An Atlas of Transcription Factors Expressed in Male Pupal Terminalia of Drosophila melanogaster

Ben J. Vincent,*,1 Gavin R. Rice,*,1 Gabriella M. Wong,* William J. Glassford,*,2 Kayla I. Downs,* Jessica L. Shastay,* Kenechukwu Charles-Obi,* Malini Natarajan,1,3 Madelaine Gogol,† Julia Zeitlinger,†,‡ and Mark Rebeiz*,4
*Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, †Stowers Institute for Medical Research, Kansas City, Missouri, and ‡Department of Pathology and Laboratory Medicine, Kansas University Medical Center, Kansas City, 66160

ABSTRACT During development, transcription factors and signaling molecules govern gene regulatory networks to direct the formation of unique morphologies. As changes in gene regulatory networks are often implicated in morphological evolution, mapping transcription factor landscapes is important, especially in tissues that undergo rapid evolutionary change. The terminalia (genital and anal structures) of Drosophila melanogaster and its close relatives exhibit dramatic changes in morphology between species. While previous studies have identified network components important for patterning the larval genital disc, the networks governing adult structures during pupal development have remained uncharted. Here, we performed RNA-seq in whole Drosophila melanogaster male terminalia followed by in situ hybridization for 100 highly expressed transcription factors during pupal development. We find that the male terminalia are highly patterned during pupal stages and that specific transcription factors mark separate structures and substructures. Our results are housed online in a searchable database (https://flyterminalia.pitt.edu/) as a resource for the community. This work lays a foundation for future investigations into the gene regulatory networks governing the development and evolution of Drosophila terminalia.

KEYWORDS gene regulation development Drosophila transcription factors terminalia

As animal development proceeds, transcription factors and signaling molecules are expressed in precise patterns to specify cell fate in space and time (Levine and Davidson 2005). These genes ultimately impinge upon cellular effectors, forming gene regulatory networks that alter cellular behavior and generate complex morphologies (Smith et al. 2018). Changes within gene regulatory networks can have cellular consequences and result in morphological differences between species (Prud’homme et al. 2007; Thompson et al. 2015; Halfon 2017). To understand how body parts are built during development and modified through evolution, we must define and dissect their relevant gene regulatory networks.

Of all the anatomical parts in the animal body plan, genitalia have been of particular interest for many evolutionary questions. Across animals, genital morphology diverges rapidly between species (Hosken and Stockley 2004; Simmons 2014). These observations have led some to theorize that male and female genitalia co-evolve under conditions of sexual conflict (Brennan and Prum 2015). Others theorize that cryptic female choice has facilitated morphological divergence (Eberhard 1985). The accumulation of divergent morphologies between species may lead to miscoupling of genitalia during interbreeding, reducing viability or fecundity (Masly 2011; Yassin and David 2016; Tanaka et al. 2018). In Drosophila, genital morphology is also critical for taxonomic classification, as it is often the only way to reliably identify species that are otherwise morphologically
indistinguishable (Okada 1954; Bock and Wheeler 1972; Kamimura and Mitsumoto 2011). Previous studies have highlighted several novel genital morphologies that may provide insights into how new traits evolve (Kopp and True 2002; Yassin and Orgogozo 2013). Despite their intensive study, the molecular basis of genital evolution remains poorly understood.

The genitalia of Drosophila melanogaster and its close relatives provide a unique opportunity to determine how gene regulatory networks build complex and evolving structures. Most previous work on genital development has focused on the larval genital disc, where transcriptomics and targeted genetic experiments have identified several genes that alter adult genitalia when perturbed (Chen and Baker 1997; Gorfinkiel et al. 1999; Keisman and Baker 2001; Chatterjee et al. 2011). However, much less is known about the genes that control genital development during metamorphosis, when many of the adult structures form through epithelial remodeling (Glassford et al. 2015). Quantitative trait locus (QTL) mapping studies have also been performed in Drosophila and have identified several large genomic regions that contribute to genital diversification between crossable sister species (Macdonald and Goldstein 1999; Zeng et al. 2000; Masly et al. 2011; McNeil et al. 2011; Tanaka et al. 2015; Takahara and Takahashi 2015). An examination of the gene regulatory networks which govern development of these structures during pupal stages may yield insights into the developmental partitioning of a complex tissue, the causative genes that underlie morphological differences between species, and the origins of novel traits.

The adult male terminalia (comprising both the genitalia and analia) of D. melanogaster are subdivided into five main structures, following recently revised nomenclature (Rice et al. 2019b): the hypandrium, phallus, surstylus (clasper), epandrial ventral lobe (EVL, also known as the lateral plate), and cercus (also known as the anal plate) (Figure 1A). By 28 hr after puparium formation (APF), four structures can be distinguished in the developing terminalia: hypandrium, phallus, cercus, and the tissue which will give rise to the EVL and surstylus (Figure 1B). By 48 hr APF, the pupal terminalia effectively prefigure adult structures – the surstylus and EVL have separated, and the epandrial posterior lobe has formed along with many other substructures associated with the hypandrium and phallus (Figure 1B). Therefore, in less than 1 day, the pupal terminalia undergo a dramatic remodeling process that builds many adult structures. This rapid transformation motivated our search for transcription factors that pattern these structures during pupal development.

In this study, we performed RNA-seq in male terminalia during early pupal development and identified highly expressed transcription factors that may operate during this stage. We then used in situ hybridization to build a gene expression atlas of 100 transcription factors in the male pupal terminalia at two time points during development. Most of these genes were highly patterned, especially at 48 hr APF, and we identified genetic markers for many structures and substructures that exhibit morphological differences between Drosophilids. Our data are housed in a searchable online database (https://flyterminalia.pitt.edu/) that will expand as new expression patterns are charted. We believe that the transcription factors characterized here draw the outlines of gene regulatory networks that control genital development and evolution in Drosophila.

**MATERIALS AND METHODS**

Detailed, formatted protocols for probe design and synthesis, sample collection, dissection and fixation, and in situ hybridization can be found at https://flyterminalia.pitt.edu/.

---

**RNA-seq and transcriptomic analysis**

RNA was isolated from single pupal terminal samples (genotype: yw;+;+;+) dissected at 24 hr APF or 28 hr APF using the Maxwell 16 Tissue RNA Purification Kit (Promega). Poly-A RNA-seq libraries were generated using a Clontech library preparation kit (040215). Individual libraries from four different samples were generated for...
each time point, and libraries were sequenced on an Illumina HiSeq 2500. Sequencing reads from 3 lanes of 51-base Hi-seq data were aligned with tophat (2.0.13) to the dm3 assembly (Trapnell et al. 2009), which was retrieved from the UCSC Genome Browser with annotations from Flybase (ftp://ftp.flybase.net/genomes/Drosophila_melanogaster/dmelR5.57_FB2014_03/gff). Reads were counted in unionized exons using bedtools count (Quinlan and Hall 2010). Genes expressed in the terminalia were compared to the FlyTF list of annotated transcription factors found at https://www.mrc-lmb.cam.ac.uk genomics/FlyTF/. (Pfreundt et al. 2010).

**Probe design and synthesis**

Templates for 200-300 basepair RNA probes were designed from a large exon present in all annotated isoforms of each examined gene. Exons were chosen by retrieving the decorated FASTA from flybase.org, and annotated isoforms were examined using the UCSC genome browser. After exon selection, Primer3Plus (Untergasser et al. 2007) was used to design PCR primers that would amplify a 200-300 base pair region, and 5-10 candidate primer pairs were screened using the UCSC In Silico PCR tool to identify sets that will amplify the region of interest from the most diverged Drosophilid species possible. This screening process was implemented to maximize the utility of any particular primer set for other species. Reverse primers were designed beginning with a T7 RNA polymerase binding sequence (TAATAC-GACTCACTATAG), and template DNA was PCR amplified from adult fly genomic DNA extracted using the DNeasy kit (QIAGEN). Digoxigenin-labeled probes were then synthesized using in vitro transcription (T7 RNA Polymerase, Promega / Life Technologies), ethanol precipitated, and resuspended in water for Nanodrop analysis. Probes were stored at -20°C in 50% formamide prior to in situ hybridization.

**Sample collection, dissection and fixation**

Male D. melanogaster white pre-pupa (genotype: yw;+;+) were collected at room temperature and incubated in a petri dish containing a moistened Kimwipe at 25°C for 24 hr or 48 hr prior to dissection. After incubation, pupae were impaled in their anterior region and immobilized within a glass dissecting well containing cold Phosphate Buffered Saline (PBS). The posterior tip of the pupa (20-40% of pupal length) was separated and washed with a P200 pipette to flush the pupal terminalia into solution. Samples were then collected in PBS with 0.1% Triton-X-100 (PBT) and 4% paraformaldehyde (PFA, E.M.S. Scientific) on ice, and multiple samples were collected in the same tube. Samples were then fixed in PBT + PFA at room temperature for 30 min, washed twice in methanol and twice in ethanol at room temperature, and stored at -20°C.

**In situ hybridization and imaging**

We used an InsituPro VSI robot to perform in situ hybridization. Briefly, dissected terminalia were rehydrated in PBT, fixed in PBT with 4% PFA and prehybridized in hybridization buffer for 1 hr at 65°C. Samples were then incubated with probe for 16h at 65°C before washing with hybridization buffer and PBT. Samples were blocked in PBT with 1% bovine serum albumin (PBT+BSA) for 2 hr. Samples were then incubated with anti-digoxigenin Fab fragments conjugated to alkaline phosphatase (Roche) diluted 1:6000 in PBT+BSA. After additional washes, color reactions were performed by incubating samples with NBT and BCIP (Promega) until purple stain could be detected under a dissecting microscope. Samples were mounted in glycerol on microscope slides coated with poly-L-lysine and imaged at 20X or 40X magnification on a Leica DM 2000 with a Leica DFC450C camera. For most images available online, extended focus compilations were acquired using the ImageBuilder module of the Leica Application Suite.

In interpreting our results, we performed several qualitative comparisons to increase our confidence in the data. First, we processed samples from both time points simultaneously in the same basket and staining well. For many genes, we observed uniform expression in 28h samples but patterned expression in 48h samples. These observations gave us confidence that the uniform early expression was not due to background staining. Similarly, we occasionally observed expression patterns in samples from one time point but not the other, which fostered confidence that the absence of expression was not due to experimental failure. As an additional safeguard, we compared results from different genes stained in the same batch to detect cross-contamination. Finally, we compared equivalent samples in annotating our results, such that the representative images presented in this manuscript were corroborated by replicates.

**Data Availability**

Images and experimental details for all samples that met the quality control standards of our experimental pipeline can be found at https://flyterminalia.pitt.edu/. Gene expression data are available at NCBI Gene Expression Omnibus (GEO) with the accession number: GSE133732. Supplemental material available at figshare: https://doi.org/10.25387/g3.9983081.

**RESULTS**

**Global measurements of gene expression levels in early pupal terminalia**

To identify transcription factors that may play a role in genital and anal development, we performed RNA-seq on early pupal terminalia dissected at 24 hr and 28 hr after puparium formation (APF). We chose these time points because we wanted to identify candidate regulators that dynamically control the development of terminal structures, and these time points immediately precede differentiation events that result in the formation of the epandrial posterior lobe, surstylius and phallic structures (Glassford et al. 2015; Smith et al. 2019). We found that 11,816 genes are expressed at levels greater than 1 read per kilobase per million reads (rpkm) in at least 1 time point, including 282 annotated transcription factors (Pfreundt et al. 2010). We found that expression measurements from both time points were broadly correlated (Figure S1), which built confidence in our results, and we focused on results at 28 hr APF due to the ease of dissection at that time point. Among the 100 most highly expressed transcription factors at 28 hr APF, the expression levels ranged from 442 to 27 rpkm (Figure 1C). These genes formed the basis for our gene expression atlas.

**An atlas of the genital transcription factor landscape**

Our transcriptomic analysis suggested that a large number of transcription factors are expressed in the pupal terminalia. In order to glean spatial and temporal expression information for these candidates, we performed in situ hybridization (ISH) in pupal terminalia at 28 hr and 48 hr APF. ISH measurements are qualitative and variable – distinguishing signal from background can be challenging, especially for genes that are uniformly expressed, and results may vary between biological replicates. We addressed these challenges through several comparisons (see Materials and Methods). In addition to the results presented here, our full dataset is housed online at https://flyterminalia.pitt.edu/. We built this database to
increase the accessibility, transparency and reproducibility of our results. We include full protocols for our methods as well as key experimental details underlying the results for each experiment. For each gene, we also include annotations of all tissues in which evidence of gene expression was observed. While we focused here on expression patterns corresponding to external structures, we also observed expression patterns associated with internal terminalia, but chose not to annotate these patterns due to the lack of morphological markers for these structures. Finally, to accurately represent the variability in our results, this database includes images of all samples that met the quality control standards of our experimental pipeline.

For the remainder of the manuscript, we organize our results by describing select transcription factors expressed in each structure of the terminalia.

The epandrial ventral lobe (lateral plate)
The epandrial ventral lobe (EVL, also called the lateral plate) is a periphallic structure lateral to the phallus (Rice et al. 2019b). The epandrial posterior lobe (hereafter referred to as the posterior lobe) develops from the EVL (Glassford et al. 2015) and is a key diagnostic feature of the melanogaster clade (Coyne 1983; Markow and O’Grady 2005). Multiple groups have attempted to map the genomic regions associated with morphological changes in the posterior lobe (Macdonald and Goldstein 1999; Zeng et al. 2000; Masly et al. 2011; McNeil et al. 2011; Tanaka et al. 2015; Takahara and Takahashi 2015). In addition, a previous study identified a gene regulatory network associated with posterior lobe development that also functions in the development of the posterior spiracle, a larval structure involved in gas exchange (Glassford et al. 2015). Multiple transcription factors within the posterior lobe network appeared among our candidates, and we used these genes as positive controls for our methods.

At 28h APF, the tissue that will form the surstylus and the EVL exists as a single continuous epithelium (Figure 1B) that later undergoes cleavage to form both structures by 48h APF (Glassford et al. 2015). Hereafter, we refer to this single structure as the epandrial ventral lobe / surstylus (EVL/S). In accordance with previous results, we found that Pox neuro (Poxn) is expressed in the EVL/S at 28h APF and the EVL at 48h APF (Figure 2A). In addition to Poxn, we found that Abdominal-B (Abd-B) and empty spiracles (ems) are expressed in the EVL/S and EVL, as well as within the posterior lobe domain (Figure 2C-E); both genes were previously identified as posterior lobe network components (Glassford et al. 2015).

In addition to these known factors, we identified many other transcription factors expressed in the EVL and posterior lobe. We found that E5 is expressed in the posterior lobe, the ventral portion of the EVL (see additional samples online), and the phallus. E5 is a homeodomain transcription factor (Dalton et al. 1989) associated with variation in posterior lobe morphology among Drosophila melanogaster populations (Takahashi et al. 2018). We also found that brother of odd with entrails limited (bowl) is expressed in the posterior lobe at 48 hr APF, as well as other tissues throughout the terminalia (Figure 2C-E). bowl is a target of Notch signaling and has been previously implicated in leg development and epithelial re-arrangements in the hindgut (Iwaki et al. 2001; de Celis Ibeas and Bray 2003).

In addition to genes localized within the posterior lobe, we found that escargot (esg) and grainy head (grh) are expressed in the EVL at
both timepoints, but occupy a compartment medial to the posterior lobe – both are expressed near the location where EVL tissue separates from the surstylus (Figure 2C-E). *grh* is a snail-related transcription factor that functions in the development of larval imaginal discs (Whiteley et al. 1992; Hayashi et al. 1993; Fuse et al. 1996), while *grh* is associated with the maternal-zygotic transition during embryonic development, as well as morphogenetic processes in several developmental contexts (Hemphälä et al. 2003; Narasimha et al. 2008; Harrison et al. 2010).

We did not identify a transcription factor that serves as a unique/non-ambiguous marker for the EVL or the posterior lobe – all genes expressed in the EVL were also expressed in at least one other tissue (Figure 2C and D). For example, *Abd-B*, *ems*, *E5* and *eg* accumulate mRNA in the posterior lobe and phallus, but within different phallic substructures (Figure 2C, see below for descriptions of phallic morphology). *grh* and *bowl* are also expressed in other, distinct terminal structures (Figure 2C). Thus, transcription factors expressed in these structures are not unique, but show patterns of co-expression that differ from factor to factor.

**The surstylus (clasper)**

The surstylus (also known as the clasper) is a curled outgrowth located medial to the EVL (Rice et al. 2019b). Like the posterior lobe, the surstylus exhibits morphological differences between species of the *melanogaster* subgroup (Bock and Wheeler 1972), and has been the focus of quantitative trait locus (QTL) mapping efforts (True et al. 1997; Tanaka et al. 2015). A recent study identified *tartan*, a cell adhesion protein, as a gene that contributes to changes in surstylus morphology between *Drosophila simulans* and *Drosophila mauritiana* (Hagen et al. 2019). However, while RNAi experiments in *Drosophila melanogaster* have identified several genes that influence surstylus morphology (Tanaka et al. 2015), little is known about the gene regulatory network that governs its development during pupal stages.

We found that *odd paired (opa)* is expressed exclusively in the surstylus at 48h APF, as well as the medial portion of the EVL/S at 28h APF (Figure 3A and B). These data suggest that *opa* is a surstylus-specific marker, and can also identify presumptive surstylus tissue prior to its cleavage from the EVL. In other tissues, *opa* controls the formation of parasegment boundaries during embryogenesis (Clark and Akam 2010), as well as morphogenetic events in the formation of the midgut and head (Cimbora and Sakonju 1995; Lee et al. 2007).

In addition to *opa*, we found transcription factors expressed in specific subcompartments of the surstylus. *Drop* (*Dr*) is expressed in presumptive surstylus tissue at 28h APF, as well as a more restricted compartment at 48h APF, which may represent the boundary between the surstylus and the EVL (Figure 3, C and E). *Dr* has been previously implicated in genital development and is expressed in larval (L3) genital discs (Chatterjee et al. 2011). We also found that *C15* is expressed in a dorsal-medial compartment of the presumptive surstylus at 28h APF, as well as at the base of the surstylus at 48h APF (Figure 3, D and F). *C15* functions in the development of the amnioserosa during embryogenesis (Rafiqi et al. 2008), as well as during leg development where it interacts with *apterous* and *bowl* (Campbell 2005), both of which exhibit patterned expression in the pupal terminalia (see https://flyterminalia.pitt.edu/). These data show that like the EVL, the surstylus can be delineated into subcompartments by the expression patterns of transcription factors during pupal development.

**The cercus (anal plate)**

The cercus (anal plate) is composed of two flat, semicircular sheets of cuticle on the dorsal side of the terminalia (Rice et al. 2019b).
The cercus is derived from abdominal segment 10 while the rest of the male terminalia originate from abdominal segment 9 (Keisman et al. 2001). This structure shows dramatic variation in bristle number and morphology within and between Drosophilid species (Lachaise et al. 1981; Kopp and True 2002), which in some cases have been implicated in reproductive incompatibility (Tanaka et al. 2018). QTL analysis for differences in the total cercus area between D. mauritiana (Lachaise et al. 1981; Kopp and True 2002), which in some cases have been implicated in reproductive incompatibility (Tanaka et al. 2018). QTL analysis for differences in the total cercus area between D. mauritiana and D. simulans (okada 1954; Kamimura 2006; Vincent et al. 2018), has been previously implicated in the anterior compartment of other abdominal segments (Kopp et al. 1997), including the terminalia (Epper and Sánchez 1983; Chen and Baker 1997; Casares et al. 1997). Finally, several genes are expressed in the developing rectum, including Dr (Figure 3C-E), knirps, and tramtrack (see https://flyterminalia.pitt.edu/).

**The hypandrium**

The hypandrium is a plate-like structure that flanks the phallus on the ventral side (Rice et al. 2019b). The hypandrium contains several substructures, including the hypandrial phragma, medial gonocoxite, pregonites, lateral gonocoxite, and transverse rod (Figure 1C). Within the hypandrium, the lateral gonocoxite and the pregonites exhibit rapid evolution across Drosophilids (Okada 1954; Kamimura and Mitsumoto 2011). While few genes have been previously implicated in hypandrial development, genetic perturbations in Dr cause changes in hypandrial morphology (Chatterjee et al. 2011), and one study localized the loss of hypandrial bristles to a cis-regulatory element of the scute gene (Nagy et al. 2018).

We found that Dichaete (D) is expressed in the hypandrial phragma (i.e., deep into the sample when viewed from the posterior) at both time points (Figure 5A and B). D is a member of the Sox family of transcription factor genes and is critical during embryogenesis (Russell et al. 1996). We also found that several transcription factors are expressed in...
hypandrial substructures. For example, Dr is expressed throughout the medial gonocoxite and weakly in the hypandrial phragma (Figure 5D). In contrast, esg is localized to the base of the pregonites as well as the posterior tip of the lateral gonocoxite (Figure 5E). Taken together, we found discrete gene expression patterns within the pupal domains of annotated hypandrial substructures.

The phallus

The phallus is the male genital organ used for intromissomion and is composed of four substructures: aedeagus, aedaegal sheath, dorsal postgonites, and ventral postgonites (Rice et al. 2019b). Each of these substructures exhibits morphological changes within the melanogaster species group (Okada 1954; Kamimura and Mitsumoto 2011), and QTL mapping has identified genomic regions associated with some of these differences (Peluffo et al. 2015). Here, we confirmed that Poxn is expressed throughout the phallus (Figure 6D), which is consistent with previous observations that Poxn is essential for phalic development (Boll and Noll 2002; Glassford et al. 2015).

The aedeagus is a phallic structure that delivers sperm and exhibits a needle-like shape in D. melanogaster. We identified genes that are expressed along the dorsal-ventral axis of the aedeagus in what appear to be non-overlapping patterns. We found that gooseberry (gsb) was exclusively expressed in the ventral portion of the aedeagus at both 28 and 48hrs APF (Figure 6A and B). gsb was previously found to be expressed in the anterior-ventral edge in L3 genital discs (Freedland and Kuhn 1996), and is a segment polarity gene that interacts with wingless during embryogenesis (Li and Noll 1993). We also found that Polycomb-like (Pcl) was expressed in the same compartment as gsb at 48h APF, but exhibits broader expression at 28h APF (Figure 6D–F). Reciprocally, we found that fd96Cb was expressed in the dorsal portion of the aedeagus. Finally, we identified genes expressed in other aedeagal subcompartments. For example, we found that esg was restricted to the anterior base of the aedeagus, while retained (retn), inv and en are expressed in the opening of the aedeagus, known as the phallotremum.

The aedeagal sheath along with the dorsal and ventral postgonites are two phallic substructures situated lateral to the aedeagus (Figure 6C). The aedeagal sheath consists of two flat, shield-like extensions that bilaterally flank the aedeagus. We found that several genes were expressed in the sheath, including fd96Cb and retn. The dorsal and ventral postgonites are two pairs of spike-like extensions that project from the aedeagal sheath. We found that esg is expressed at the base of both pairs of postgonites, while fd96Cb was expressed throughout both pairs of postgonites. We also found that retn (Figure 6F) and dsx (see images at https://flyterminalia.pitt.edu/) are expressed in the ventral postgonites, but not the dorsal pair, and we note that dsx has a known enhancer that drives expression in this region (Rice et al. 2019a). Taken together, we identified genes that are expressed in distinct phallic structures, as well as within subcompartments of individual structures.

DISCUSSION

In this study, we profiled the transcriptome of the male pupal terminalia in D. melanogaster at critical timepoints when major adult structures form. We then determined the spatiotemporal gene expression patterns of the 100 most highly expressed transcription factors during this stage. We identified transcription factors expressed in five major terminal structures, as well as several substructures that exhibit morphological diversity between species. The entirety of this dataset can be browsed and searched at https://flyterminalia.pitt.edu/. We discuss the implications of our results for the development and evolution of terminalia in Drosophilids.

Drosophila terminalia as a model system

To appreciate the transformative power of a gene expression atlas, we need to look no further than the Drosophila melanogaster embryo. Beginning with the iconic Heidelberg screen (Nüsslein-Volhard and Wieschaus 1980; Wieschaus and Nüsslein-Volhard 1996), which identified genes that control embryonic patterning, many groups have contributed to the development and dissemination of genetic
resources for studies in embryogenesis. These resources include transcriptomic profiling (Lott et al. 2011) and expression atlases of nearly all genes detectable during this stage of development (Tomancak et al. 2007; Lécuyer et al. 2007). Quantitative gene expression atlases are now available at cellular resolution for multiple genetic backgrounds and in different species (Fowlkes et al. 2008, 2011; Pisarev et al. 2009; Staller et al. 2015; Karaikos et al. 2017). These atlases enable computational models of gene regulatory networks and enhancer function that have provided insights into the evolution of patterning networks (Wunderlich et al. 2012; Wotton et al. 2015). However, these resources have revealed that the gene regulatory network which patterns the embryo evolves slowly, producing subtle quantitative changes in gene expression even between distantly related Drosophilids (Fowlkes et al. 2011; Wunderlich et al. 2019). In contrast, the terminalia contain multiple rapidly evolving structures which can illuminate important and under-explored aspects of gene regulatory network evolution.

We envision this atlas of 100 transcription factors as a first step toward building a comprehensive system for the study of developmental network function and evolution. Our RNA-seq data suggest that additional transcription factors are expressed at 28 hr APF, and it is likely that transcriptomic measurements at other time points or with different methods will reveal additional candidates. We will continue to add additional gene expression measurements to the FlyTerminalia database (https://flyterminalia.pitt.edu/) as these candidates are pursued. In particular, our atlas provides a foundation for performing and analyzing single-cell RNA-seq experiments on developing pupal terminalia. While single-cell RNA-seq data provide more highly-resolved information on cell types, they do not contain anatomical information on the spatial organization of those cell types. We therefore anticipate that this atlas will permit the annotation and interpretation of single-cell RNA-seq data. In the future, we hope to expand FlyTerminalia to include expression patterns in the developing female terminalia, which are historically understudied (Hosken and Stockley 2004; Ah-King et al. 2014), as well as expression measurements in other species. By continuing to develop these resources, we hope that Drosophila terminalia will become a premiere model system to address many questions in developmental evolution.

**Implications for genital evolution**

Most of the recent work on the genetic basis of genital evolution has been confined to variation within species and between crossable species (True et al. 1997; Zeng et al. 2000; Masly et al. 2011; Tanaka et al. 2015, 2018; Peluffo et al. 2015). However, even for the most extensively studied genital traits, only a portion of the heritable changes have been resolved to the level of individual genes (Hagen et al. 2018; Nagy et al. 2018). This atlas may thus provide useful candidates for numerous unresolved QTL peaks. In addition, many traits evolve on macroevolutionary time scales, precluding the possibility of QTL analysis. Previous work used a comparative analysis of gene expression to identify a network of genes that was co-opted to the posterior lobe – a novel trait restricted to the melanogaster clade (Glassford et al. 2015). However, the D. melanogaster clade contains other unique traits, including

---

**Figure 6** Transcription factors expressed in the phallus. A) Left: schematic of major terminal structures at 28 hr APF with the phallus indicated in orange. Right: Light microscope image of in situ hybridization data gooseberry gsb mRNA at 28 hr APF. B) Left: schematic of major terminal structures at 48 hr APF with the phallus indicated in orange. Right: Light microscope image of in situ hybridization data for (gsb) mRNA at 48 hr APF. C) Cartoon representation of the substructures of the phallus: ventral postgonite (orange), aedeagus (yellow), phalloerema (brown), dorsal postgonites (pink), and aedeagal sheath (red). Additional in situ hybridization data for transcription factors Poxn, esg, fd96Cb, retn, Pcl, and en at 28 hr APF (D) and 48 hr APF (E). F) Top: High magnification images of the samples shown in (E) to illustrate details of phallus expression patterns. Bottom: Cartoon representation of the substructures of the phallus, with shading indicating expression within each substructure. Note that for en, light shading indicates weak expression throughout the phallus.
structures whose gene regulatory networks have not been previously characterized. In this study, we found several genes that are expressed in lateral gonocoxite (esg, inv, cn), and postgonites (esg, fd96Ch, crp, mod, retn and dsx), both of which exhibit morphological changes between species. Furthermore, a ventral postgonite enhancer was recently identified for the gene doublesex (Rice et al. 2019a) which may serve as a useful driver to manipulate this structure in future studies. Other enhancers that drive expression in the larval genital disc may persist in the pupal terminalia and prove useful as drivers to target other structures (Jory et al. 2012). To assess the functional roles of individual genital structures in copulation, genetic disruption may help complement other techniques such as laser ablation (Polak and Rashed 2010; Kamimura and Polak 2011; LeVasseur-Viens et al. 2015).

Rapid morphological changes between species hamper the identification of homology relationships among terminal structures. Structural homology has previously been defined by similarities in adult morphology, but structures that appear similar may nevertheless not be related by common descent. As a result, there are conflicting claims of homology—the same structure in one species has been called homologous to different structures in other species (McAlpine, et al. 1981; Grimaldi 1987; Grimaldi 1990). Based on our results, we suggest that gene expression profiles may be useful in reconciling contradicting claims of homology. For example, homology is difficult to establish for the postgonites, often referred to as parameres or branches (Kamimura 2007; Yassin and Orgogozo 2013; Peluffo et al. 2015). Here, we identified genes expressed in both pairs of postgonites (fd96Ch, and esg), which may help to define homologous structures in other species.

**Implications for genital development**

In mapping the transcription factor landscape in the pupal terminalia, we have begun defining the gene regulatory networks that operate during the development of these structures. Identifying relevant transcription factors and measuring their gene expression patterns is an important first step, but we must also determine how these genes interact. At this point, we can infer regulatory interactions by documenting incidences of co-expression or reciprocal expression. For example, it would be interesting to test whether transcription factors expressed in the entirety of particular structures, such as the surstylus marker odd paired, are required for expression of other genes deployed in more restricted subcompartments, such as C15. Some of these genes have known regulatory interactions in other contexts, such as apterous, C15, and bowl (Campbell 2005). While this atlas can be a tool for generating hypotheses regarding how these gene regulatory networks are wired, these hypotheses must ultimately be tested via genetic perturbation.

Locating the regulatory DNA that controls these expression patterns will also be critical for defining relevant gene regulatory networks. One notable feature of our results is that most of the identified transcription factors are expressed in multiple locations throughout the pupal terminalia, especially at 48h APF. It remains unclear whether these patterns are controlled by multiple regulatory elements, or if disparate patterns are generated by the same enhancer region (Small et al. 1996). It is possible that the enhancers controlling these patterns also operate in other tissues or at different developmental stages (Preger-Ben Noon et al. 2018; Sabaris et al. 2019), as is the case for the posterior lobe enhancer of Pox neuro (Glassford et al. 2015) and the hypantrial enhancer of scute (Nagy et al. 2018). By finding the regulatory sequences that control these gene expression patterns, we can determine the direct targets of transcription factors in this system.

Epithelial remodeling is a critical component of many developmental events, including gastrulation, neural tube formation, and organogenesis (Neumann and Affolter 2006). Studying these processes in Drosophila tissues, such as the wing disc and the trachea, has yielded insights into similar processes in mammals (Affolter et al. 2003). We focus here on patterned transcription factors because morphogenetic processes are tightly regulated at the level of gene expression. However, we are ultimately interested in the connections between transcription factors and the effectors that dictate cell behavior (Smith et al. 2018). Recent work has implicated a variety of cellular mechanisms in the formation of genital structures, including changes in cell size and cell intercalations in the developing ovipositor (Green et al. 2019) and the influence of the apical extracellular matrix in the developing posterior lobe (Smith et al. 2019). In the future, we hope to characterize the functional roles of transcription factors in both cellular dynamics and adult morphology, and elucidate how the expression and function of these genes are tuned to generate new or different structures over evolutionary time.

**ACKNOWLEDGMENTS**

We thank the entire Rebeiz lab for assistance optimizing the in situ hybridization protocol and helpful discussions, especially Eden McQueen and Donya Shodja for helpful comments on the manuscript. Artoo Situ deserves special recognition for autonomous assistance with in situ hybridization. We also thank the University of Pittsburgh Center for Research Computing, especially Kim Wong for technical help on building and hosting our online database. Finally, we thank our editor and two anonymous reviewers, whose feedback substantially improved the clarity of the paper. This project was supported by the Stowers Institute for Medical Research and the National Institutes of Health (1F32GM130034 to BJV, 1F32GM125329 to WJG, and GM107387 to MR).

**LITERATURE CITED**

Affolter, M., S. Belluscio, N. Itoh, B. Shilo, J.-P. Thiery et al., 2003 Tube or not tube: remodeling epithelial tissues by branching morphogenesis. Dev. Cell 4: 11–18. https://doi.org/10.1016/S1534-5807(02)00410-0

Ah-King, M., A. B. Barron, and M. E. Herberstein, 2014. Genital evolution: why are females still understudied? PLoS Biol. 12: e1001851. https://doi.org/10.1371/journal.pbio.1001851

Baker, B. S., and K. A. Ridge, 1980 Sex and the single cell. I. On the action of major loci affecting sex determination in Drosophila melanogaster. Genetics 94: 383–423.

Bock, I. R., and M. R. Wheeler, 1972 The Drosophila melanogaster species group. University of Texas Publication 7213: 1–102.

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

Brennan, P. L. R., and R. O. Prum, 2015 Mechanisms and Evidence of Genital Coevolution: The Roles of Natural Selection, Mate Choice, and Sexual Conflict. Cold Spring Harb. Perspect. Biol. 7: a017749. https://doi.org/10.1101/cshperspect.a017749

Campbell, G., 2005 Regulation of gene expression in the distal region of the Drosophila leg by the Hox11 homolog, C15. Dev. Biol. 278: 607–618. https://doi.org/10.1016/j.ydbio.2004.12.009

Casares, F., L. Sánchez, I. Guerrero, and E. Sánchez-Herrero, 1997 The genital disc of Drosophila melanogaster. I. Segmental and compartmental organization. Dev. Genes Evol. 207: 216–228. https://doi.org/10.1007/s004270050110

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

Chatterjee, S. S., L. D. Uppendahl, M. A. Chowdhury, P.-L. Ip, and M. L. Siegal, 2011 The female-specific doublesex isoform regulates pleiotropic transcription factors to pattern genital development in...
