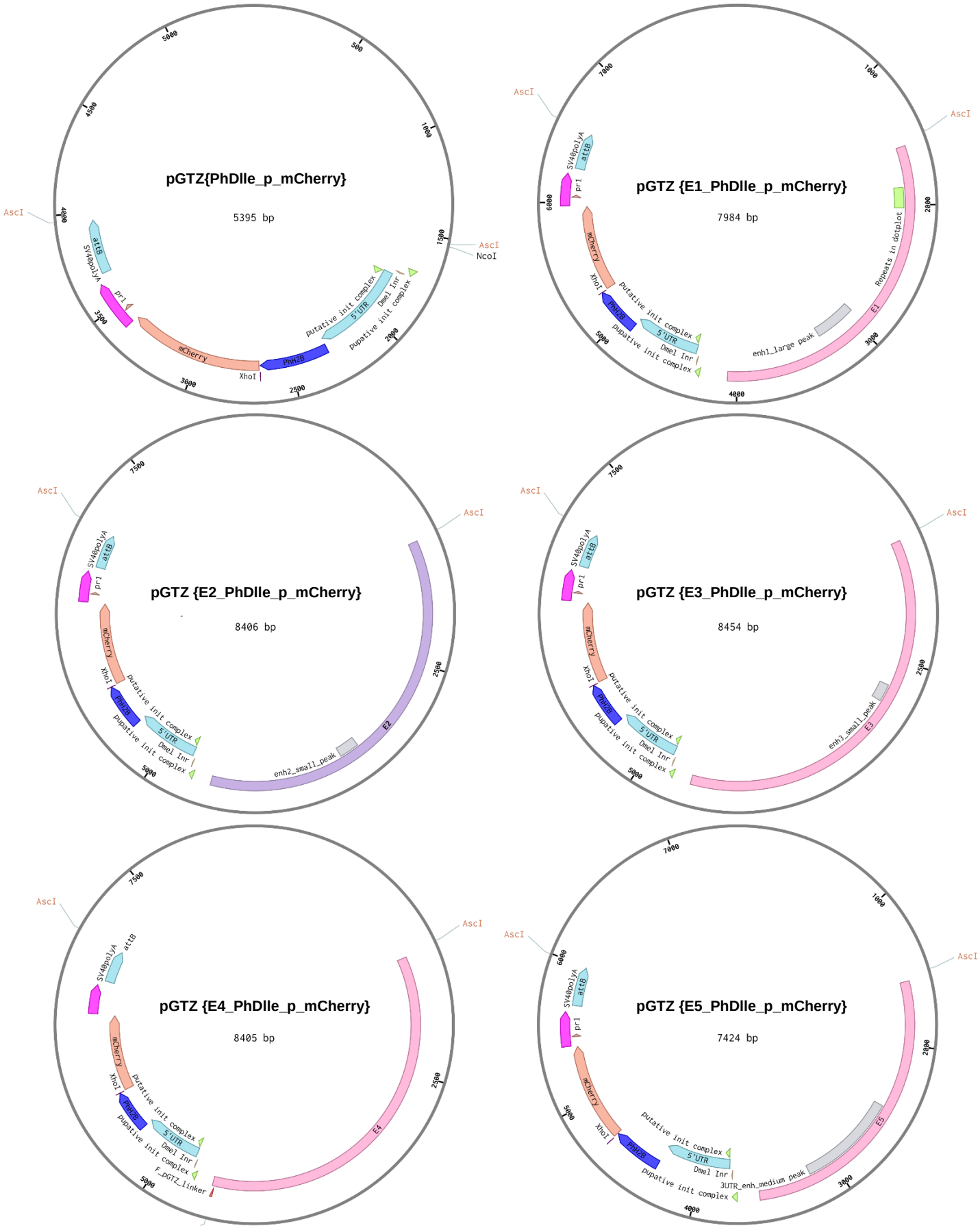

Figure 12: *Reporter constructs in p GEM T easy bearing the cloned ATAC-seq peaks upstream of Parhyale Dlle endogenous promoter and nuclear mCherry.*

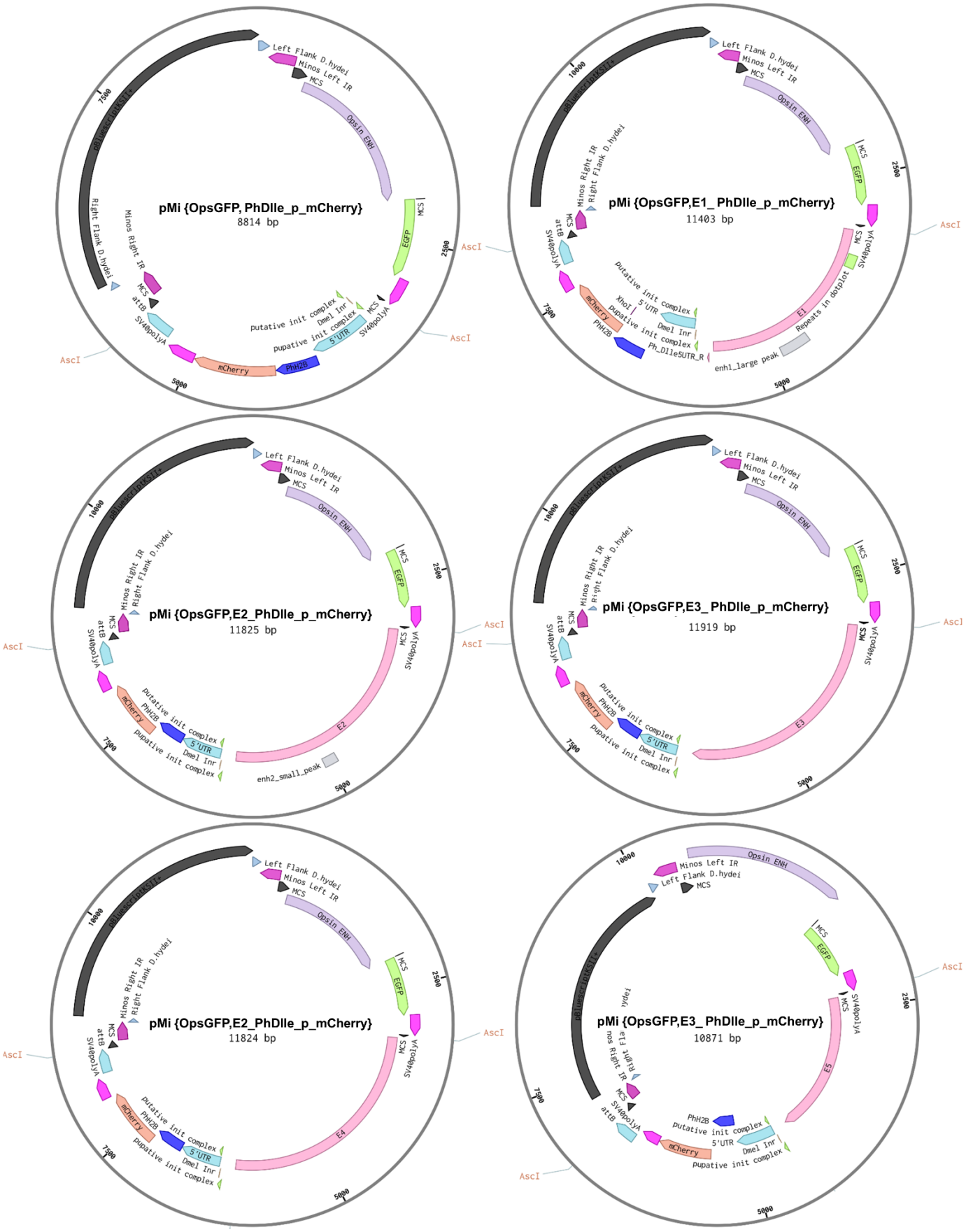

***Figure 13**: Reporter gene cassettes sub-cloned into pMi_OpsGFP vector that carries the Minos Left and Right inverted repeats*

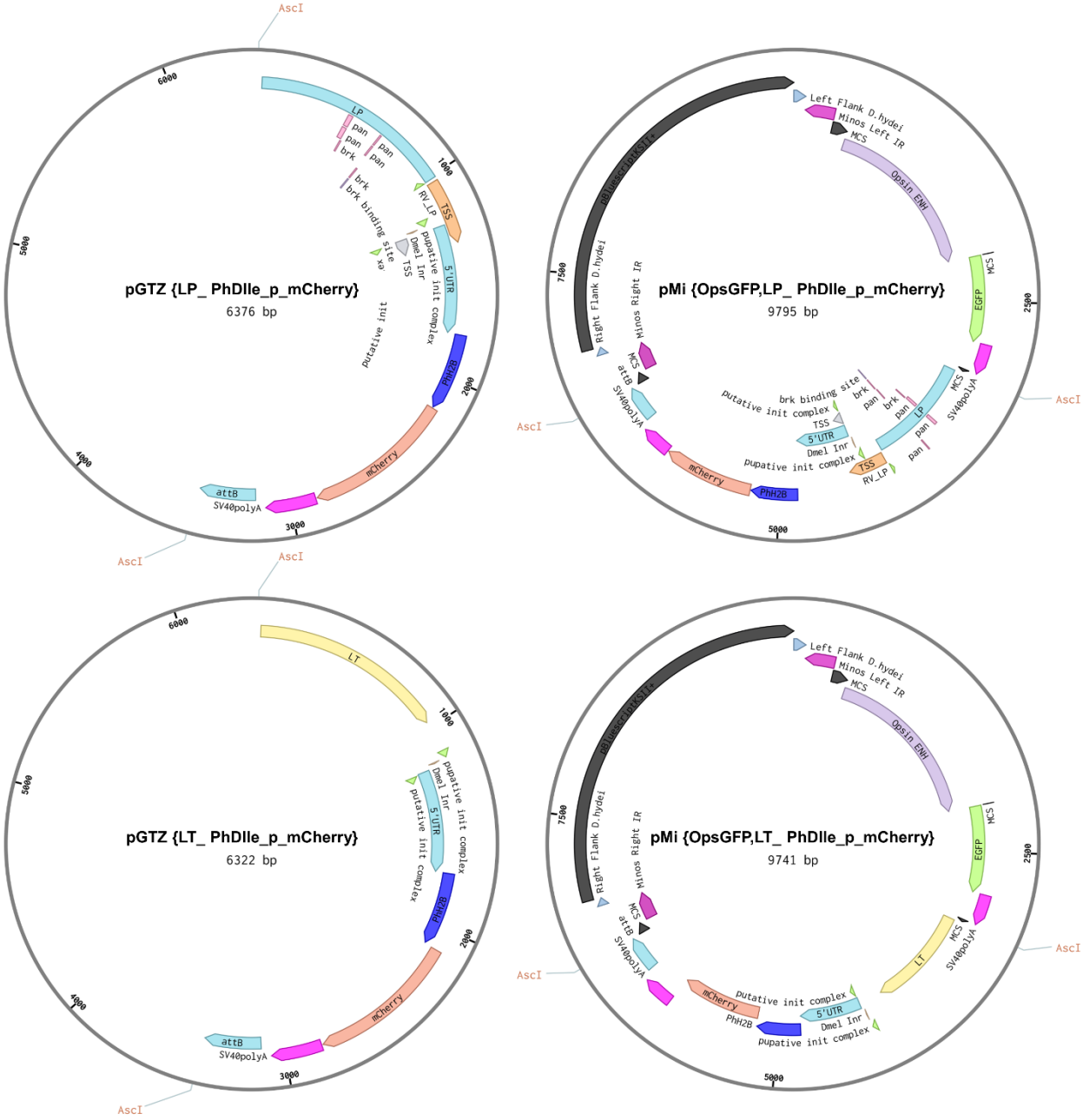

Figure 14: Maps of plasmid constructs used in this study. (Left hand side) *Reporter constructs in p GEM T Easy bearing the cloned Drosophila LP and LT enhancers upstream of Parhyale Dlle endogenous promoter and nuclear mCherry.* ***(Right hand side)*** *Reporter gene cassettes sub-cloned into pMi_OpsGFP vector that carries the Minos Left and Right inverted repeats.*

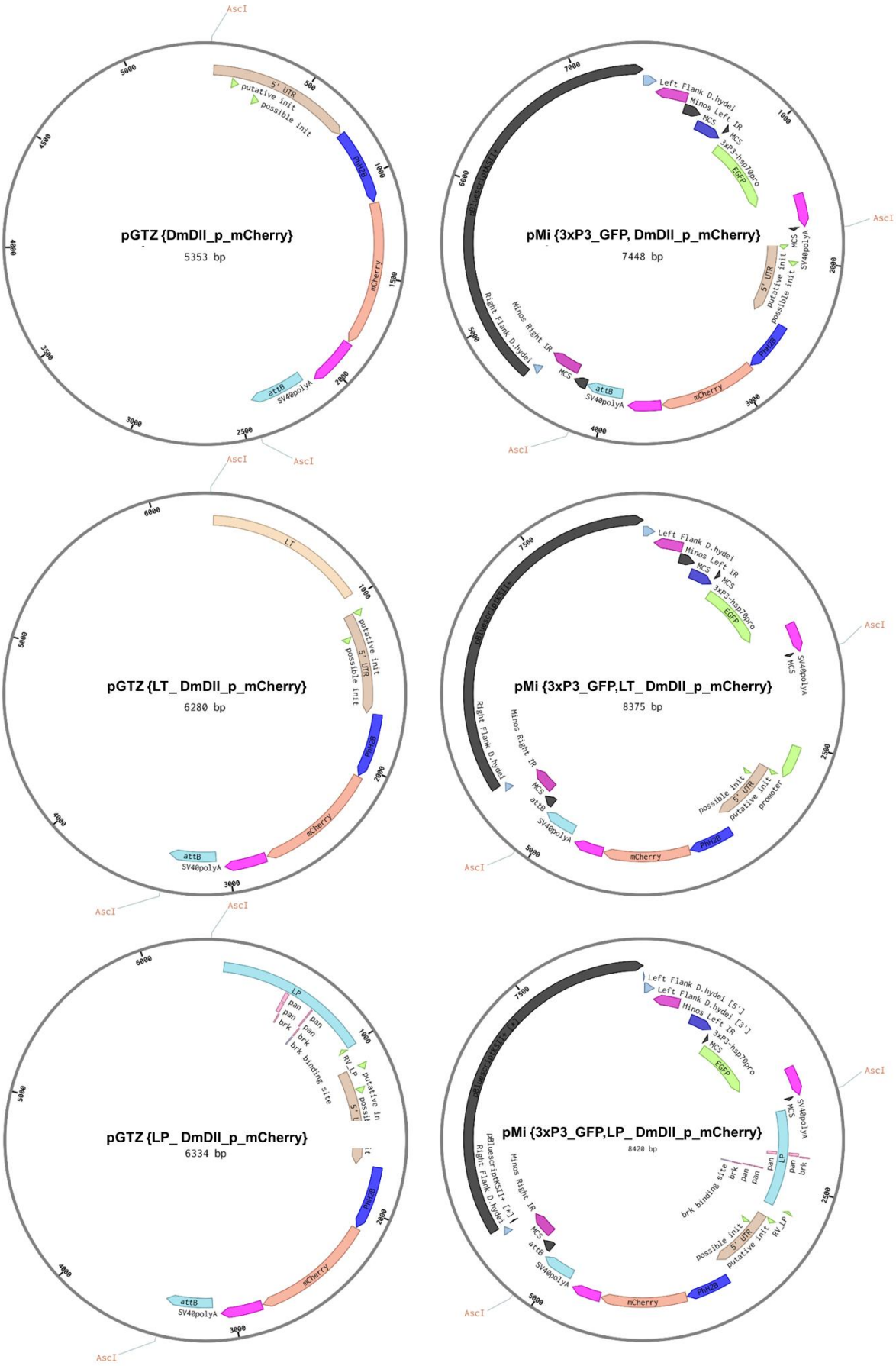

Figure 15: Maps of plasmid constructs used in this study. (Left hand side) *Reporter constructs in p GEM T easy bearing the cloned Drosophila LP and LT enhancers upstream of Drosophila Dll endogenous promoter and nuclear mCherry.* ***(Right hand side)*** *Reporter gene cassettes sub-cloned into pMi_OpsGFP vector that carries*

Parhyale hawaiensis aquacultures

Parhyale were maintained in dense cultures at 25–26°C. Thousands of animals were kept in 10–20 l plastic containers on a bottom layer of crushed corals, covered in 4-5 l of 34‰ artificial seawater (ASW) on a rich diet made of ground tropical and spirulina fish flakes, wheat germ pellets, vitamins and fatty acid cocktails. Cultures were aerated continuously with submerged water or air pumps and kept waste-free with phosphate and nitrate absorbing bags and weekly changes of the seawater. These cultures provided daily accessibility to hundreds of embryos at all stages of development all year round.

Parhyale 1-cell stage embryo collection for micro-injections

Adult *Parhyale* form mating pairs. For each micro-injection procedure, we had to collect more than 150 pairs from the wild-type cultures, the day before injections. The pairs were transferred into 150 mm Petri dishes with ASW and few corals. After mating, gravid females carry their eggs in a ventral brood pouch and can be easily identified by naked eye. Using a Pasteur pipette, gravid females were transferred into 90mm Petri-dishes with Filtered Artificial Sea Water (FASW) and were anesthetized by bubbling CO_2 gas for 30–60 s. Anesthetized females were then transferred into 60 mm tissue culture Petri dishes that contained FASW with antibiotics (penicillin-streptomycin diluted 1/100) and antimycotics (fungizone-amphotericin B diluted 1/200) (FASWA) (Kontarakis and Pavlopoulos, 2014). To remove the eggs from the ventral pouch, each female was held gently on its back under a dissecting scope with one pair of blunt forceps, while the brood pouch was nicked and the eggs were pushed gently out with another pair or forceps. This way, the fertilized eggs from all anaesthetized females were collected into a 60 mm tissue culture Petri dish in FASWA. Anaesthetized females were left to recover before returning them into the main *Parhyale* cultures. From each batch, only 1-cell stage embryos were sorted out and used for microinjections.

Preparation of the microinjection mix

To proceed with micro-injections for stable transposon mediated transgenesis, we synthesized *in-vitro* capped mRNA encoding the *Minos* transposase. We used the pBlueSKMimRNA plasmid template that contains a T7 promoter driving the expression of the *Minos* transposase coding sequence flanked with 5′ and 3′ UTRs from the *Drosophila hsp70* and *inflated* (αPS2 integrin) genes, respectively. The pBlueSKMimRNA plasmid was linearized with NotI and used as template for capped *Minos* mRNA synthesis using the HiScribe™ T7 ARCA mRNA Kit (New England Biolabs) as previously described (Kontarakis and Pavlopoulos, 2014). Each in vitro transcription reaction yielded approximately 25ug of capped mRNA encoding the *Minos* transposase. Each injection mix contained 500 ng/uL of the *Minos* vector of interest, 300 ng/uL of the *Minos* transposase mRNA, and 0.1 volumes phenol red in a total volume of 5uL.

Parhyale hawaiensis microinjections

Commercially available microneedles (Eppendorf Femtotips II) or customized microneedles prepared from borosilicate glass capillaries with OD/ID 1.0/0.58mm on a P-97 Flaming-Brown needle puller (Sutter Instruments) were used for microinjection of *Parhyale* embryos. The capillaries were pulled using a box filament and the following settings: heat 503, pull 70, velocity 80, time 250. Right before microinjections, the tip of each needle was broken off by gently touching it against a glass slide. During microinjection, embryos were immobilized and kept moist by placing then on agarose steps prepared in (FASWA) (Kontarakis and Pavlopoulos, 2014).

Our microinjection set-up was composed of an automated microinjector (Eppendorf FemtoJet or Narishige IM300), a micromanipulator (Narishige M-152), and a Leica MZ9.5 stereo microscope equipped with 25X eye pieces. The microinjector allowed us to deliver small volumes of about 100 pl injection mix with precision through the microneedle by applying a certain amount of pressure for a small period of time. Each microneedle was filled with 1-2uL of injection mix and was mounted to a needle holder whose movement was controlled by the micromanipulator.

We loaded the agarose step with 10-15 1-cell stage embryos in 10uL of FASWA on a glass slide, and placed it in a way that its long edge faced the needle coming from the side of the microscope. The embryos were loaded onto the agarose step with a P20 pipette tip coated with BSA. Adjusting the micromanipulator, we arranged the needle to pierce the center of the embryo and triggered the injection pressing the foot pedal. We then removed the needle from the first embryo and proceeded with the rest. Once finished with all embryos on the agar step, we collected them with a fine brush into a 35mm tissue culture Petri dish, and repeated the process with the rest of the embryos as previously described (Kontarakis and Pavlopoulos, 2014).

Transient and stable transgenesis in *Parhyale*

At least 100 1-cell-stage *Parhyale* embryos were collected and microinjected as described above. The injected embryos were transferred into 35 mm tissue culture Petri dishes with FASWA and incubated at 25–26 °C. During days 1-3 of embryogenesis, we were discarding the dead embryos and transferring the living ones (using a micropipette with a BSA-coated plastic tip) into a new 35 mm tissue culture dish in FASWA on a daily basis. During days 4-10, we repeated the previous step once every 2 days. FASWA stock was freshly made twice a week in a sterile bottle. Ten days after injection, late-stage *Parhyale* embryos were screened under a fluorescence stereoscope for expression of the transgenesis marker gene *Opsin1-EGFP* in the eyes. Unilaterally- or bilaterally-EGFP-expressing embryos were selected and kept in separate 35 mm dishes in FASWA. Under ideal microinjection conditions, almost 30% of injected *Parhyale* embryos hatched 10–12 days after injection. About 20-30% of these G0 embryos exhibited bilateral or unilateral *Opsin1-EGFP* expression. Each Opsin1-EGFP positive hatchling was raised indivisually to adulyhood in a 60mm Petri dish with FASW, a few pieces of coral and ground fish flakes. Water and food were replaced regularly every 4 days by transferring each G0 into a new petri dish until it reached sexual maturation about 2 months after injection. Upon sexual maturation, single backcrosses of transgenic G0s with similar-sized wild-type *Parhyale* adults of the opposite sex were set in 60 mm Petri dishes in FASW with frequent water and food changes as described above. After

copulation, gravid females were transferred into separate 60mm Petri dishes and developing embryos were dissected out at the appropriate stage and kept in 35 mm tissue culture Petri dishes. At least 50 G1 embryos were screened from each G0 and stable transgenic individuals were selected based on bilateral *Opsin1-EGFP* expression. To create stable lines, transgenic G1 siblings from the same G0 founder were raised to adulthood and were crossed to recover homozygous G2s for each insertion.

Drosophila transgenesis

I used the *Drosophila* line VK00013 for microinjections tha carries a maternally expressed *ΦC31* integrase gene under the control of the *nanos* promoter on the X chromosome and is homozygous for *attP* landing sites on the third chromosome. Preblastoderm embryos were injected with the pMi{3xP3-EGFP; DmDll-LT-mCherry} and pMi{3xP3-EGFP; DmDll-LP-mCherry} constructs, which carried the *3xP3-EGFP* transformation marker, *attB* recombination sites and the mCherry fluorescent reporter under the control of the *Drosophila* LT and LP enhancers upstream of the *Drosophila Dll* promoter.

The genotype of the injected flies was:

$$nos{:}\Phi C31, y, w\ ;;\ attP$$

Survivors (about 10% of the injected embryos) were kept until adulthood in vials with *Drosophila* food.

♂ $\frac{nos{:}\Phi C31,\ y,\ w}{Y}$; ; $\frac{attP}{Mi\{3xP3\_eGFP,Dm5UTR\_LT\ or\ LP\_mCh\}}$ x ♀ y, w (cross 1)

Each F1 adult was screened for EGFP expression in the adult *Drosophila* eye. The transformation efficiency of this procedure was around 25-30%. Single transformed males were then crossed with $y,\ w$ virgin females, to remove vasa:Φc31 integrase from the genome.

Single F1 males that expressed 3xP3-EGFP were crossed with virgin females that carried chromosome three balancers TM3 and TM6B.

♂ $\frac{y,w}{Y}$; ; $\frac{+}{Mi\{3xP3\_eGFP,Dm5UTR\_LT\ or\ LP\_mCh\}}$ x ♀ ; ; $\frac{TM3}{TM6B}$ (cross 2)

Following, males and virgin females from the progeny of cross 2, that expressed 3xP3-EGFP and had the same genotype were crossed to establish stocks.

$$♂\frac{y,w}{Y};;\frac{TM3}{Mi\{3xP3\_eGFP,Dm5UTR\_LT\ or\ LP\_mCh\}} x ♀\ y,w;;\frac{TM3}{Mi\{3xP3\_eGFP,Dm5UTR\_LT\ or\ LP\_mCh\}}$$

(cross 3) or

$$♂\frac{y,w}{Y};;\frac{TM6B}{Mi\{3xP3\_eGFP,Dm5UTR\_LT\ or\ LP\_mCh\}} x ♀\ y,w;;\frac{TM6B}{Mi\{3xP3\_eGFP,Dm5UTR\_LT\ or\ LP\_mCh\}}$$

(cross 4)

Once the insertion was balanced, I established three separate lines for each transgene: Lines 5.1, 6.1, 6.2 carrying *Mi{3xP3-EGFP; DmDll-LT-mCherry}* and Lines 10.1, 14.2, 21.1 carrying *Mi{3xP3-EGFP; DmDll-LP-mCherry}* inserions.

Drosophila immunohistochemistry

For the collection of *Drosophila* embryos, I prepared cherry juice plates containing 3% agar dissolved in 50% cherry juice and water. Then we set up cages with 50-80 adults for each line that we created, using the agar juice plates coated with a layer of yeast. The embryo stages we wanted to collect were S11 to S15. To do so, I had to let flies lay eggs for 6 hours and let the eggs develop for 7 hours at 25^{o}C. Then I proceeded with the egg collection. Eggs were transferred into small baskets using a fine brush. For the dechorionation, I incubated the baskets with the embryos in 50% bleach for 2’ and washed them thoroughly with tube water and distilled water to remove bleach and prevent desiccation. I then transferred the embryos into glass 4mL vials that contained the fixation mix (2ml heptane,1.2ml PBS x1, 0.2ml 40% PFA). I fixed them for 18’ applying shaking at 200rpm. After that, I removed the lower phase with a glass Pasteur pipette. I added 2ml methanol and shook vigorously for 30’’-60’’ to allow embryo devitalinization. All embryos that sunk to the bottom of the vial were collected into a 2ml Eppendorf tubes. Next, I washed the embryos with methanol three times and stored at -20^{o}C.

For embryo rehydration, I removed half of the methanol and rinsed them with 1mL of PT (1xPBS + 0.1% Triton-X-100). Next, I removed half of the volume of the tube and add 1mL of PT. I repeated this process 3-4 times. Next, I washed the embryos twice with PT for 15’. I removed all of the PT carefully and blocked

the embryos with PBT for 30' (PT + 0,1%BSA) followed by 30' with PBT-5%NGS at 4ºC. Next, I incubated with primary antibodies (1:200 M-Wg and 1:400 R-mCherry in PBT-5%NGS) overnight at 4ºC. The primary antibodies were stored for reuse and embryos were rinsed three times with PT, washed three times with PT for 15' and blocked with the same procedure as before. The secondary antibodies that I used for our experiments were anti-M-AlexaFluor488 and anti-R-AlexaFluor555 diluted 1:500 in PBT-5% NGS. Embryos were incubated with secondary antibodies for 2-3 hours. After that, the secondary antibodies were removed and embryos were rinsed three times with PT and washed three times with PT for 15' (DAPI was diluted 1:1000 in PT and used in the second wash). After the washes, I removed the remaining PT and added 50% glycerol in PBS. I let the embryos sink, removed the glycerol and add 70% glycerol in PBS. Again, I let them sink overnight at 4ºC and mount in glass slide using 20-30uL of ProLong Antifade Mountant solution before adding a coverslip. Embryos were observed under a Leica SP8 confocal microscope and the acquired images were processed on Fiji.

Results

Search for *Parhyale* early-*Dll* cis-regulatory elements

ATAC-seq analysis

There are three *Distal-less* gene homologs in *Parhyale,* I will focus on the one that is expressed early in development and is called early-*Dll (Dlle)* (Liubicich *et al.*, 2009). To find putative cis-regulatory element (CREs) of the *Dlle* I used the ATAC-seq technique. My working hypothesis was that the identified open chromatin regions in the neighborhood of *Dlle* could act as its putative CREs.

Whole late-stage embryo (S22) DNA libraries were created and sequenced by Averof's Lab, who kindly shared the reads with us. The starting point of this ATAC-seq pipeline was the control of the quality of reads using the FastQC tool. To fix the "per base sequence content" failure I trimmed the reads using Trimmomatic-0.36. The trimmed reads were further mapped using the Burrows-Wheeler Aligner (BWA-MEM) against the index of *Parhyale* genome. Using

samtools I converted the SAM format of the alignments into its binary version BAM. Last but not least I merged and sorted them to one BAM file and used the flagstat command to see the percentage of uniquely mapped reads. Following the genome coverage file was created, using the bedtools genomecov command. The provided bdg formatted file was used for our further analysis. I then used macs2 algorithm to call the ATAC-seq peaks. For more details go to materials and methods.

To find the *Dlle* genomic region I acquired local BLAST. I created a database from the whole genome of *Parhyale* and as a query sequence I used the *Dlle* CDS against the whole genome database. After finding the exact genomic scaffold and the coordinates of *Dlle* I used the Integrative genome viewer (IGV) to visualize the ATAC-seq data by loading the *Parhyale* genome and the gene model file before loading the bdg format file containing the ATAC-seq peaks. I concluded by identifying four distinct peaks approximately 5kb 10kb 12kb 15kb of *Dlle* transcription start site, and one peak approximately 15kb downstream of the 3UTR, that could act as *Dlle* CREs. I also identified one high peak overlapping with the transcription start site (TSS) of *Dlle* (Fig. 13C).

MEME-Suite analysis

MEME Suite is a software toolkit for performing motif-based sequence analysis, which is valuable in a wide variety of scientific contexts. The MEME Suite software has played an important role in the study of biological processes involving DNA, RNA and proteins in over 9800 published studies. The MEME Suite is freely available for academic use at http://meme-suite.org, and source code is also available for download and local installation (Bailey *et al.*, 2009).

The question I wanted to answer was whether enrichment of known binding motifs could be found, in the ATAC-seq peaks, of Transcription Factors that regulate *Distal-less* expression in *Drosophila* such as pangolin (*pan*), brinker (*brk*), and Mothers against *dpp* (Mad). pan is a downstream transcriptional activator of the Wg pathway, *brk* is a transcriptional repressor and Mad is a transcriptional activator of the Dpp pathway. To do so, I used the MEME-Suite tool AME. AME identifies known motifs provided by the user which are relatively enriched in our sequences compared with control sequences (McLeay

and Bailey, 2010). The disadvantages of this approach are that, first, I had to provide candidate known motifs, adding a high-level of bias to our analysis to get our results, and second, the TF binding motifs that are known in arthropods come from *Drosophila*, which derived from *Parhyale* approximately 500 million years ago, which means that the sequences of DNA binding motifs of *Dll* regulators are probably changed.

To overcome these problems, I tried to find *de-novo* putative TF binding motifs, which would be over-represented in the ATAC-seq peaks and compare them with known motifs. For this purpose, I used three MEME-suite tools (MEME, TOMTOM and FIMO). MEME can discover novel, un-gapped motifs in our sequences. MEME splits variable-length patterns into two or more separate motifs. TOMTOM compares motifs against a database of known motifs (e.g., JASPAR). Subsequently, it ranks the motifs in the database and produces an alignment for each significant match. Last, but not least, FIMO scans a set of sequences to find individual matches to each provided motif. The combination of these three tools found two, nineteen nucleotides long, de-novo motifs (MEME) (Bailey *et al.*, 2009). The first one showed significant similarity to *Drosophila brk*, and Mad binding motif as well as similarity to *pan,* amongst other

TF name	p-value
opa	0.004
odd	0.0129
onecut	0.0152
cad	0.0165
brk	0.0281
sna	0.0325
Mad	0.0488
OdsH	0.0613
Trl	0.0626
en	0.0659
unc-4	0.0659
achi	0.0662
pan	0.0679
Rx	0.0708

Table 2: *List of existing arhthropod TF binding motifs similar with the de-novo identified motif showed in Fig. 16A after TOMTOM analysis.*

TF name	p-value
bap	0.000143
Dll	0.00571
btd	0.0072
NK7.1	0.0082
B-H2	0.00873
Hmx	0.0104
tup	0.0104
Lim1	0.0125
al	0.0137
unc-4	0.0137
Vsx2	0.0142
repo	0.015
kni	0.0154
Dr	0.0177

Table 3*: List of existing arhthropod TF binding motifs similar with the de-novo identified motif showed in Fig. 16B after TOMTOM analysis.*

transcription factors that have not been associated with *Dll* regulation in *Drosophila* (Fig. 16A), (Table 2). It has been shown that *pan, brk* and Mad transcriptional regulators directly regulate *Dll* in *Drosophila*. (Estella, McKay and Mann, 2008).The second motif showed similarity between others, to the leg gap gene Lim1 and *Dll* binding sites in a statistically significant manner (TOMTOM) (Gupta *et al.*, 2007) (Fig16. B), (Table 3). Both were over-represented in the ATAC-seq peak regions (FIMO) (Grant, Bailey and Noble, 2011) (Fig. 13C).

Vista plot analysis

On a complementary approach using local blast I searched the genomic region of *Dlle* homolog of the two available amphipod genomes: *Trinorchestia longiramus* and *Hyalella azteca.* Both amphipods belong to the superfamily of Talitroidea, as *Parhyale hawaiensis* does. Then we used the mVista online tool to compare those genomic regions. Both genomes were download from i5k genomic database, whose aim is to reach 5000 sequenced arthropod genomes (Robinson *et al.*, 2011). mVISTA is a set of tools that is used for comparing DNA sequences from two or more species up to several Mb in length. It also allows visualization of these alignments with annotation information (Brudno *et al.*, 2003; Frazer *et al.*, 2004). mVISTA has a clear output, that allows us to sequence similarities and differences. It is implemented as an online server that provides access to global pairwise, multiple and glocal (global with rearrangements) alignment tools. As expected, the Vista plots showed us peaks of high similarity in the exon regions and lower similarity in the intron regions. Surprisingly, Vista uncovered two low, but distinguishable peaks with over 50% similarity overlapping with two of the identified ATAC-seq peaks (Fig. 16E). Those conserved between the three species non-coding regions might have a functional role. Thus, this finding strengthens our hypothesis that the ATAC-seq peaks have a functional role as CREs.

Our three-step bioinformatic analysis allowed us to discover: (1) several open chromatin regions in the neighborhood of the *Dlle* genomic region in late-stage embryos, indicating their putative role as CREs, (2) two novel ungapped DNA-motifs one of which shared similarity with the DNA binding motifs of *pan, brk,* Mad, and *Dll* transcriptional regulators, associated with *Dll* regulation in *Drosophila* (3) two conserved regions in Talitroidea that overlapped with different ATAC-seq peaks

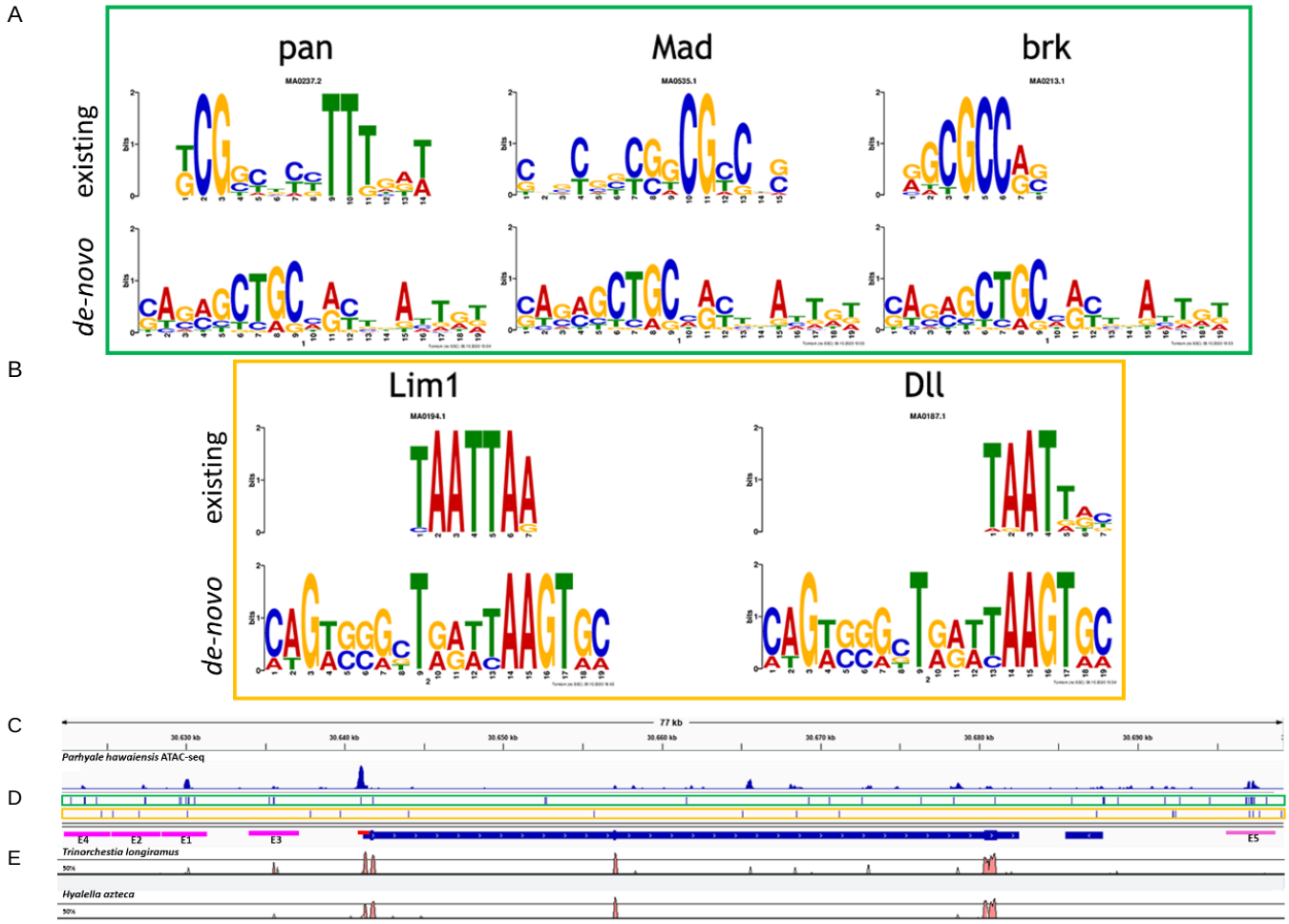

Figure 16: ***Representation of the bioinformatic analysis results from Parhyale Dlle locus***. ***(A, B)*** *de-novo identified transcription factor binding motifs intersected with similar existing insect motifs extracted from JASPAR database.* ***(C)*** *ATAC-seq peaks identified in the Parhyale Dlle locus visualized on IGV.* ***(D)*** *Enrichment for the de-novo identified motifs in blue vertical lines in the Dlle locus, green represents motif* ***A*** *and yellow represents motif* ***B****, blue horizontal line indicates the Dlle gene, while blue thick boxes represent the exons, red line represents the Dlle promoter and 5'UTR and the magenta lines are the regions used for creation of reporter constructs.* ***(E)*** *Vista plots comparing the Parhyale Dlle locus with the two available amphipod genomes. High similarities are observed in the exon regions as well as in the region of E1 and E3.*

Expression analysis of *Parhyale Dlle* putative CREs

To test the biological functionality of each putative CRE of *Dlle* I selected regions of length 2-3kb that included each ATAC-seq peak and created reporter constructs (Fig. 13D). As a promoter I used *Parhyale Dlle* endogenous promoter, a 200bp long sequence upstream of 5'UTR (which included the TSS ATAC-seq peak), as well as the *Dlle* 5'UTR region (Fig. 13D). As a reporter gene, I used the monomeric, bright and fast-maturing mCherry fluorescent protein fused to histone H2B for nuclear localization. The construct maps can be found in materials and methods (Fig. 12).

The six reporter gene cassettes so far created - five containing putative CREs upstream the *Dlle* promoter and one control only containing the *Dlle* promoter - were sub-cloned into pMi {PhOps1-EGFP} transgenesis vector which carries the Minos Left and Right Inverted Repeats. Minos is able to carry relatively large insert sizes. PhOpsin1-EGFP was used as a marker gene. Opsin-1 enhancer strongly drives the expression of GFP in the photoreceptors of the transformed animals, replacing the traditionally used Pax6-responsive promoter (*3xP3*) (Pavlopoulos and Averof, 2005). These constructs were micro-injected together with mRNA encoding the Minos transposase into *Parhyale* embryos for Minos transposon mediated transgenesis as described in (Pavlopoulos and Averof, 2005; Kontarakis and Pavlopoulos, 2014). See materials and method (Fig. 13)**.**

The surviving injected Go embryos were screened live under fluorescent stereoscope for the putative CRE-H2B-mCherry expression throughout embryogenesis, and for Opsin1-GFP positive transformants on the tenth day when the compound eye of the embryo is fully formed. The injection results can be found in the following table (Table 4).

***Parhyale hawaiensis* injection table**				
	Surviving/Injected	**Transformed G0 (ops+)**	**Surviving G0**	**Surviving (Ops+) G1**
pMi {OpsEGFP,E1_PhDlle_p_mCherry}	**86/324**	**17**	**13**	**3/46**
pMi {OpsEGFP,E2_PhDlle_p_mCherry}	**46/230**	**5**	**2**	
pMi {OpsEGFP,E3_PhDlle_p_mCherry}	**16/50**	**3**	**3**	
pMi {OpsEGFP,E4_PhDlle_p_mCherry}	**20/70**	**5**	**3**	**1/10**
pMi {OpsEGFP,E5_PhDlle_p_mCherry}	**36/152**	**5**	**5**	

Table 4: Parhyale hawaiensis injection table*. Successful steps are indicated with green boxes, yellow boxes indicate that this step is still in progress, whereas red boxes indicate failure of transgenesis procedure.*

The majority of the transformed (Opsin-GFP positive) individuals from CRE lines all did not exhibit any detectable by the fluorescent stereoscope mCherry signal. All transformed animals were crossed to create transgenic lines, and once established,I am going to proceed with staining using an Anti-mCherry antibody, that will enhance the mCherry signal, in fixed embryos of early and late stages, in order to achieve a more detailed picture of the expression pattern of all chosen putative *Dlle* cis-regulatory elements. The only identified transformants that we were able to express nuclear mCherry

in detectable amounts, and in spatiotemporal pattern similar to the *Dlle* expression (Liubicich *et al.*, 2009; Kao *et al.*, 2016) were carrying the **E1_PhDlle_p_mCherry** cassette in their genome. The observed Go embryos showed a mosaic expression pattern of mCherry expression in the distal parts of the antennae, and thoracic appendages (Fig. 17,18,19). The earliest detected expression of mCherry by fluorescent stereoscope was in embryos approximately five days old. This signal became stronger by the sixth and seventh day of embryogenesis and could be detected in lower levels until the last day of embryogenesis. As would somebody expect, the

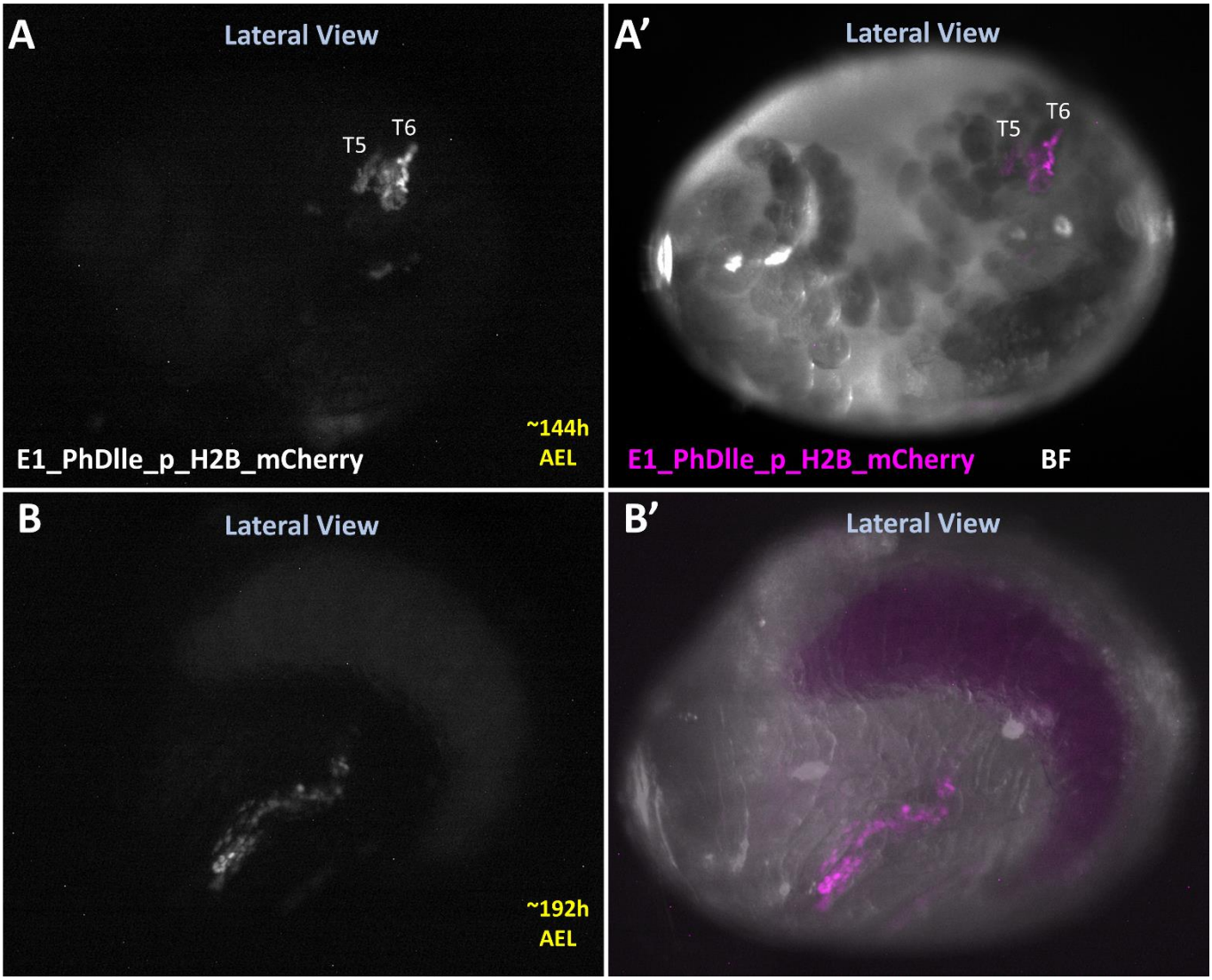

Figure 17: Parhyale embryo transformed with the E1_PhDlle_p_H2B_mCherry construct. The A-P orientation of the embryo is Anterior-left. (A, B) *Nuclear localization of mCherry in the T5 and T6 thoracic appendages in embryos 144 and 192 hours after egg lay (AEL).* ***(A', B')*** *Merged brightfield image of the developing embryo in grey and nuclear mCherry signal in magenta in embryos 144 and 192 hours AEL*

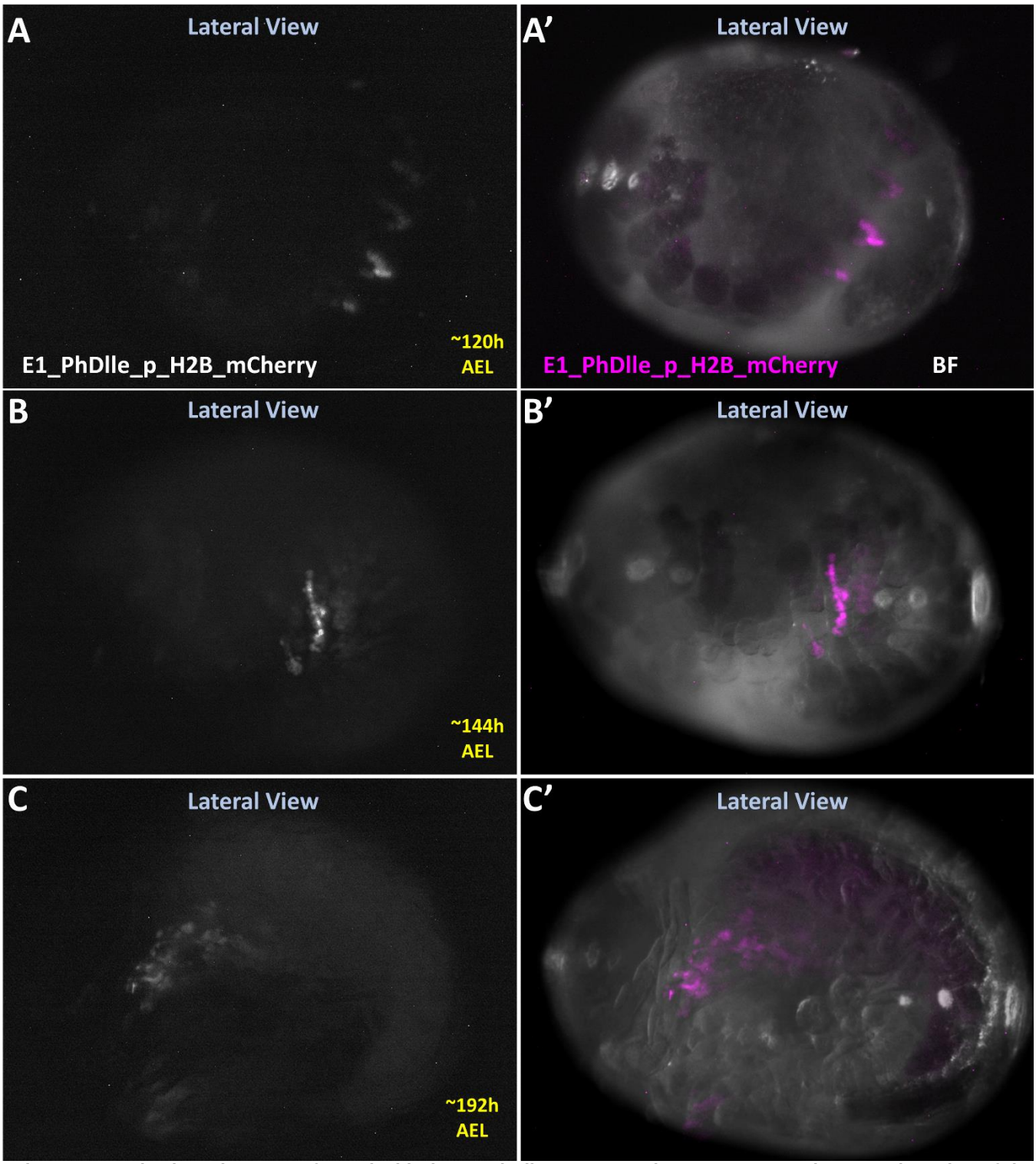

Figure 18: Parhyale embryo transformed with the E1-PhDlle_p_H2B_mCherry construct. The A-P orientation of the embryo is Anterior-left. (A, B, C) *Nuclear localization of mCherry in the T2-T5 thoracic appendages in embryos 120, 144, and 192 AEL.* ***(A', B', C')*** *Merged brightfield image of the developing embryo in grey and nuclear mCherry signal in magenta in embryos 120, 144, and 192 hours AEL*

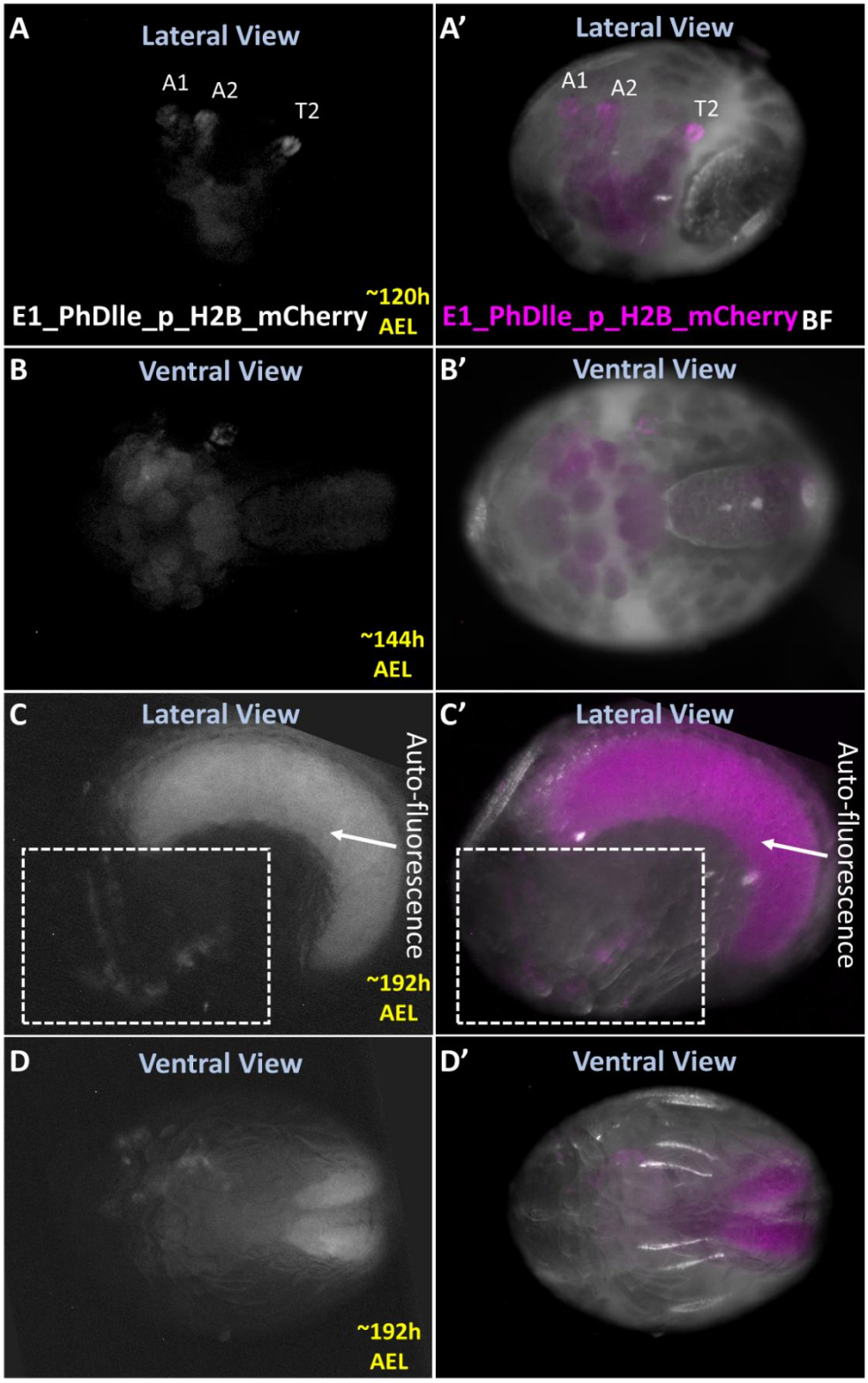

Figure 19: Parhyale embryo transformed with the E1-PhDlle_p-H2B-mCherry construct. The A-P orientation of the embryo is Anterior-left. ***(A, C)*** *Lateral view of embryo expressing nuclear-localized mCherry in the A1, A2 and T2 appendages in embryos 120 and 144 hours AEL.* ***(A', C')*** *Merged lateral view of the brightfield image of the developing embryo in grey and nuclear mCherry signal in magenta 120 and 144 hours AEL.* ***(B, D)*** *Ventral view of embryo expressing nuclear-localized mCherry in the A1, A2 and T2 appendages 120 and 144 hours AEL.* ***(A', C')*** *Merged ventral view of the brightfield image of the developing embryo in grey and nuclear mCherry signal in magenta 120 and 144 hours AEL.*

ATAC-seq peak included in the E1 cloned region contained many repeats of the *de-novo* identified TF binding motif that shared similarity with *pan, brk,* and Mad binding motifs and recognized as a small peak in the Vista plots that compared the *Dlle* loci of three amphipod species. Those evidence strengthen the assumption that this ATAC-seq peak region has a cis-regulatory role in *Dlle* expression across amphipods.

Cross-species test of the *Drosophila Dll* late CREs in *Parhyale*

In parallel with the previous experiments, I also conducted a cross-species functional analysis of the late *Drosophila Dll* CREs in *Parhyale*. To do so, I created reporter constructs of the well-characterized *Drosophila* late *Dll* enhancers LT and LP as described before, excluding the previously tested *Dll304* early enhancer from our analysis. I hypothesized that the very early embryonic phase of broad *Dll* activation and the later restriction in *Drosophila,* as described in the introduction section, are evolutionary innovations not present in *Parhyale*. I assumeed that *Parhyale* embryos may have patterning mechanisms similar to those of the early *Drosophila* leg disk for appendage specification and subdivision into proximal and distal fates. If any activity of those elements is detected in *Parhyale* embryos, this would be a strong indication that they share common regulatory mechanisms in both species.

These constructs were injected into 1-cell stage *Parhyale* embryos (Fig. 14 right side). Although I managed to create one transgenic line (Opsin positive Go and G1) containing the *Drosophila* LP enhancer I detected no mCherry expression under the fluorescence stereoscope, and antibody staining will be needed to uncover whether it is expressed in *Parhyale*. Unfortunately, I was not able to acquire transformed embryos carrying the LT enhancer no matter how many embryos I injected. There are two explanations to this problem. The first one is that all replicates of the plasmid Midi-preparations that I tried were contaminated with RNAses that degraded the *Minos* transposase mRNA used. The second explanation is that this *Drosophila* enhancer element can attract as a silencer in *Parhyale* and prevent the activity of Opsin enhancer and thus the expression of

GFP in the eye. The injection results can be found in the following table (Table 5).

Parhyale hawaiensis injection table				
	Surviving/Injected	Transformed G0 (ops+)	Surviving G0	Surviving (Ops+) G1
pMi {OpsEGFP, PhDlle_p_mCherry}	41/170	2	1	
pMi {OpsEGFP,LP_PhDlle_p_mCherry}	23/95	1	1	3/16
pMi {OpsEGFP,LT_PhDlle_p_mCherry}	94/337	1		

Table 5: Parhyale hawaiensis injection table. *Successful steps are indicated with green boxes, yellow boxes indicate that this step is still in progress, whereas red boxes indicate failure of transgenesis procedure.*

To verify the functionality of the LT and LP enhancers in *Drosophila*, I created modified reporter constructs appropriate for *Drosophila* injections. This means that I used the *Drosophila* endogenous *Dll* promoter upstream of mCherry and the 3xP3-GFP transformation marker in our constructs, see materials and methods (Fig. 15). Additionally, all our constructs contained *AttB* landing sites for integrase mediated transgenesis. After creating stable lines, I wanted to verify the LT and LP expression patterns. To do so, I proceeded with Anti-mCherry antibody staining to amplify the mCherry signal and images the embryos with confocal microscopy. As expected, the LT enhancer could drive the expression of mCherry in the thoracic limb primordia in clusters of cells that formed ring like structures adjacent to cells that express high levels of Wg, as described in the literature (McKay, Estella and Mann, 2009) (Fig. 20 A,B)**.** I was unable to verify the pattern of LP. A possible explanation would be that LP

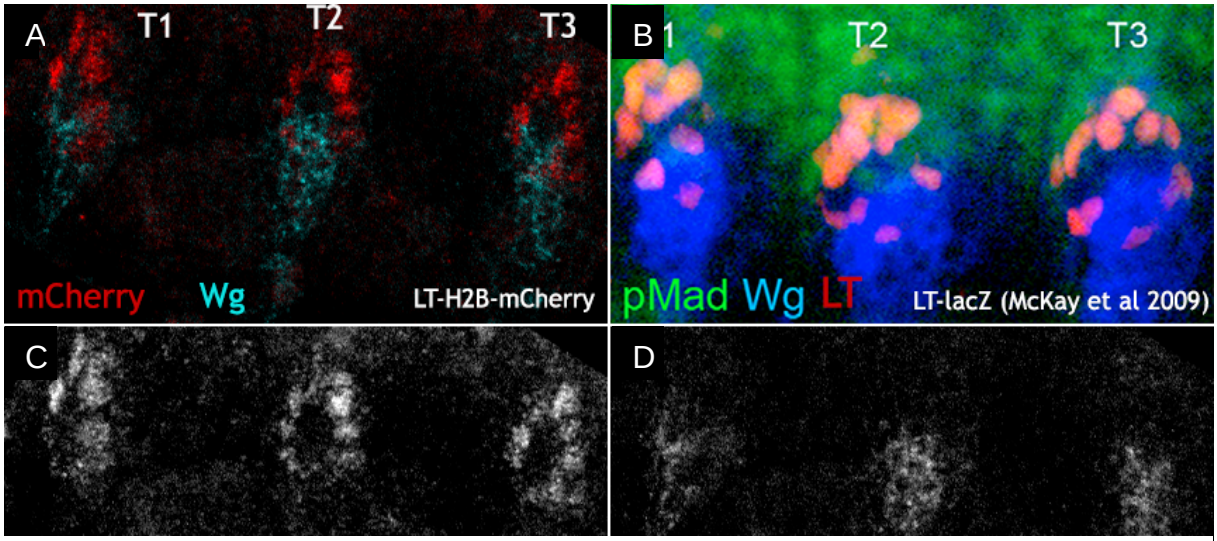

Figure 20: (A) *Nuclear-localized mCherry signal driven by the LT enhancer in cells forming ring like structures in the thoracic limb primordia close to where Wg protein is localized in S14 Drosophila embryos.* ***(B)*** *The original LT enhancer expression pattern published in (McKay et al 2009).* ***(C)*** *mCherry channel.* ***(D)*** *Wg channel.*

enhancer is found 20kb downstream of the 3'UTR of *Dll* I should had cloned him downstream of the reported gene. Another more realistic explanation is that it could drive mCherry expression in undetectable levels. The injection results can be found in the following table (Table 6).

***Drosophila melanogaster* injection table**				
	Injected	**G0 crossed**	**G1 transgenic**	**Lines**
pMi{3xP3_EGFP_DmDlle_p}	**0**			
pMi{3xP3_EGFP_LP_DmDlle_p}	**~600**	**20**	**4**	**3**
pMi{3xP3_EGFP_LT_DmDlle_p}	**~600**	**35**	**6**	**3**
pMi{Ops1_EGGFP_E1_DmDlle_p}	**~200**			

***Table 6: Drosophila melanogaster injection table**. Successful steps are indicated with green boxes, yellow boxes indicate that this step is still in progress, whereas red boxes indicate failure of transgenesis procedure.*

Discussion

Arthropods were able to adapt to variable environments because they diversified their appendages and thus were very successful in colonizing the Earth. Crustaceans in particular, show an immense diversity in the morphology and function of their appendages across species and even within species. Unfortunately, very few animals, hold the beneficial position of being model organisms. The vast technological gaps and lack of resources are separating them from numerous non-model species has biased evolutionary developmental comparisons towards the discovery of similarities rather than differences in the developmental mechanisms that pattern metazoans. One striking example is the common genetic toolbox for patterning outgrowing limbs used by representatives across all three Bilaterian clades (Pueyo and Couso, 2005; Tarazona *et al.*, 2019). In contrast, the mechanisms that specify the limbs in the early embryos can vary dramatically even within one animal group, for example the pancrustaceans (Angelini and Kaufman, 2005b). *Drosophila melanogaster* has been used as the primary model to understand limb specification, patterning, growth and differentiation. These processes take place at different stages of *Drosophila* development throughout embryonic, larval, and pupal development (indirect development), while in most other pancrustaceans (and arthropods), appendages develop as direct outgrowths of body wall of the embryo similar to the vertebrates (direct development).

This study was the first step to our main project, which is to understand the mechanistic basis for the evolutionary transformation of direct into indirect developing appendages by comparing the regulation of the widely conserved limb patterning gene *Distal-less* between the well-studied *Drosophila* and the crustacean amphipod *Parhyale hawaiensis.* This cross-species comparison will provide insights both into the conservation as well as divergence of limb patterning mechanisms.

Distal-less (Dll) encodes a homeo-domain transcription factor that has conserved role in appendage development, and plays a key role in limb specification and PD patterning. *Dll* is first expressed in the limb primordia and later in the distal parts of developing limbs in all arthropods. *Dll* has also another conserved, ancestral, role in the development of the peripheral sensory organs in proximal and distal positions of arthropod appendages but we will focus on its key role in limb specification and PD patterning (Panganiban *et al.*, 1997).

Many details are known regarding regulation of *Dll* in *Drosophila. Dll304* enhancer drives *Dll* expression in the early *Drosophila* embryo in the thoracic appendage primordia and its activity is lost within some hours (Vachon *et al.*, 1992). The *Dll LT* enhancer acts late in *Drosophila* embryogenesis driving *Dll* expression in a ring of cells in the appendage primordia which will develop into the telopod cell fate. The *Dll* LP enhancer is a recently discovered element located 3' of *Dll* transcriptional unit. Both enhancers get positive inputs from the Wg, Dpp, and EGFR pathways (Estella and Mann, 2008; McKay, Estella and Mann, 2009; Galindo *et al.*, 2011).

Many expression studies of *Dll* conducted in diverse crustaceans show that the early clusters of Dll-expressing cells appear in the nascent appendage primordia in an anterior-posterior manner. *Dll* expression initiates in a small number of cells and immediately expands laterally, anteriorly, and posteriorly (Williams, Nulsen and Nagy, 2002; Hejnol and Scholtz, 2004; Prpic, 2008). In contrast with the available data from *Drosophila*, crustacean *Dll* expression becomes confined early to the cells that contribute to the distal region of the developing appendages. Thus, it appears that crustacean limb specification and PD domain identification happen simultaneously. Until today, we lack the knowledge of how

Dll expression initiates in arthropods beyond *Drosophila*, which of the Wg, Dpp, or other pathways are required for limb specification and early PD patterning, and which are the *cis*-regulatory elements that are sensitive to those signaling pathways. It is quite difficult to reach solid conclusions regarding the genetic toolkit that orchestrates appendage specification and early PD patterning in arthropods, due to the diverging expression patterns and the lack of functional data beyond *Drosophila*, especially for pleiotropic genes such as *wg* and *dpp*. The up-to-date data imply that severe modifications and heterochronic shifts of these regulatory mechanisms have occurred during the evolutionary transformation of direct into indirect developing limbs.

Throughout the past year, for the purpose of my Master's thesis, I used a combination of bioinformatic, genomic, and functional genetics approaches to characterize the cis-regulatory elements that mediate the different phases of *Distal-less* regulation in *Parhyale hawaiensis. Parhyale* has proven itself to be a powerful genetic model that is supported by many experimental resources. It is susceptible to embryological manipulations, functional genetic approaches, has the ideal optical properties for advanced microscopic imaging in live or fixed specimens and an increasing palette of genomic resources. To uncover the putative cis-regulatory element (CREs) of the (*Dlle*) I exploited ATAC-seq genomic resources and visualized them using IGV. I assumed that open chromatin regions in the neighborhood of (*Dlle*) could act as its putative CREs. With the assistance of MEME-suite toolkit I identified bioinformatically two *de-novo* transcription factor binding motifs, one of which showed significant similarity to *pan, brk*, and Mad binding motifs, which correspond to downstream transcription factors of *wg* and *dpp* pathways that directly regulate *Dll* in *Drosophila*. The other binding motif resembled the ones of the leg gap gene *Lim1*, and *Dll*, which is also known to self-regulate its expression. Both motifs were also similar to other factors that until now were not involved in *Dll* regulation. In addition, both *de-novo* identified motifs were over-represented in the ATAC-seq peak regions. By Vista-plot analysis comparing the *Parhyale Dlle* genomic locus with two other available amphipod genomes, the one of *Trinorchestia longiramous* and *Hyalella azteca* I identified two distinct similarity peaks overlapping with two of the identified ATAC-seq peaks those peaks. I also

observed enrichment of the *de-novo* motifs at those Vista plot peaks. Those conserved between the three species non-coding regions that came in agreement with our genomic approach might have a functional role, regulating *Dlle*. The biological functionality of each putative CRE of *Dlle* was tested by creating reporter constructs of selected regions of length 2-3kb found in *Dlle* locus. Those constructs were microinjected in 1-cell stage *Parhyale* embryos. The well characterized LT and LP *Drosophila* enhancers were also inserted in reporter constructs and tested for their cross-species expression in *Parhyale.*

Concluding, I showed that the combination of ATAC-seq and bioinformatics approaches can identify putative CREs in the *Parhyale hawaiensis Dlle* locus. Additionally, I identified at least one CRE (E1) that drives reporter gene expression in *Parhyale* limbs from early limb bud stage onwards, recapitulating the endogenous *Dlle* expression pattern. Using the bioinformatic tool MEME-suite I found that E1 contains putative binding sites for known regulators of *Dll* in *Drosophila* (*pan*, *brk*, Mad, *Dll*) as well as new regulators that have not been implicated in *Dll* regulation before. Last but not least, phylogenetic foot-printing analysis between amphipod genomes with Vistal tool, further corroborates the involvement of identified CREs in *Dll* regulation.

Future perspectives

I plan to finish the creation of transgenic lines and proceed with a more detailed expression analysis of all identified CREs. I also plan to acquire ATAC-seq profiles for earlier developmental stages of *Parhyale*, and I will further proceed with the analysis of more ATAC-seq peaks that might act as *Dlle* CREs in earlier developmental stages, using similar experimental strategies. For the detailed live-imaging analysis of the CRE reporter expression pattern, at the high spatial and temporal resolution, I will use light-sheet fluorescence microscopy together with advanced imaging techniques for cell tracking. Once I identify which cloned fragments can represent the *Dlle* expression pattern I will repeat the transgenesis procedure creating deletions to achieve the minimum functional size of the CREs. Subsequently, the expression of all the functional minimal *Parhyale* CRE reporters will be analyzed in *Drosophila*. Identifying *Dll* CREs of *Parhyale* that can be functional in *Drosophila,* can as well help us to use the

sophisticated genetics of *Drosophila* for the identification of the upstream regulators of *Parhyale Dll* CRE*s.*

Next, I will continue with the characterization of the putative molecular pathways involved in the *Parhyale* early *Dlle* regulation. To do so, I are planning to create and analyze RNA-seq datasets from FACS sorted cell populations that express mCherry under the control of the minimal identified CREs. This will allow us to identify which of the 1143 transcription factors of *Parhyale* (Kao *et al.*, 2016) are overrepresented in each dataset. Once I confirm that the *Parhyale Dll* CRE reporters can be active in *Drosophila* I plan to continue with functional analysis doing RNAi screens in *Drosophila* against the identified TF homologs. To avoid pleiotropic effects, I will work in a tissue-specific manner. To do so, I will create GAL4 lines in which GAL4 will be driven by our minimal CREs. The CRE-GAL4 lines will be then crossed with a UAS:mCherry line to create CRE-GAL4;UAS:mCherry stocks. To knock-down the *Drosophila* homologs of the identified TFs, in the CRE-expressing cells, will use UAS:RNAi stocks supplied from the large VDRC RNAi stock collection. The readout of these experiments is going to be altered expression of mCherry. On a complementary approach, I will use commercially available small-molecule drugs or peptides to inhibit or increase (using agonists) the activity of candidate molecular pathways such as Wnt, TGF-beta, EGFR that are known to regulate *Dll* in *Drosophila* (Estella and Mann, 2008; Galindo *et al.*, 2011) as well as other pathways indicated by the tissue specific RNAseq dataset bioinformatic analysis. Transgenic animals carrying the minimal CRE reporters will be drug-treated or injected, in a temporally controlled manner few hours before the appropriate stage, and screened for the expression of the CRE reporters. The results of this objective will help us uncover which molecular pathways are conserved and have a key role in the crustacean direct limb development regulation machinery.

After identifying which molecular pathways and downstream effectors are involved in the regulation of *Dll* through its cis-regulatory elements, I will aim to characterize the specific DNA binding sites of the downstream Transcription Factors. For this purpose, I will conduct Electrophoretic Mobility Shift Assays (EMSA). I will clone the selected *Parhyale* TFs that were able to perturb the reporter activity, in frame with affinity tags for further expression and purification

and amplify the selected minimal CRE sequences together with their shortened versions by creating 50-100bp deletions. This assay will show us the regions on the CREs at which the purified proteins directly bind to. After that, I can use PCR directed mutagenesis to insert mutations on the characterized wild-type binding sites to score for reduced TF binding, biochemically, and *in vivo* by altered reporter gene expression. An alternative approach will be to CRISPR mutate the endogenous CREs at the specific TF binding sites of *Dll* using the already available *Dlle-KI* line (Kao *et al.*, 2016) and score for altered *Dll* expression or mild to severe phenotypes such as those of *Dll* mutants, if shadow enhancers are absent.

This study addresses if, how, and when widely conserved limb patterning genes and signalling pathways are being used in appendage specification and outgrowth of *Parhyale*. After fulfilling all our tasks, I will be able to number which pathways are involved, which genes regulate *Dll* directly as well as their exact binding sites. I will be also capable of providing strong evidence about the conservation or divergence of mechanisms that underlie *Parhyale* early *Dll* regulation at the level of cis-regulatory elements.

References

Alwes, F., Hinchen, B. and Extavour, C. G. (2011) 'Patterns of cell lineage, movement, and migration from germ layer specification to gastrulation in the amphipod crustacean Parhyale hawaiensis', *Developmental Biology*. Elsevier Inc., 359(1), pp. 110–123. doi: 10.1016/j.ydbio.2011.07.029.

Angelini, D. R. and Kaufman, T. C. (2005a) 'Functional analyses in the milkweed bug Oncopeltus fasciatus (Hemiptera) support a role for Wnt signaling in body segmentation but not appendage development', *Developmental Biology*, 283(2), pp. 409–423. doi: 10.1016/j.ydbio.2005.04.034.

Angelini, D. R. and Kaufman, T. C. (2005b) 'Insect appendages and comparative ontogenetics', *Developmental Biology*, 286(1), pp. 57–77. doi: 10.1016/j.ydbio.2005.07.006.

Bailey, T. L. *et al.* (2009) 'MEME SUITE: tools for motif discovery and searching', *Nucleic acids research*. 2009/05/20. Oxford University Press, 37(Web Server

issue), pp. W202–W208. doi: 10.1093/nar/gkp335.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) ‘Trimmomatic: a flexible trimmer for Illumina sequence data’, *Bioinformatics (Oxford, England)*. 2014/04/01. Oxford University Press, 30(15), pp. 2114–2120. doi: 10.1093/bioinformatics/btu170.

Browne, W. E. *et al.* (2005) ‘Stages of embryonic development in the amphipod crustacean, Parhyale hawaiensis’, *Genesis*, 42(3), pp. 124–149. doi: 10.1002/gene.20145.

Brudno, M. *et al.* (2003) ‘LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA.’, *Genome research*, 13(4), pp. 721–731. doi: 10.1101/gr.926603.

Chaw, R. C. and Patel, N. H. (2012) ‘Independent migration of cell populations in the early gastrulation of the amphipod crustacean Parhyale hawaiensis’, *Developmental Biology*. Elsevier, 371(1), pp. 94–109. doi: 10.1016/j.ydbio.2012.08.012.

Cohen, B., Simcox, A. A. and Cohen, S. M. (1993) ‘Allocation of the thoracic imaginal primordia in the Drosophila embryo’, *Development*, 117(2), pp. 597–608.

Constantinou, S. J. *et al.* (2016) ‘Wnt repertoire and developmental expression patterns in the crustacean Thamnocephalus platyurus’, *Evolution and Development*, 18(5–6), pp. 324–341. doi: 10.1111/ede.12204.

Dohle, W. and Scholtz, G. (1988) ‘Glonal analysis’, 160, pp. 147–160.

Duman-Scheel, M., Pirkl, N. and Patel, N. H. (2002) ‘Analysis of the expression pattern of Mysidium columbiae wingless provides evidence for conserved mesodermal and retinal patterning processes among insects and crustaceans’, *Development Genes and Evolution*, 212(3), pp. 114–123. doi: 10.1007/s00427-002-0215-6.

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

Estella, C., McKay, D. J. and Mann, R. S. (2008) 'Molecular Integration of Wingless, Decapentaplegic, and Autoregulatory Inputs into Distalless during Drosophila Leg Development', *Developmental Cell*, 14(1), pp. 86–96. doi: 10.1016/j.devcel.2007.11.002.

Extavour, C. G. (2005) 'The fate of isolated blastomeres with respect to germ cell formation in the amphipod crustacean Parhyale hawaiensis', *Developmental Biology*, 277(2), pp. 387–402. doi: https://doi.org/10.1016/j.ydbio.2004.09.030.

Franz, G. and Savakis, C. (1991) 'Minos, a new transposable element from Drosophila hydei, is a member of the Tc1-like family of transposons.', *Nucleic acids research*, 19(23), p. 6646. doi: 10.1093/nar/19.23.6646.

Frazer, K. A. *et al.* (2004) 'VISTA: computational tools for comparative genomics.', *Nucleic acids research*, 32(Web Server issue), pp. W273-9. doi: 10.1093/nar/gkh458.

Galindo, M. I. *et al.* (2011) 'Control of Distal-less expression in the Drosophila appendages by functional 3' enhancers', *Developmental Biology*. Elsevier B.V., 353(2), pp. 396–410. doi: 10.1016/j.ydbio.2011.02.005.

Gerberding, M., Browne, W. E. and Patel, N. H. (2002) 'Cell lineage analysis of the amphipod crustacean Parhyale hawaiensis reveals an early restriction of cell fates', *Development*, 129(24), pp. 5789–5801. doi: 10.1242/dev.00155.

Gilles, A. F., Schinko, J. B. and Averof, M. (2015) 'Efficient CRISPR-mediated gene targeting and transgene replacement in the beetle Tribolium castaneum', *Development*, 142(16), pp. 2832–2839. doi: 10.1242/dev.125054.

González-Crespo, S. and Morata, G. (1996) 'Genetic evidence for the subdivision of the arthropod limb into coxopodite and telopodite.', *Development (Cambridge, England)*. England, 122(12), pp. 3921–3928.

Goto, S. and Hayashi, S. (1997) 'Specification of the embryonic limb primordium by graded activity of Decapentaplegic', *Development*, 124(1), pp. 125–132.

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

10.1093/bioinformatics/btr064.

Gupta, S. *et al.* (2007) ‘Quantifying similarity between motifs’, *Genome Biology*, 8(2), p. R24. doi: 10.1186/gb-2007-8-2-r24.

Hannibal, R. L., Price, A. L. and Patel, N. H. (2012) ‘The functional relationship between ectodermal and mesodermal segmentation in the crustacean, Parhyale hawaiensis’, *Developmental Biology*, 361(2), pp. 427–438. doi: https://doi.org/10.1016/j.ydbio.2011.09.033.

Hejnol, A. and Scholtz, G. (2004) ‘Clonal analysis of Distal-less and engrailed expression patterns during early morphogenesis of uniramous and biramous crustacean limbs’, *Development Genes and Evolution*, 214(10), pp. 473–485. doi: 10.1007/s00427-004-0424-2.

Jockusch, E. L., Williams, T. A. and Nagy, L. M. (2004) ‘The evolution of patterning of serially homologous appendages in insects’, *Development Genes and Evolution*, 214(7), pp. 324–338. doi: 10.1007/s00427-004-0412-6.

von Kalm, L., Fristrom, D. and Fristrom, J. (1995) ‘The making of a fly leg: a model for epithelial morphogenesis.’, *BioEssays : news and reviews in molecular, cellular and developmental biology*. United States, 17(8), pp. 693–702. doi: 10.1002/bies.950170806.

Kao, D. *et al.* (2016) ‘The genome of the crustacean Parhyale hawaiensis, a model for animal development, regeneration, immunity and lignocellulose digestion’, *eLife*, 5, pp. 1–45. doi: 10.7554/elife.20062.

Kontarakis, Z., Pavlopoulos, A., *et al.* (2011) ‘A versatile strategy for gene trapping and trap conversion in emerging model organisms’, *Development*, 138(12), pp. 2625–2630. doi: 10.1242/dev.066324.

Kontarakis, Z., Konstantinides, N., *et al.* (2011) ‘Reconfiguring gene traps for new tasks using iTRAC’, *Fly*, 5(4), pp. 352–355. doi: 10.4161/fly.5.4.18108.

Kontarakis, Z. and Pavlopoulos, A. (2014) *Transgenesis in non- model organisms: the case of Parhyale.* Methods in Molecular Biology 1196:145– 181. doi: 10.1007/978-1-4939-1242-1.

Kubota, K. *et al.* (2000) ‘EGF receptor attenuates Dpp signaling and helps to

distinguish the wing and leg cell fates in Drosophila', *Development*, 127(17), pp. 3769–3776.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools.', *Bioinformatics (Oxford, England)*, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.

Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows–Wheeler transform', *Bioinformatics*, 25(14), pp. 1754–1760. doi: 10.1093/bioinformatics/btp324.

Liubicich, D. M. *et al.* (2009) 'Knockdown of Parhyale Ultrabithorax recapitulates evolutionary changes in crustacean appendage morphology', *Proceedings of the National Academy of Sciences of the United States of America*, 106(33), pp. 13892–13896. doi: 10.1073/pnas.0903105106.

Martin, A., Serano, Julia M., *et al.* (2016) 'CRISPR/Cas9 Mutagenesis Reveals Versatile Roles of Hox Genes in Crustacean Limb Specification and Evolution', *Current Biology*. Elsevier Ltd, 26(1), pp. 14–26. doi: 10.1016/j.cub.2015.11.021.

Martin, A., Serano, Julia M., *et al.* (2016) 'CRISPR/Cas9 Mutagenesis Reveals Versatile Roles of Hox Genes in Crustacean Limb Specification and Evolution', *Current Biology*, 26(1), pp. 14–26. doi: https://doi.org/10.1016/j.cub.2015.11.021.

McKay, D. J., Estella, C. and Mann, R. S. (2009) 'The origins of the Drosophila leg revealed by the cis-regulatory architecture of the Distalless gene', *Development*, 136(1), pp. 61–71. doi: 10.1242/dev.029975.

McLeay, R. C. and Bailey, T. L. (2010) 'Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data', *BMC Bioinformatics*, 11(1), p. 165. doi: 10.1186/1471-2105-11-165.

Niwa, N. *et al.* (2000) 'Niwa et al. 2000', 4381, pp. 4373–4381.

Ober, K. A. and Jockusch, E. L. (2006) 'The roles of wingless and decapentaplegic in axis and appendage development in the red flour beetle, Tribolium castaneum', *Developmental Biology*, 294(2), pp. 391–405. doi: 10.1016/j.ydbio.2006.02.053.

Panganiban, G. *et al.* (1995) ‘The development of crustacean limbs and the evolution of arthropods’, *Science*, 270(5240), pp. 1363–1366. doi: 10.1126/science.270.5240.1363.

Panganiban, G. *et al.* (1997) ‘The origin and evolution of animal appendages’, *Proceedings of the National Academy of Sciences of the United States of America*, 94(10), pp. 5162–5166. doi: 10.1073/pnas.94.10.5162.

Panganiban, G. (2000) ‘A PEER REVIEWED FORUM Distal-less Function During Drosophila Appendage and’, *Developmental Dynamics*, 562(June), pp. 554–562.

Parchem, R. J. *et al.* (2010) ‘BAC library for the amphipod crustacean, Parhyale hawaiensis’, *Genomics*, 95(5), pp. 261–267. doi: https://doi.org/10.1016/j.ygeno.2010.03.005.

Pavlopoulos, A. *et al.* (2009) ‘Probing the evolution of appendage specialization by Hox gene misexpression in an emerging model crustacean’, *Proceedings of the National Academy of Sciences of the United States of America*, 106(33), pp. 13897–13902. doi: 10.1073/pnas.0902804106.

Pavlopoulos, A. and Averof, M. (2005) ‘Establishing genetic transformation for comparative developmental studies in the crustacean Parhyale hawaiensis’, *Proceedings of the National Academy of Sciences of the United States of America*, 102(22), pp. 7888–7893. doi: 10.1073/pnas.0501101102.

Pavlopoulos, A. and Wolff (2020) *CRUSTACEAN LIMB MORPHOGENESIS DURING NORMAL*.

Prpic, N. M. *et al.* (2003) ‘Gene expression in spider appendages reveals reversal of exd/hth spatial specificity, altered leg gap gene dynamics, and suggests divergent distal morphogen signaling’, *Developmental Biology*, 264(1), pp. 119–140. doi: 10.1016/j.ydbio.2003.08.002.

Prpic, N. M. (2004) ‘Homologs of wingless and decapentaplegic display a complex and dynamic expression profile during appendage development in the millipede Glomeris marginata (Myriapoda: Diplopoda)’, *Frontiers in Zoology*, 1, pp. 1–12. doi: 10.1186/1742-9994-1-6.

Prpic, N. M. (2008) ‘Parasegmental appendage allocation in annelids and arthropods and the homology of parapodia and arthropodia’, *Frontiers in Zoology*, 5, pp. 1–5. doi: 10.1186/1742-9994-5-17.

Pueyo, J. I. and Couso, J. P. (2005) ‘Parallels between the proximal-distal development of vertebrate and arthropod appendages: homology without an ancestor?’, *Current opinion in genetics & development*. England, 15(4), pp. 439–446. doi: 10.1016/j.gde.2005.06.007.

Quinlan, A. R. and Hall, I. M. (2010) ‘BEDTools: a flexible suite of utilities for comparing genomic features’, *Bioinformatics*, 26(6), pp. 841–842. doi: 10.1093/bioinformatics/btq033.

Rehm, E. J., Hannibal, R. L., Chaw, R. C., *et al.* (2009) ‘In situ hybridization of labeled RNA probes to fixed Parhyale Hawaiensis embroys’, *Cold Spring Harbor Protocols*, 4(1), pp. 1–6. doi: 10.1101/pdb.prot5130.

Rehm, E. J., Hannibal, R. L., Chaw, C. R., *et al.* (2009) ‘Injection of Parhyale hawaiensis blastomeres with fluorescently labeled tracers’, *Cold Spring Harbor Protocols*, 4(1), pp. 1–6. doi: 10.1101/pdb.prot5128.

Rieckhof, G. E. *et al.* (1997) ‘Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein.’, *Cell*. United States, 91(2), pp. 171–183. doi: 10.1016/s0092-8674(00)80400-6.

Robinson, G. E. *et al.* (2011) ‘Creating a Buzz About Insect Genomes’, *Science*. Edited by J. Sills, 331(6023), pp. 1386 LP – 1386. doi: 10.1126/science.331.6023.1386.

Serano, J. M. *et al.* (2016) ‘Comprehensive analysis of Hox gene expression in the amphipod crustacean Parhyale hawaiensis’, *Developmental Biology*. Elsevier, 409(1), pp. 297–309. doi: 10.1016/j.ydbio.2015.10.029.

Stamataki, E. and Pavlopoulos, A. (2016) ‘Non-insect crustacean models in developmental genetics including an encomium to Parhyale hawaiensis’, *Current Opinion in Genetics and Development*. Elsevier Ltd, 39, pp. 149–156. doi: 10.1016/j.gde.2016.07.004.

Sun, D. A. and Patel, N. H. (2019) ‘The amphipod crustacean Parhyale

hawaiensis: An emerging comparative model of arthropod development, evolution, and regeneration', *Wiley Interdisciplinary Reviews: Developmental Biology*, 8(5), pp. 1–20. doi: 10.1002/wdev.355.

Tarazona, O. A. *et al.* (2019) 'Evolution of limb development in cephalopod mollusks', *eLife*, 8, pp. 1–19. doi: 10.7554/eLife.43828.001.

Vachon, G. *et al.* (1992) 'Homeotic genes of the bithorax complex repress limb development in the abdomen of the Drosophila embryo through the target gene Distal-less', *Cell*, 71(3), pp. 437–450. doi: 10.1016/0092-8674(92)90513-C.

VanHook, A. M. and Patel, N. H. (2008) 'Crustaceans', *Current Biology*, 18(13), pp. R547–R550. doi: https://doi.org/10.1016/j.cub.2008.05.021.

Williams, T. A., Nulsen, C. and Nagy, L. M. (2002) 'A complex role for Distal-less in crustacean appendage development', *Developmental Biology*, 241(2), pp. 302–312. doi: 10.1006/dbio.2001.0497.

Wolff, C. *et al.* (2018) 'Multi-view light-sheet imaging and tracking with the MaMuT software reveals the cell lineage of a direct developing arthropod limb', *eLife*, 7, pp. 1–31. doi: 10.7554/eLife.34410.

Zhang, Y. *et al.* (2008) 'Model-based Analysis of ChIP-Seq (MACS)', *Genome Biology*, 9(9), p. R137. doi: 10.1186/gb-2008-9-9-r137.