CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS
DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD IRAPUATO

UNIDAD DE GENÓMICA AVANZADA

“Comparación de la co-expresión génica durante la embriogénesis de dos especies de protostomados”

Tesis que presenta

René Alexander Ramos Díaz

Para obtener el grado de

Maestro en Ciencias

Con la especialidad de

Biología Integrativa

Directora de tesis

Selene Lizbeth Fernández Valverde

Irapuato, Guanajuato Octubre, 2019

CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS
DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD IRAPUATO

UNIDAD DE GENÓMICA AVANZADA

“Comparing gene co-expression during the embryogenesis of two protostome species”

THESIS

Presents

René Alexander Ramos Díaz

To obtain the degree of

Master of Sciences in Integrative Biology

Thesis director

Selene Lizbeth Fernández Valverde

Irapuato, Guanajuato October, 2019

El presente trabajo de investigación fue realizado bajo la dirección de la Dra. Selene Lizbeth Fernández Valverde en el Laboratorio de Genómica Funcional y Evolutiva de ARNs regulatorios en la Unidad de Genómica Avanzada del Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional CINVESTAV, Unidad Irapuato, Guanajuato. El proyecto de maestría se realizó en el período de agosto del 2017 a octubre de 2019.

Se agradece al Consejo Nacional de Ciencia y Tecnología (CONACYT) por la beca de maestría número 636143 otorgada durante el período de realización del presente trabajo de investigación.

AGRADECIMIENTOS

Agradezco a cada una de las personas que han contribuido a mi formación académica y crecimiento personal durante la maestría. Quisiera primero agradecer a la Dra. Selene Lizbeth Fernández Valverde por haberme dado la oportunidad de trabajar en el laboratorio de Genómica Funcional y Evolutiva de ARNs regulatorios. Gracias a su apoyo, recomendaciones, consejos y observaciones fue posible concluir exitosamente este proyecto.

A la Dra. Kasia Oktaba y al Dr. Eugenio Mancera por haber formado parte del cómite evaluador, que con sus observaciones, sugerencias y críticas hicieron posible que este trabajo de investigación fuera mejorando desde el inicio.

A los integrantes del laboratorio de biología computacional: Aracely, Josian, Irving, Paco, Emiliano, Jaime, Fani, América, Beto, Víctor, Falcón, Jordan, Atenea, Evelia, Gilberto, Pablo, Obed, Lorena, Esteban y al Dr. Cei Abreu. Gracias por todo el apoyo y por compartir su tiempo mientras estuve en el laboratorio. También quiero agradecer a Mayra Flores Barraza por su paciencia y ayuda con los procesos administrativos.

A los amigos que hice durante el posgrado en Biología Integrativa: Cori, Judith, Luisa, Michelle, Freddy, Josúe, Karina, Raúl, Paco, Astrid, Juan Esteban, Rosy, Paul, César y Javier. Gracias por ayudarme a aprender biología y compartir conmigo durante este tiempo.

A mis amigos y todas las personas que me han apoyado durante estos años. Especialmente a Eugenia, Ana María, Lilian, Ovidio, Rita, Rodrigo, Gabriela, Juan Fernando, José, Omar, Laura, Eli, Linmar y a la Dra. Karen Salomé Caballero Mora por haberme alentado a seguir trabajando en investigación.

Finalmente quiero dar las gracias a mi madre Bertila y a mi hermano Manuel quienes nunca han dejado de apoyarme incondicionalmente en cada proyecto que he emprendido a pesar de todas las dificultades que han habido a lo largo del camino.

Contents

List of Figures

List of Tables

1 Abstract

Embryogenesis or embryonic development encompasses the stages of an animal's life cycle between fertilization and hatching of larva or birth. Previous comparative transcriptomic studies of embryogenesis in species of the same phylum have found that gene co-expresion is more conserved during a stage called phylotypic stage. Orthologs co-expressed during this stage are related to conserved developmental processes within the phylum. On the other hand, ortholog expression has been found to be most divergent between different phyla during the phylotypic stage, suggesting this divergent expression is fundamental to the establishment of divergent body plans. In spite of these general observations, it is not well understood if some orthologous families (or orthogroups) conserve their co-expression between animals of different phyla. In order to identify orthogroups that are co-expressed between species of different phyla, we analyzed transcriptomic data of whole embryos encompassing all embryogenesis in two prostostome species: *Drosophila melanogaster* and *Caenorhabditis elegans*. We found *1,009* orthogroups are co-expressed in both species, the majority of which are expressed at different developmental times between these species. From these, seven groups of homeobox, zinc-finger, paired-box (Pax), sine oculis homeobox (SIX), and LIM-homeobox transcription factors preserve their coexpression in spite of shifts in their time of expression between these species. Our findings suggest the co-regulation of these genes may predate the divergence of these protostome species.

Resumen

La embriogénesis o desarrollo embrionario abarca las etapas del ciclo de vida de un animal entre la fertilización y la eclosión de la larva o nacimiento. Estudios transcriptómicos comparativos previos de embriogénesis en especies del mismo filo han encontrado que la coexpresión de genes está más conservada durante una etapa denominada etapa filotípica. Los ortólogos coexpresados durante esta etapa están relacionados con procesos de desarrollo conservados dentro del filo. Por otro lado, se ha encontrado que la expresión de ortológos es más divergente entre diferentes filos durante la etapa filotípica, lo que sugiere que esta expresión divergente es fundamental para el establecimiento de planes corporales divergentes. A pesar de estas observaciones generales, no se entiende bien si algunas familias de genes ortólogos (u ortogrupos) conservan su coexpresión entre animales de diferentes filos. Para identificar los ortogrupos que se coexpresan entre especies de diferentes phyla, analizamos datos transcriptómicos de embriones completos que abarcan toda la embriogénesis en dos especies de protostomados: *Drosophila melanogaster* y *Caenorhabditis elegans*. Encontramos que *1,009* ortogrupos se coexpresan en ambas especies, la mayoría de los cuales se expresan en diferentes momentos del desarrollo entre estas especies. De estos, siete grupos de factores de transcripción homeobox, zinc-finger, paired-box (Pax), sine oculis homeobox (SIX) y LIM-homeobox preservan su coexpresión a pesar de los cambios en su tiempo de expresión entre estas especies. Nuestros resultados sugieren que la coregulación de estos genes puede ser anterior a la divergencia de estas especies de protostomados.

2 Introduction

2.1 General aspects of animal embryogenesis

Embryogenesis is defined as the stages of an animal's life cycle between fertilization and hatching or birth [Gilbert 2011, chapter 1 p. 4-5]. These stages can be defined according to which major molecular and phenotypical transitions are observed in the embryo, this process of dividing embryogenesis into stages is known as "periodization" [Hall 1999, chapter 8 p. 127]. An important remark related to periodization is that: "All subdivisions of embryogenesis into stages are necessarily artificial, in the sense that they imply a beginning and an end of the embryogenetic process at each stage. In spite of the artificiality that it imposes, however, staging is very useful for describing embryonic development, since it provides a temporal framework to which embryogenetic events can be referred" [Campos-Ortega and Hartenstein 1997 chapter 1 p. 1]. In general animal embryogenesis can be divided into the following stages [Gilbert 2011 chapter 1, 2, 4, 7, 8, 9, 21 p. 4-15, 38-41, 95-142, 217-250, 251-275, 277-309, 672-692; Hall 1999, chapter 8 p. 129-134]:

1. Fertilization. Fusion of the egg (oocyte) with sperm, which are haploid cells. Each of these gametes provides half of the chromosomes that will make the diploid genome of the embryo. The single cell formed after fertilization is called zygote. Egg activation can occur before (insects) or after fertilization (nematodes, echinoderms, and vertebrates) and usually requires the release of calcium ions across species [Horner and Wolfner 2008; Runft et al. 2002]. The IP3 pathway regulates the release of ions stored in the endoplasmic reticulum to promote the initiation of maternal regulation in many species including Drosophila melanogaster [Gilbert 2011 chapter 7 p. 234-239; Kaneuchi et al. 2015]. In the nematode Caenorhabditis elegans this initial calcium release is regulated by the calcium channel TRP-3 [Takayama and Onami 2016].

2. Cleavage. 2. Cleavage. This stage encompasses the rapid mitotic divisions that follow after fertilization. However, a zygote could remain without dividing for a small amount of time as discussed later. Once cleavage ends the zygote has divided into smaller cells called blastomeres forming a structure called the blastula. Cleavage patterns are different across species and depend on yolk protein content and mitotic spindle formation, and can be classified in two general types: (a) holoblastic which is common in embryos with little yolk that acquire nutrients from other sources such as mammals that obtain nutrients from the placenta, it is characterized by a uniform blastula; (2) meroblastic cleavage occurs in embryos in which most of the cell volume is yolk used as the only source of nutrients throughout embryonic development, it is observed for example in insects and birds, only a portion of the cytoplasm is cleaved [Gilbert 2011 chapter 1, p. 11-13]. A syncytium is a cell containing many nuclei. In *Drosophila melanogaster* and other insects a syncytium is formed and hence nuclear divisions are occurring instead of cytoplasmic cleavage, the embryo forms a structure called syncytial blastoderm which is observed between after 13

nuclear cycles [Campos-Ortega and Hartenstein 1997 chapter 1 p. 1-2; Gilbert 2011 chapter 2, p. 38; Hales et al. 2015].

3. Gastrulation. Rearrangement of the blastula by cell movements that organize the embryo in three germ layers: endoderm (inner), mesoderm (middle) and ectoderm (outer) from which tissues will form later. Endoderm for example generates the digestive tuve and pharynx. Muscles and blood cells arise from mesoderm. Central nervous system and skin are specified from the ectoderm. Germ cells are the only cell type which is separated during development and does not arise from any germ layer [Gilbert 2011 chapter 1, p. 15]. There are five basic cell movements that animals use during gastrulation: invagination, involution, ingression, delamination, and epiboly. Patterns of gastrulation are different across species but usually include combinations of these basic cell movements [Gilbert 2011 chapter 1, p. 13-14; Wolpert 1992]. Triploblastic animals or bilaterians (species with bilateral symmetry along the anterior-posterior axis), for example vertebrates, nematodes and arthropods have these three germ layers. Diploblastic animals like cnidarians and ctenophores have only endoderm and ectoderm [Gilbert 2011 chapter 8, p. 252-253].

4. Organogenesis. Regions of the embryo that were previously not associated, become closer and allow the formation of organ systems by exchanging molecular signals between these new sites generated from germ layers. Additionally cells can migrate from an initial location to a final site. Migrating cells include precursors of blood cells, lymph cells and gametes [Gilbert 2018 chapter 1, p. 4]. Cell adhesion molecules are important for mantaining tissue organization, and the most important are cadherins or "calcium-dependent adhesion molecules". The extracellular matrix is a network of structural proteins including collagen, proteoglycans, and glycoproteins present in all animal tissues. Cell-to-cell communication can be divided into two general categories: (a) by direct contact or juxtacrine signaling and (b) paracrine signaling in which cells communicate by secreting proteins into the extracellular matrix. Signaling proteins that induce a response in other cells are called ligands and membrane proteins that bind other membrane-associated proteins (juxtacrine signaling) or ligands (paracrine signaling) are called receptors. [Gilbert 2018 chapter 4, p. 100-104]. Morphogens are "diffusable biochemical molecules that can determine the fate of a cell by its concentration. That is, cells exposed to high levels of a morphogen activate different genes than those cells exposed to lower levels" and can be transcription factors (more common in syncytium embryos as in *Drosophila melanogaster*) or paracrine factors [Gilbert 2018 chapter 4 p. 115]. Most paracrine factors required for animal organogenesis can be classified according to their structure into one of four families : (a) Fibroblast growth factor FGF, (b) Hedgehog, (c) Wnt, and (d) TGF-β superfamily which includes the families: TGF-β, activin, bone morphogenetic proteins or BMPs, Nodal proteins, Vg1, and other related proteins [Gilbert 2018 chapter 4 p. 116-142]. These signaling pathways are present across animals, but their individual components can vary between species [Babonis and Martindale 2016].

5. Larval or juvenile stages. After hatching or born, many organisms are not sexually mature and their morphology is different compared to adult individuals. This stage of the life cycle is called larva. Species that go through larval stages are collectively called indirect developers and require to undergo metamorphosis to become a sexually mature adult. Larval stages are common in insects (*Drosophila melanogaster*), nematodes (*Caenorhabditis elegans*), amphibians (*Xenopus tropicalis*), and many marine organisms such as sea urchins (*Strongylocentrotus purpuratus*). Some species do not go through larval stages, instead the juvenile organism hatches or borns as a miniature adult. This type of organisms are called direct developers, examples include mammals (e.g *Mus musculus*) and birds (e.g. *Gallus gallus*) [Gilbert 2018 chapters 1, 21 p. 4, 672-692].

Most stages occur within specific time intervals that can vary widely across species [Hall 1999, chapter 23 p. 365-373]. Additionally, other metazoan embryonic stages have been proposed and are related to more specific developmental transitions or events, the most important are [Hall 1999, chapters 7, 8, 14 p. 115-117, 127, 227-231; Slack et al. 1993]:

1. Maternal-to-zygotic transition (abbreviated as MZT) in which maternal regulation ends, zygote genome activation occurs, and zygotic transcription begins [Baroux et al. 2008; Langley et al. 2014; Lee et al. 2014; Palfy et al. 2017; Tadros and Lipshitz 2009]. Early zygotic genes have been studied in model organisms such as *Drosophila melanogaster* [De-Renzis et al. 2007] and *Danio rerio* [Jukam et al. 2017]. Even though most of these genes are not conserved across phyla [Heyn et al. 2014], similar mechanisms regulate the transition. For example in *Danio rerio* transcription factors *Nanog*, *Pou5f3*, and *Sox19b* are required to initiate zygotic gene expression, they activate the microRNA *miR-430* which participates in clearance of maternal RNAs, however other factors are thought to participate in this process. In *Drosophila melanogaster* the transcription factor *Zelda* (originally named *Vielfältig*) is a zinc finger transcription factor that activates early zygotic transcripts including the microRNA *miR-309* with an equivalente role in clearance of maternal RNAs along with the RNA-binding protein *Smaug* and probably other factors [Lee et al. 2014].

2. The zootype, a stage of expression of Hox cluster genes was originally proposed as characteristic of metazoa [Slack et al. 1993]. The generalization of this stage as characteristic of animals has been criticized because diploblastic animals such as cnidarians (*Bellonella rigida* and *Hydra vulgaris*) have different expression patterns of orthologues of the Hox cluster compared to bilaterians (triploblastic species with bilateral symmetry across the anterior-posterior axis), which suggest that they do not have a role in regulating segment identity across the oral-aboral axis (similar to the anterior-posterior axis), and hence this function is an innovation of bilaterians [Martínez et al. 1998; Schierwater and Desalle 2001]. Some authors propose to discard the concept of zootype altogether based on genomic data of Hox genes across phyla and comparative morphological studies of bilaterian and non-bilaterian

animals (including species of Porifera, Cnidaria, and Placozoa) [Ryan and Baxevanis 2007]. In summary the homology between oral-aboral axis of cnidarians and anterior-posterior axis of bilaterians is not supported by morphological, genomic, and gene expression data [Martindale 2005; Ryan and Baxevanis 2007], and it is still unclear how bilaterial symmetry has evolved in animals [Genikhovich and Technau 2017]. Hence similar expression patterns of Hox cluster genes regulating regional specification across the anterior-posterior axis are observed in bilaterians, but the available evidence indicates that this is not sufficient to define a developmental stage across animals.

3. Phylotypic stage was defined as "the stage at which all members of the phylum show the maximum degree of similarity (tailbud or pharyngula stage in vertebrates, fully-segmented stage in insects or nematode after the completion of most embryonic cell divisions)" [Slack et al. 1993]. While the original definition of this stage is based on morphology of embryos [Irie and Kuratani 2014; Kalinka and Tomancak 2012], comparative transcriptomic analyses in vertebrates [Comte et al. 2010; Domazet-Loso and Tautz 2010; Irie and Kuratani 2011], flies [Kalinka et al. 2010], and nematodes [Levin et al. 2012] support a scenario in which this stage is more conserved than earlier or later stages. Several sources of evidence such as comparative transcriptomics of developmental stages [Piasecka et al. 2013], constraints in stages depending on knock-out experiments [Roux and Robinson-Rechavi 2008], and comparative embryology [Richardson et al. 1997] provide conflicting results for a more precise definition of the phylotypic stage. It is unclear which events initiate or which are the mechanisms that regulate the phylotypic stage [Hall 1999 chapter 14 p. 230-231; Willmore 2012; Yanai 2018].

2.2 Embryogenesis of *Drosophila melanogaster*

Embryogenesis is completed approximately in 22 to 24 hours after fertilization (or 1,320-1,440 minutes post-fertilization) and it is divided into 17 stages following the nomenclature proposed by Campos-Ortega and Hartenstein [Campos-Ortega and Hartenstein 1997 chapters 1,2 p. 1-100; Hales et al. 2015]. As mentioned before the early embryo in this species is a syncytial cell containing many nuclei. Here we only summarize the most important aspects of these stages, their timing and important events associated during embryogenesis. Stage 1 lasts for about 25 min and begins once the egg has been laid after fertilization, and ends after the first two nuclear cycles have been completed [Campos-Ortega and Hartenstein 1997 chapter 2 p. 9-19]. Early nuclear divisions are very fast, and last in average about 8 minutes each [Gilbert and Barresi 2018 chapter 9 p. 279-280]. Stages 2 and 3 encompass nuclear cycles 3 to 9 [Campos-Ortega and Hartenstein 1997 chapter 2 p. 19-22; Gilbert and Barresi 2018 chapter 2 p. 38-39]. During stage 4 when cycle 10 is completed, the nuclei start migrating to the periphery of the egg forming a syncytial blastoderm [Campos-Ortega and Hartenstein 1997 chapter 2 p. 22-24; Gilbert and Barresi 2018 chapter 9 p. 280]. At the end of nuclear cycles 13 and 14 during stage 5 cellularization occurs when nuclei organize in single cells forming a layer of cells around a core of yolk which is called cellular blastoderm [Campos-Ortega and Hartenstein 1997 chapter 2 p. 25-30; Gilbert and Barresi 2018 chapter 2,9 p. 38-39, 280-281].The maternal to zygotic transition is a gradual process in all species, specifically in *Drosophila melanogaster*, clearance of maternal transcripts starts approximately 48 minutes post-fertilization (nuclear cycle 6) and zygote genome activation occurs approximately 2.5 hours after fertilization (150 minutes post-fertilization, nuclear cycle 14) [Lee et al. 2014; Palfy et al. 2017; Tadros and Lipshitz 2009].

Gastrulation encompasses stages 6 and 7 and lasts approximately 20 minutes [Campos-Ortega and Hartenstein 1997 chapter 2 p. 30-36]. The layered invaginated ventral area consisting of ectoderm and mesoderm which forms during gastrulation is called germ band [Brody 2019 The Interactive Fly]. Germ band elongation occurs during stages 8 to 9, at the end of germ band elongation the cephalic furrow is formed. At the end of stage 10 neuroblasts (neural progenitors) and the procephalic region are distinguishable [Campos-Ortega and Hartenstein 1997 chapter 2 p. 36-50]. During stage 11 embryo segmentation is observed [Campos-Ortega and Hartenstein 1997 chapter 2 p. 50- 65]. Germ band retraction occurs during stages 12 and 13 and central and peripherical nervous system differentiate [Campos-Ortega and Hartenstein 1997 chapter 2, 9, 10, 11 p. 65-77, 175-208, 209-233, 233-266]. After stages 14 and 15, dorsal closure and head involution respectively, morphogenesis is basically complete: muscles, epidermal tissue, and sensory organs become distinguishable. During stage 16 the synthesis and secretion of the cuticle that will protect the larva starts.

At the end of stage 17 the embryo is morphologically similar to the larva, tracheal tree fills with air and retraction of the ventral nerve cord is observed [Campos-Ortega and Hartenstein 1997 chapter 2,8 p. 78-98, 166-169]. figure 1 shows a diagram summarizing the embryonic stages of discussed above, was adapted from [Hales et al.2015] with additional data for stages and their estimated time intervals from [Brody 2019 The Interactive Fly] and [Campos-Ortega and Hartenstein 1997 chapters 1,2 p. 1-100].

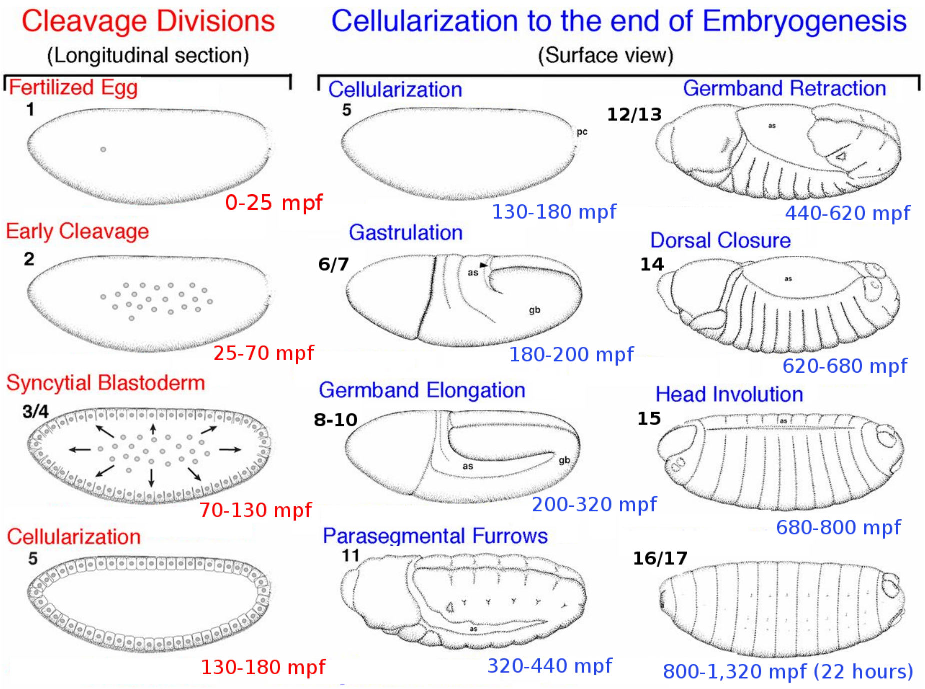


Figure 1. Embryonic stages of *Drosophila melanogaster*. Stage numbers are indicated in the upper left of each embryo along with the major event or characteristic of each stage. Time intervals in minutes post-fertilization (mpf) are shown in the lower right of each embryo. Adapted from [Hales et al.2015], stage data and intervals in which they are observed retrieved from [Brody 2019 The Interactive Fly; Campos-Ortega and Hartenstein 1997 chapters 1,2 p. 1-100].

As in other indirect developers most structures such as organs and limbs are not fully developed in the larva. There are three larval stages called first (lasts 1 day), second (1 day) and third instar (2 days). Larval tissues are organized in imaginal discs that will become specific adult structures later. Temperature can influence the duration of embryonic and larval stages. Adults eclose after metamorphosis of pupae between 4 to 5 days after the third larval stage [Hales et al. 2015] and live approximately 80 days or less depending on environmental conditions [Brummel et al. 2004; Linford et al. 2013].

2.3 Embryogenesis of *Caenorhabditis elegans*

Cell specification in this species occurs very fast [Gilbert and Barresi 2018 chapter 8 p. 265-267; Herman 2006 WormBook; Rose and Gonczy 2014 WormBook], and its complete cell lineage tree is known and available online at WormAtlas: `https://www.wormatlas.org/celllineages.html` [Altun and Hall 2019 WormAtlas]. figure 2 shows the embryonic stages of this species as described above. A summarized cell lineage tree representing the origins of tissues is shown in figure 3. Highly stereotyped cell divisions are observed approximately 50 minutes after fertilization, and last until 150 minutes-post fertilization (mpf), this process takes place in the utero. Maternal to zygotic transition occurs between 70 to 90 mpf [Lee et al. 2014; Tadros and Lipshitz 2009] during these early cleavages. Gastrulation occurs after the embryo is laid between 150 to 330 mpf starting with 26 cells.

Organogenesis-morphogenesis stages begins after gastrulation and ends approximately between 12 to 14 hours after fertilization (approximately 720-800 mpf), and it is subdivided into other stages. The first is called "lima bean" or bean stage which starts around 360 mpf and ends around 400 mpf, spontaneous muscle activity is observed but connections in the nervous system are still incomplete. Elongation or elongation phase follows inmediately and encompasses the comma stage, 1.5-fold stage, and 3-fold stage which ends around 640 mpf. During this phase the embryo changes its morphology to a more elongated form. At the end of 3-fold stage coordinated movement is observed along the anterior-posterior axis which indicates that motor neurons and their connections are developed. After this stage quickening starts, larval cuticle synthesis is observed around 690 mpf, pharyngeal pumping is observed at 760 mpf and finally hatching occurs around 800 mpf. Time intervals for each stage vary depending on temperature [Altun and Hall 2019 WormAtlas; Gilbert and Barresi 2018 chapter 8 p. 265-273; ; Hall et al. 2017 WormAtlas].

An important aspect of *Caenorhabditis elegans* is that most individuals are herma -phrodite that can self-fertilize and only a small number are male. There are four larval stages L1, L2, L3 (8 hours each) and L4 (10 hours), at the end onf each stage cuticle is replaced with a new stage-specific cuticle (moulting). Additionally, at the end of the L2 stage, a larva can enter an arrested state called dauer larva induced by unfavorable conditions such as high temperature or lack of nutrients. Feeding is arrested and motion is restricted, this type of larva can survive up 4 weeks to until favorable conditions are detected and enters L4 larval stage for 4 hours and then enters the normal adult stage. Adults live for approximately 10-15 days [Altun and Hall 2019 WormAtlas].

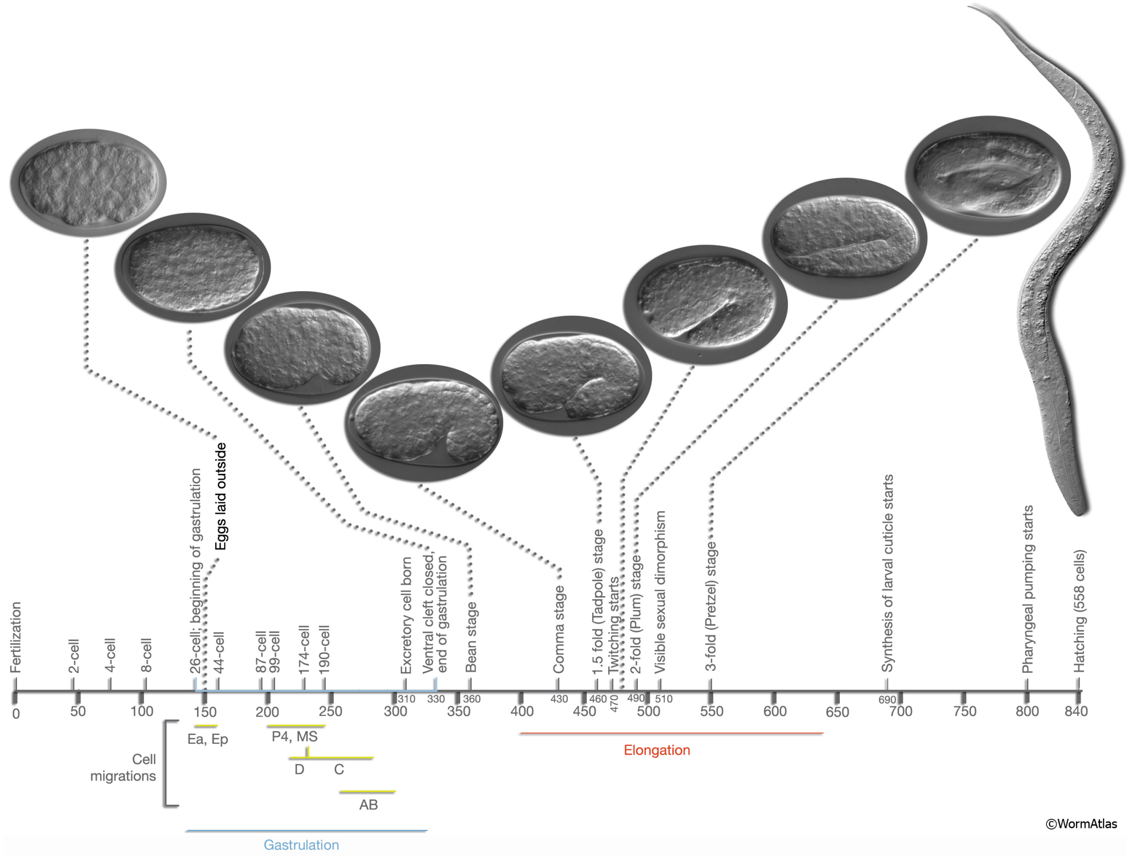


Figure 2. Time is in minutes post-fertilization (mpf). Blue bar represents gastrulation (150-330 mpf), the first cells that move inwards from the ventral surface are gut precursors (E), followed by mesoderm (MS), germline precursors (P4), and muscle precursos D and C, while cells of the AB lineage (neurons, pharynx and other ectodermal tissues) are organized in the outer region. Bean stage occurs between 360 to 400 mpf. Elongation encompasses three stages: comma stage, 1.5-fold stage, and 3-fold stage and ends around 640 mpf it is indicated as a red bar. After elongation, quickening stage starts and lasts until hatching. Larval cuticle synthesis is observed around 690 mpf, first pharyngeal pumping is observed around 760 mpf, and hatching usually occurs between 880 to 840 mpf depending on the temperature (common growth temperature range is from 20 to 25 Celsius). Adapted from [Altun and Hall 2019 WormAtlas] available at `https://www.wormatlas.org/embryo/introduction/EIntroframeset.html`

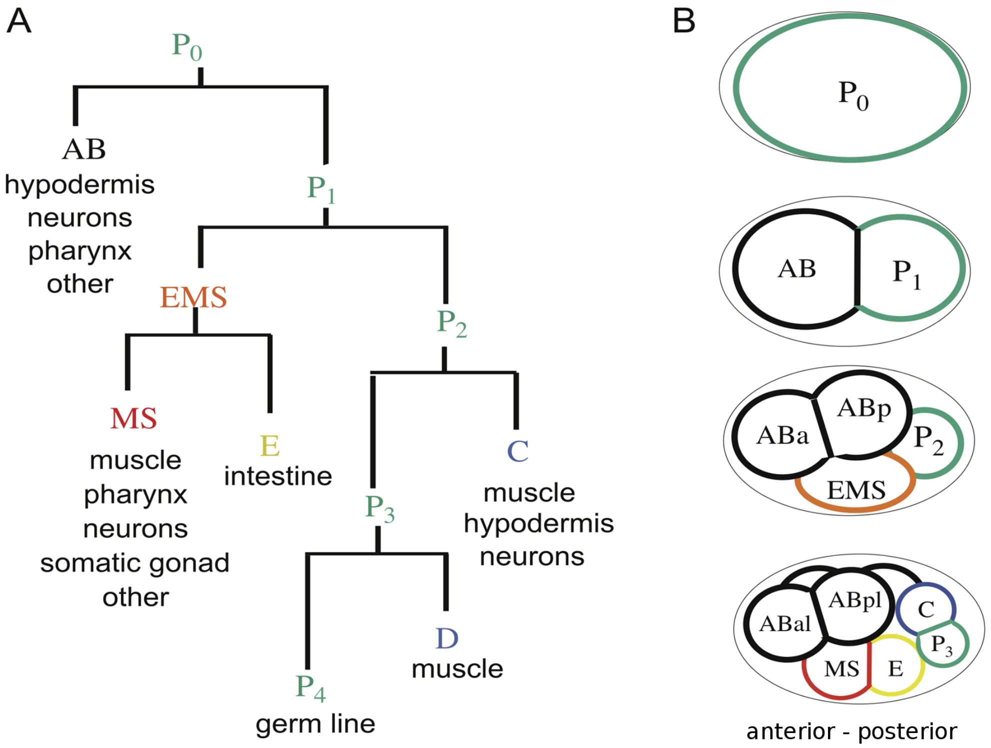


Figure 3. Summary of Cell lineages of *Caenorhabditis elegans*. (A) Simplified cell lineage tree highlighting the six founder cells AB, MS, E, P4, D, C and the major tissues that they generate. (B) Diagram representing the relative positions of cells in the embryo during the first three cell divisions. The embryo is shown in ventral view, anterior to the left, the eggshell is represented by the ellipse surrounding the embryo. Adapted from [Rose and Gonczy 2014 WormBook], available online at `http://www.wormbook.org/chapters/www_asymcelldiv.2/asymcelldiv.2.html`

2.4 Phylogeny of protostomes

Bilaterians are divided into two groups of species depending if the mouth or the anus forms first at or near the opening of the gut (blastopore) during gastrulation, the first group are called protostomes (Greek "mouth first") and the second group is known as deuterostomes ("mouth second") [Gilbert 2018 chapter 8, p. 252-253]. Protostomes encompass two subgroups: Lophotrochozoans and Ecdysozoans. Lophotrochozoans also known as Spiralia are characterized by spiral cleavage and a distinctive planktonic (free-swimming) larva called trochophore (Greek trochos "wheel") in marine or freshwater species, and includes species such as snails and flatworms. Ecdysozoans (Greek ecdysis “to get out of” or “shed”) are invertebrates that moult their exoskeletons [Aguinaldo et al. 1997; Gilbert 2018 chapter 8, p. 253; Giribet and Edgecombe 2017; Telford et al. 2015].

Representative members include Arthropoda and Nematoda, which include species such as *Drosophila melanogaster* and *Caenorhabditis elegans* respectively. Nematoda are part of a larger group called Cycloneuralia which includes Nematomorpha, Kinorhyncha, Loricifera, and Priapulida. Arthropoda is part of Panarthropoda which also includes Tardigrada and Onychophora [Giribet and Edgecombe 2017; Schumann et al. 2018]. Molecular and fossil evidence suggests that most ecdysozoan lineages probably appeared during the Ediacaran approximately 587–543 million years ago (Mya), and major radiations occurred during the Cambrian (539–511 Mya) and Ordovician (510–471 Mya) [Erwin 2015; Giribet and Edgecombe 2017; Rota-Stabelli et al. 2013; Wang et al. 2019]. A phylogeny of protostomes is shown in figure 4.

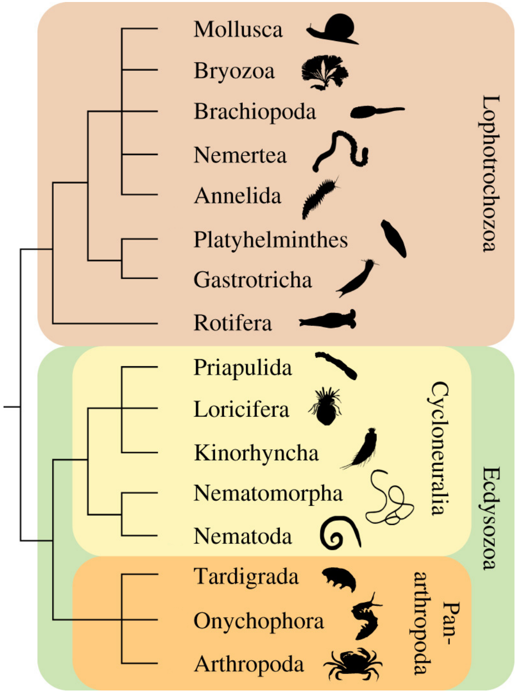


Figure 4. Phylogeny of protostomes. Lophotrochozoans also known as Spiralia are characterized by spiral cleavage. Ecdysozoans or moulting animals encompasses two subclades: Cycloneuralia and Panarthropoda. Nematoda includes the round worm *Caenorhabditis elegans* and Arthropoda includes the fruit fly *Drosophila melanogaster*. Adapted from [Schumann et al. 2018].

2.5 Comparative transcriptomics of embryogenesis within a genus

Previous comparative analyses of gene expression during embryogenesis have shown that there are three general patterns of expression for one-to-one orthologs: (1) highly-correlated across all stages, (2) highly correlated during specific stages, and (3) divergent or non-correlated across all stages [Drost et al. 2017; Irie and Kuratani 2014; Kalinka and Tomancak 2012; Schep and Adryan 2013; Yanai 2018]. Most of these studies have been performed between species of the same genus in flies [Kalinka et al. 2010], nematodes [Levin et al. 2012], and vertebrates [Owens et al. 2016; Yanai et al. 2011]. Highly correlated ortholog expression during the phylotypic stage had been reported for comparisons between members of the same genus. Kalinka et al. 2010 compared the expression profiles of 3,019 single-copy orthologs in six *Drosophila* species (*D. ananassae, D. melanogaster, D. persimilis, D. pseudoobscura, D. simulans*, and *D. virilis*). Stages were divided in time intervals of two hours starting from 0-2 hours to 14-16 hours, encompassing a total of eight intervals. They concluded that the expression profiles of 1,188 of these single-copy orthologs are highly correlated across stages and their variance is minimal during germ band retraction (stages 12-13 approximately 8-10 hours after fertilization in *D. melanogaster*).

Additionally, based on GO enrichment analysis they reported that these genes are related to developmental processes and regulation of transcription [Kalinka et al. 2010 supplementary file p. 8-25]. Figure 5 shows nine examples of genes following this expression pattern along with a short description of their functions, whereas figure 6 provides nine examples of genes with divergent expression profiles across stages. Levin et al. 2012 used a different approach based on PCA analysis to compare the expression profiles of 2,095 single-copy orthologs of five *Caenorhabditis* species (*C. remanei, C. briggsae, C. brenneri, C. elegans,* and *C. japonica*) across eight stages (4-cell, first E-lineage division, fourth to seventh AB lineage divisions, ventral enclosure, comma stage, movement or quickening, and L1 larva). In summary they determined that during ventral enclosure transcriptomes are more similar among species, and hence this stage corresponds to the nematode phylotypic stage [Levin et al. 2012]. Ventral enclosure occurs approximately 365–375 minutes post-fertilization in *C. elegans* [Chisholm and Hsiao 2012]. They also reported that as in the case of flies, most of these genes are highly enriched for GO terms associated to developmental processes and regulation of transcription and additionally they generated a list of 294 genes that are higly expressed during ventral enclosure in all nematode species included in their analysis [Levin et al. 2012 supplementary table S3].

Figure 7 shows six example genes that are highly correlated across stages. Results in favor [Drost et al. 2017 Irie and Kuratani 2011; Piasecka et al. 2013] and against [Comte et al. 2010; Richardson et al. 1997; Roux and Robinson-Rechavi 2008] the phylotypic stage (pharyngula or tail-bud stage in vertebrates) as the stage of highest expression similarity are common in studies comparing vertebrates. Other analyses have focused on comparing species of different genera in echinoderms [Gildor and Ben-Tabou de-Leon 2015; Israel et al. 2016] and nematodes [Macchietto et al. 2017] or even between species of different phyla [Levin et al. 2016]. Results related to the stages in which the highest expression similarity occurs differ across studies [Drost et al. 2017]. In the case of nematodes in a study comparing two species of *Caenorhabditis* clade (*Caenorhabditis elegans, Caenorhabditis angaria*) and two species of insect pathogenic nematodes of the genus *Steinernema* (*Steinernema carpocapsae, Steinernema feltiae*) concluded that similar expression patterns are more common during late embryogenesis approximately between comma stage and L1 larva than during larval stages or earlier embryonic stages. A total of 4,164 single-copy orthologs were analyzed in this study [Macchietto et al. 2017].

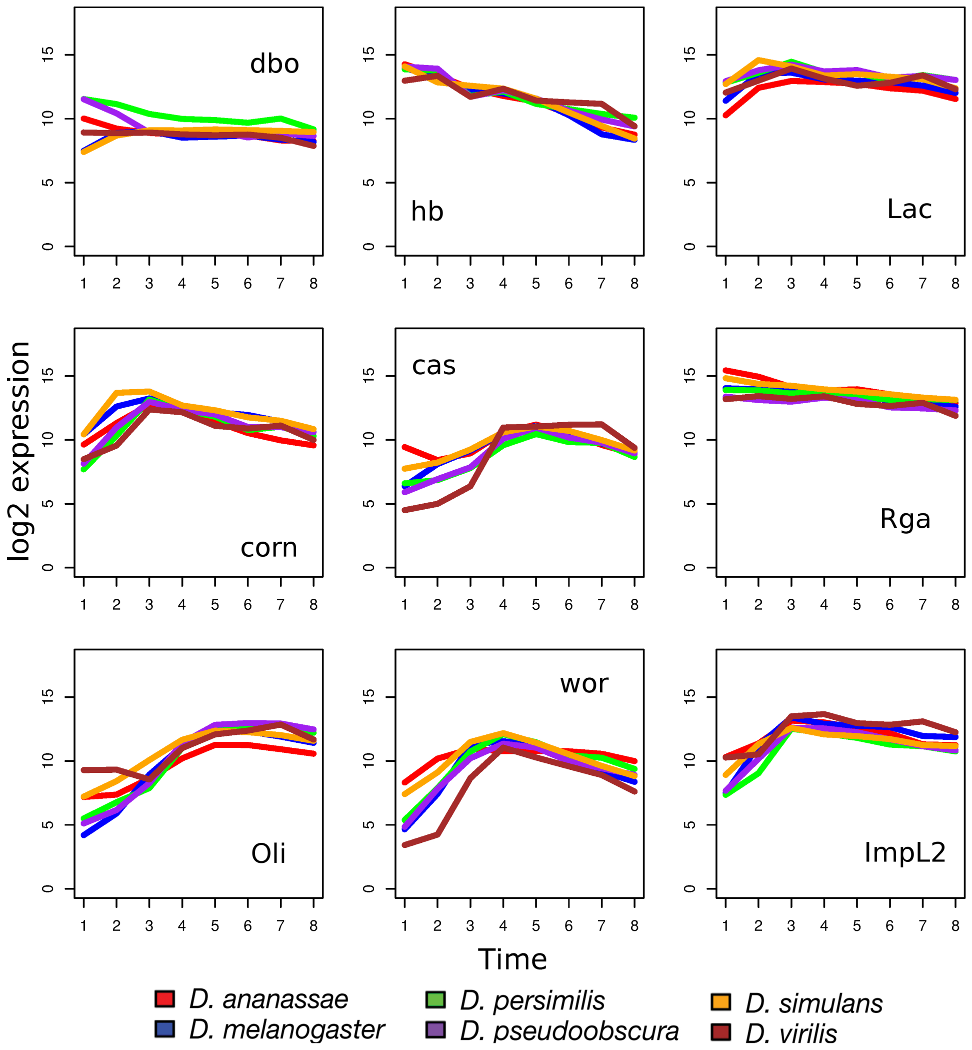


Figure 5. Expression profiles of nine one-to-one orthologs with similar expression profiles across embryonic stages in six *Drosophila species*. Each time corresponds to an interval of two hours from 0-2 to 14-16 hours after fertilization. First row: *db* (*diablo*) interacts with E3 ubiquitin-protein ligase complex which mediates ubiquitination, *hb* (*hunchback*) encodes a zinc finger C2H2 transcription factor involved in the establishment of anterior-posterior gradient, *Lac* (*Lachesin*) encodes a cell surface protein which is required for normal tracheal development. Second row: *corn* (*cornetto*) is a microtubule binding protein, *cas* (*castor*) encondes a zinc finger C2H2 transcription factor and regulates late neuron differentiation, *Rga* (Regena) is a component of the CCR4-NOT complex which is a mRNA deadenylase. Third row: *Oli* (*Olig* family) encodes a bHLH transcription factor which regulates motor neuron axon guidance, *wor* (*worniu*) encodes a zinc finger C2H2 transcription factor that controls neuroblast divisions, *ImpL2* is a insulin-binding protein is a suppressor of insulin-mediated growth in *Drosophila*. Adapted from [Kalinka et al. 2010] supplementary file page 21. Gene functions were retrieved from FlyBase [Thurmond et al. 2019].

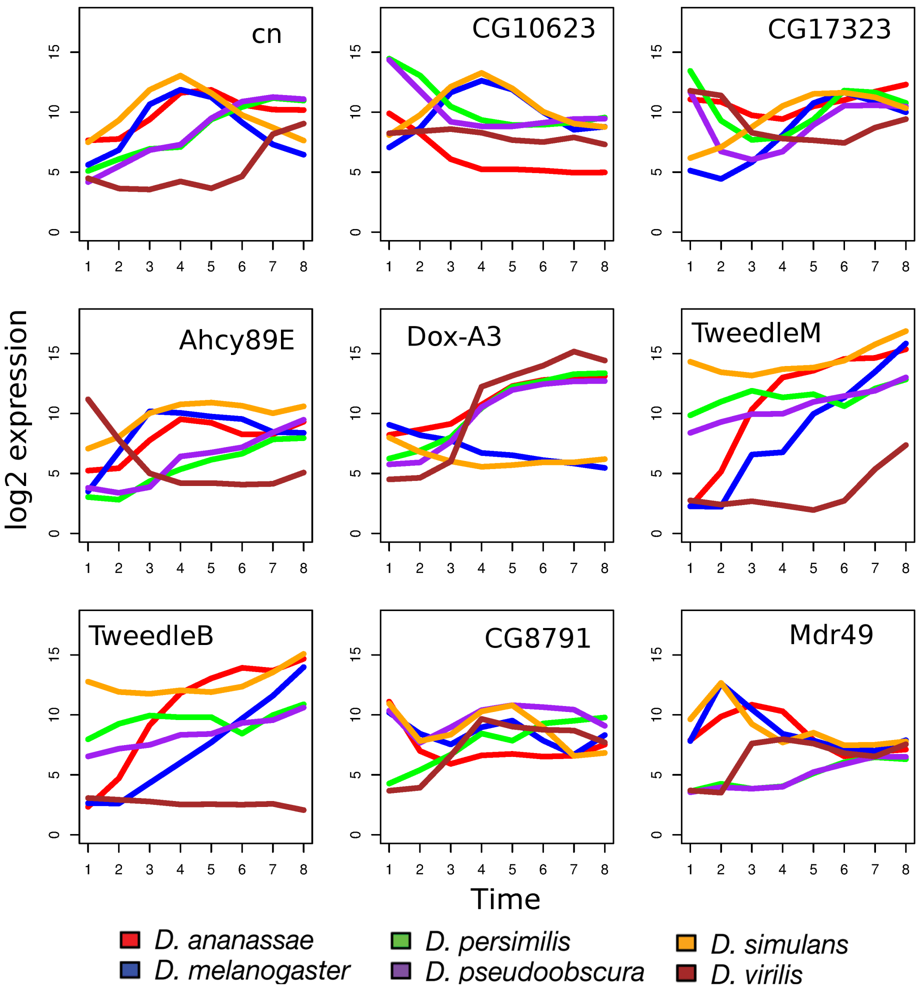


Figure 6. Expression profiles of nine one-to-one orthologs with similar expression profiles across embryonic stages in six *Drosophila species*. Times are intervals of two hours from 0-2 to 14-16 hours after fertilization. First row *cn* (*cinnabar*) is an enzyme with 3-monooxygenase activity, CG10623 homocysteine S-methyltransferase, CG17323 (UDP-glycosyltransferase family 36 member D1). Second row: *Ahcy89E* (Adenosylhomocysteinase like 2), *Dox-A3* (Prophenoloxidase 3) is a copper-containing oxidase involved in the formation of pigments such as melanins, *TweedleM* (domain of unknown function DUF243) its function is unknown. Third row: *TweedleB* (domain of unknown function DUF243) its function is unclear, CG8791 is a solute carrier transmembrane transporter, *Mdr49* (Multi drug resistance 49) is a transmembrane protein that transports substrates which contributes to insecticide resistance. Adapted from [Kalinka et al. 2010] supplementary file page 24. Gene functions were retrieved from FlyBase [Thurmond et al. 2019].

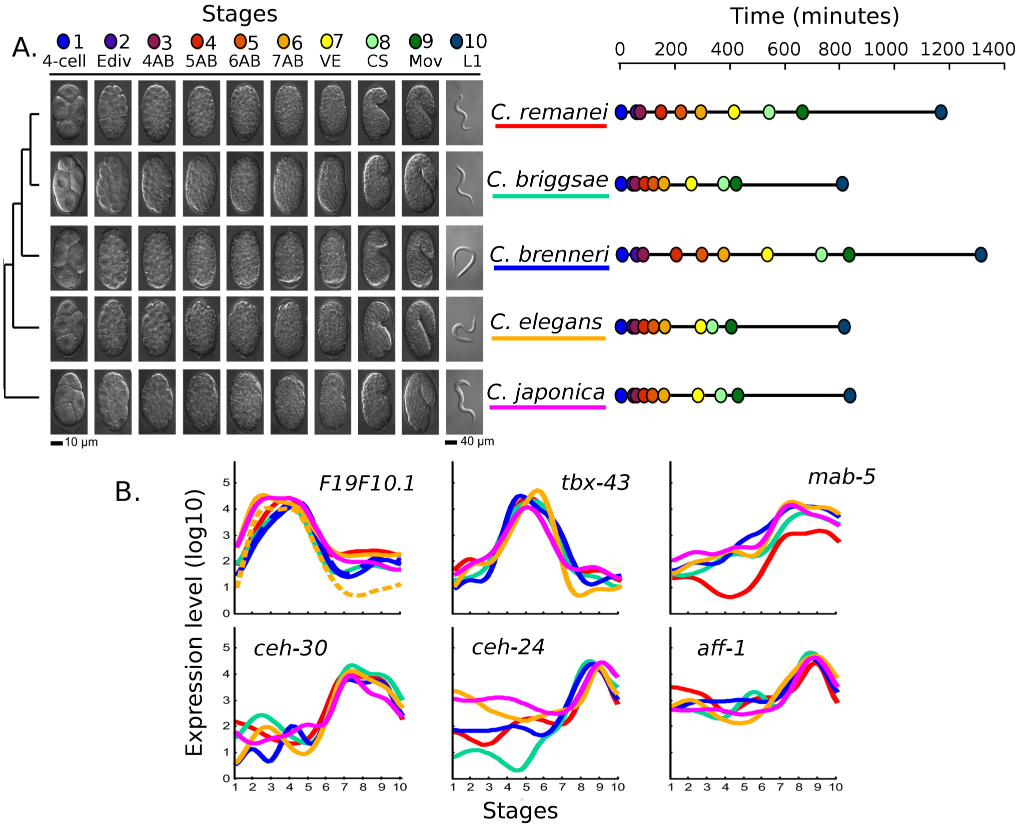


Figure 7. Expression profiles of six single-copy orthologs in five *Caenorhabditis* species. (A) Phylogeny summarizing the evolutionary relationships within the clade, embryonic stages used for the analysis. *C. elegans* and the four other species shared a common ancestor approximately 30 million years ago. Timing of each stage differs among species as shown in the right panel. Samples correspond to 4-cell, first E-lineage division, fourth to seventh AB lineage divisions, ventral enclosure (VE), comma stage (CS), movement (Mov), and L1 larva. (B) Expression profiles for six orthologs. Only *F19F10.1* has a paralog in *C. elegans* represented as a dashed line, its function is unknown. All other genes are single-copy orthologs, *tbx-43* is a T-box transcription factor but is function in unknown. The gene *mab-5* encodes a homeobox transcription factor and its involved in neuronal differentiation, *ceh-30* is another homeobox transcription factor that has been reported to participate in neuron differentiation, both genes are highly expressed during ventral enclosure stage. The NK-homeobox transcription factor *ceh-24* is expressed in pharynx muscles, *aff-1* encodes a surface protein involved in cell fusion during *C. elegans* vulval development. Adapted from [Levin et al. 2012]. Gene functions retrieved from WormBase [Lee et al. 2017].

2.6 Comparative transcriptomics of embryogenesis across phyla

The definition of animal phyla is usually based on morphological data. A body plan is the set of morphological features that define a phylum [Hall chapter 2 p. 18-37]. For example insects are members of Arthropoda which includes animals with segmented bodies and jointed limbs. Because this definition of phylum is based only on morphology, Levin et al. 2016 compared transcriptomes encompassing all embryogenesis for species of ten different phyla to search patterns of conservation and divergence in gene expression. This comparative approach can provide information about differences in the molecular processes underlying the body plans associated to each phyla. The ten phyla included are: [1] Annelida (*P. dumerilii*), [2] Platyhelminthes (*S. polychroa*) , [3] Nematoda (*C. elegans*), [4] Tardigrada (*H. dujardini*), [5] Arthropoda (*D. melanogaster*), [6] Echinodermata (*S. purpuratus*), [7] Chordata (*D. rerio*), [8] Cnidaria (*N. vectensis*), [9] Porifera (*A. queenslandica*), and Ctenophora [10] (*M. leidyi*). Figure 8 (a) shows the phylogeny of the species included in the study, (b) represents how embryos were collected in each species vertical lines are individual time points were embryos were collected, embryos above the time course are representative of the known stage, a solid arrow indicates direct development, and a dashed arrow indirect development.

Comparisons were performed using a set of 11,139 orthologous protein families obtained with OrthoMCL [Li et al. 2003]. Single-copy orthologs between each pair of species were used to compare orthologous gene expression across embryogenesis. In order to compare ortholog expression they used a correlation between windows containing the same number of time points for each pair of species. The most important finding related to this analysis is that ortholog expression is more similar during two phases that the authors called early and late. And the stage in which ortholog expression is less similar is a stage defined as the mid-developmental transition. Interestingly the authors report that this transition overlaps with the known phylotypic stages. As discussed before, the phylotypic stage corresponds to germ band stages in *D. melanogaster* and ventral enclosure in *C. elegans* respectively [Kalinka et al. 2012; Levin et al. 2012].

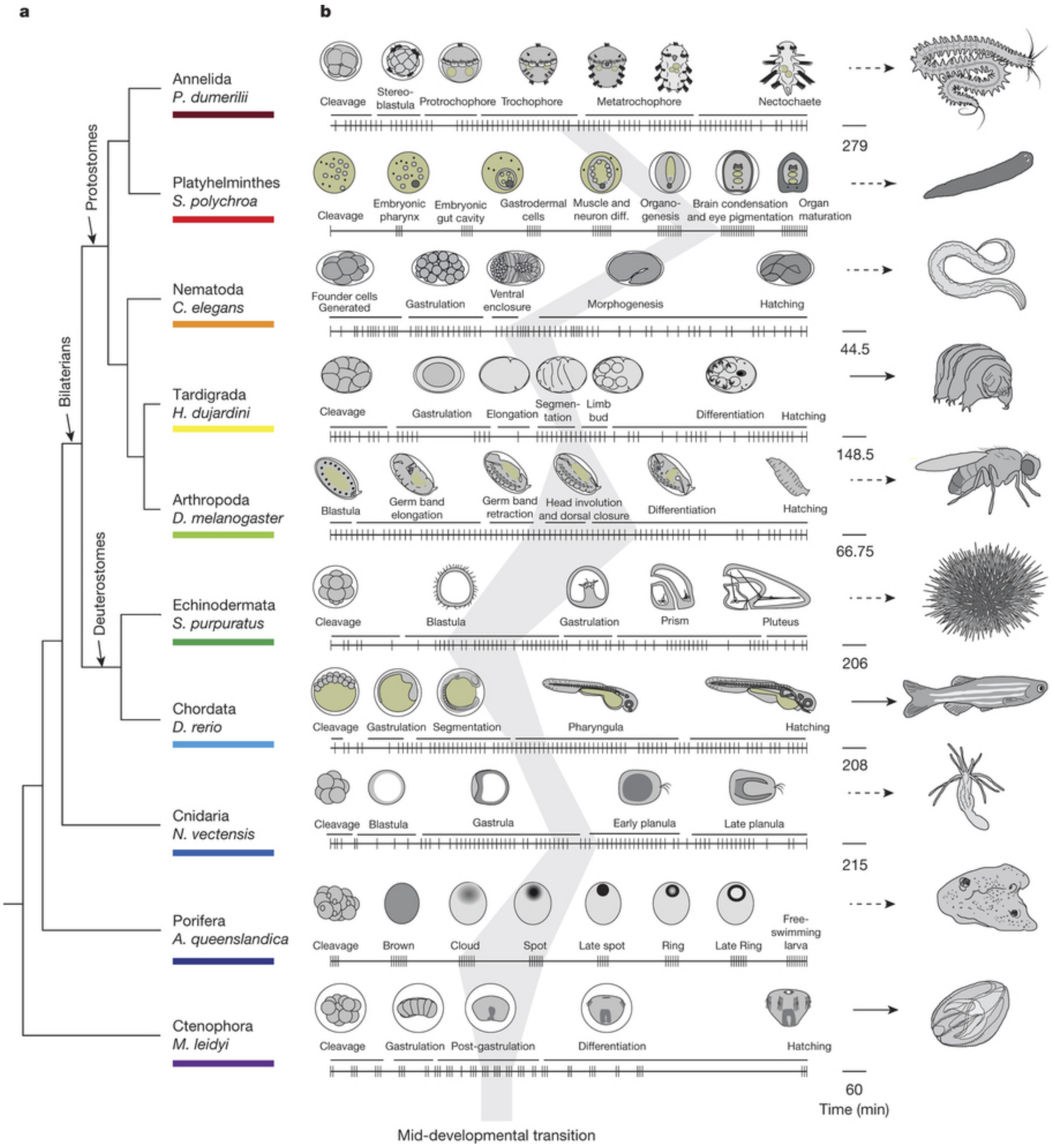


Figure 8. Phylogeny and developmental time courses of the species included in the analysis of Levin et al. (2016). (a) Phylogeny of the ten species included in the analysis. (b) Time course of the sampled embryos across the embryogenesis of each species. Vertical lines represent individual time points in which embryos were collected, embryos are representative of the known stages associated to these time points, lines with a number below are the time scales in minutes for each species, solid arrows represent direct development, and dashed arrows indicate indirect development. The grey area represents the mid-developmental transition in each species. In D. melanogaster the phylotypic stage corresponds to germ band stages, and in C. elegans to the ventral enclosure stage [Kalinka et al. 2012; Levin et al. 2012]. Adapted from [Levin et al. 2016].

Next genes in each species were assigned to one of three temporal categories: early, mid-developmental transition, and late. This classification was made by obtaining the correlation of each gene expression profile with three idealized expression profiles and then selecting the category in which the maximum correlation coefficient was observed. For all 45 pairwise comparisons between species the total number of orthologs that were assigned to the same category was counted. In summary what they found is that the number of orthologs associated to the mid-developmental transition is the lowest compared to early or late orthologs. Figure 9 summarizes these results.

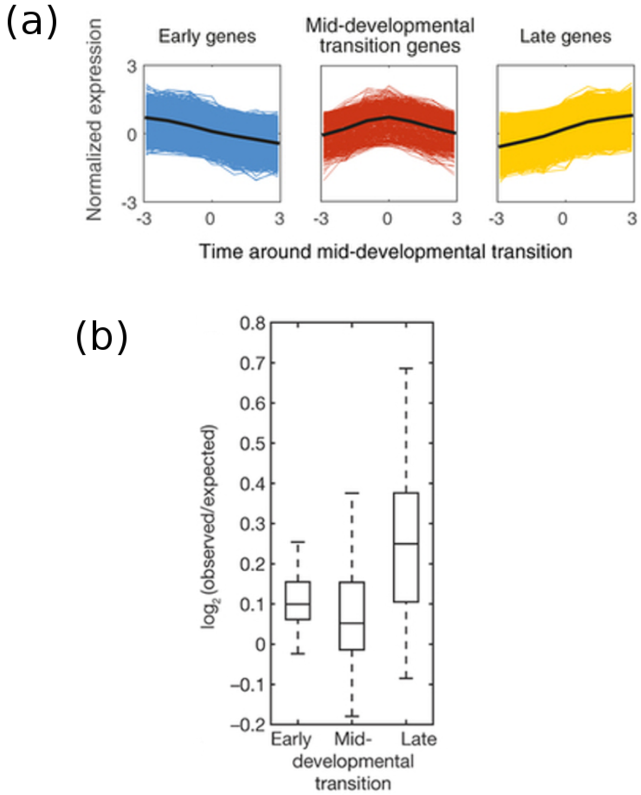


Figure 9. Classification of genes expressed during embryogenesis into three temporal categories. (a) Idealized expression profiles for classifying genes in three temporal categories. Genes in the early category tend to decrease their expression across embryogenesis. Mid-developmental transition genes are highly expressed only during this stage, and late genes increase their expression after the transition. (b) Summary of the ortholog temporal associations for all the 45 pairwise species comparisons in the ten species described above. Odds ratio (observed/expected) greater than 1 means that the number of observed orthologs in the same category is higher than expected by chance, and hence log(observed/expected) $>$ 0. To evaluate the differences between the distribution of mid-developmental transition genes and the other two categories the authors used a Kolmogorov-Smirnov test that resulted in P $<$ 10e-6 for early genes and P $<$ 10e-12 for later genes indicating that both categories are significantly different compared to the mid-developmental transition category. Adapted from Levin et al. 2016.

The authors propose an inverse hourglass model for cross-phyla similarity. In this model ortholog expression tends to more conserved during early and late stages outside the phylotypic stage of each phyla. Figure 10 represents this model compared to the hourglass model for within-phylum similarity. Orthologs expressed during the early phase across the ten species are enriched for chromatin changes, cell cycle, and regulation of gene expression and the authors propose that this phase is associated to cell proliferation. For the late phase they found enrichment related to protein transport, metabolic enzymes, and synaptic factors and propose that these might be related to differentiation processes. In the case of the mid-developmental transition they found enrichment related to signaling pathways such as Wnt, Notch and JAK-STAT and transcription factor enrichment only for the homeobox family. They propose that each phyla uses a different combination of signaling pathways and transcription factors and that this might be related to the development of different body plans across phyla.

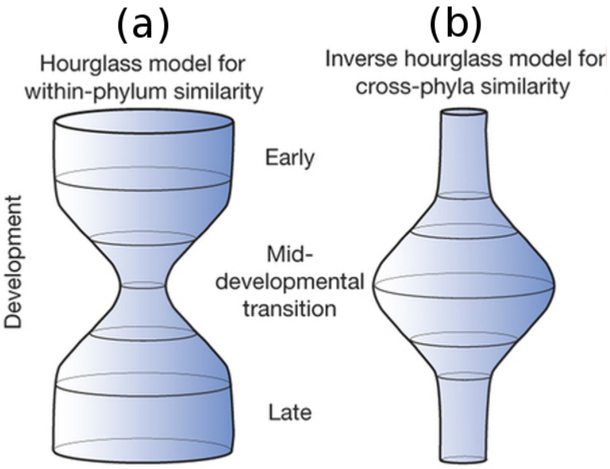

(a)
Hourglass model for within-phylum similarity
(b)
Inverse hourglass model for cross-phyla similarity
Early
Mid-developmental transition
Late
Development

Figure 10. Models for within-phyla and cross-phyla similarity of ortholog expression during embryogenesis. In both models similarity implies high correlation of ortholog expression profiles and highest number of expressed orthologs during the conserved phases. (a) Hourglass model for within-phylum similarity. In this model ortholog expression is less variable during the phylotypic stage or mid-developmental transition. (b) Inverse hourglass model for cross-phyla similarity. Expression of orthologs is more similar outside the mid-developmental transition in early or late phases. Adapted from Levin et al. 2016.

While this study provides a big picture of inter-species ortholog expression patterns during embryogenesis and a model to explain divergent body plans across phyla, it is still unclear if there are specific regulatory events that initiate the processes related to the phylotypic stage. Moreover it is possible that genes that did not correlated with any of the three idealized temporal expression profiles have restricted expression during specific stages. These type of genes could have important roles despite not correlating with these three general expression profiles. It is also possible that these genes are just expressed at different times or stages even in species of the same phyla or genus. This type of change in the timing of expression of orthologs is called a heterochronic shift and it is related to two possible scenarios: (1) the developmental processes in which the genes are involved occurs at different stages but the function of the gene is not altered, (2) the gene might have a different function and probably a larger heterochronic shift is observed, for example a gene expressed only maternally is expressed during organogenesis in the other species [Israel et al. 2016]. The authors of this study also proposed a parameter called the "jump score" to measure how the expression of orthologs shifts between stages of embryogenesis stages in sea urchin species, considering seven expression clusters or co-expression modules for seven known stages from unfertilized egg to early larva [Israel et al. 2016].

3 Justification and aims

3.1 Justification

Considering the previous evidence of how genes with important developmental roles can mantain their expression dynamics among species of the same genus during embryogenesis [Kalinka et al. 2010; Levin et al. 2012; Macchietto et al. 2017], the importance of heterochronic shifts and timing of expression during embryogenesis [Hashimshony et al. 2014; Israel et al. 2016; Gildor and Ben-Tabou de-Leon], and the observation that conservation of co-expression might occur between distant phyla [Levin et al. 2016]. Four important aspects related to the findings of these previous studies can be summarized as follows:

1. Most studies have only compared the expression of single-copy orthologs during embryogenesis but not orthologous groups. Comparing orthologous groups instead of single-copy orthologs could be useful to understand if paralogous genes have divergent or conserved expression profiles compared to other members of the group.

2. Expression profiles can be obtained with several visualization tools from normalized expression data. However estimating the times in which the expression of a gene or set of genes increases or decreases across stages could help to understand their regulation.

3. Although it is not possible to determine all the molecular interactions occuring in a particular developmental stage or process only from transcriptomic data, comparative transcriptomic analyses can be used as an initial exploratory approach to search for interesting genes or processes.

4. Heterochronic shifts of ortholog expression might be related to conserved processes that are occurring at different times and stages or to changes in gene function between species. An initial exploratory analysis might provide information on how this processes could occur.

Therefore we wanted to know if there are orthologous groups that preserve their co-expression during embryogenesis between species with larger divergence times, not only at the genus level. Additionally without making any assumptions about pre-defined expression profiles. For this exploratory analysis we only wanted to determined which are the representative developmental co-expression modules (groups of genes that co-express across embryogenesis) in two species and determine if there are orthogroups (orthologous families) present in both sets of modules.

Two protostome species were selected for this project: *Caenorhabditis elegans* and *Drosophila melanogaster*. These species shared a common ancestor approximately 587-543 million years ago [Giribet and Edgecombe 2017; Rota-Stabelli et al. 2013; Wang et al. 2019]. Moreover both species have well annotated genomes and functional annotation [Raymond et al. 2017; Thurmond et al. 2019], their stages of embryogenesis are well characterized [Gilbert and Barresi 2016], and time course transcriptomes encompassing all stages of embryogenesis are available [Levin et. al 2016]. Moreover using data from these model organisms in the analysis is useful to understand how known developmental processes are related to the genes in the co-expression modules based on their expression patterns and GO enrichment.

3.2 Aims

3.2.1 General aim

Determine if there are orthologous groups that are co-expressed during the embryogenesis of *Caenorhabditis elegans* and *Drosophila melanogaster*.

3.2.2 Specific aims

1. Obtain developmental gene co-expression modules in *Drosophila melanogaster* and *Caenorhabditis elegans*.

2. Obtain the changepoints of the expression profiles and the GO functional enrichments of all modules.

3. Compare co-expression modules in order to identify orthologous groups that are co-expressed in both species.

4 Methods

4.1 General pipeline for data analysis

In this analysis it is necessary to first classify groups of genes that are co-expressed in each species before performing any comparisons, and after this identify orthogroups that are present between modules of different specie. Additionally the expression profile of all modules as well as their GO enrichment should be obtained. In order to achieve this, we developed a pipeline consisting of seven general steps (see figure 12). Datasets are time-course gene expression matrices downloaded from NCBI Gene Expression Omnibus (project accession code GSE70185), accession codes for *Drosophila melanogaster* and *Caenorhabditis elegans* datasets are GSE60471 and GSE60755 respectively, stage annotation for samples (sample mapped to minutes post-fertilization) is available as a supplementary file in the original article [Levin et al. 2016] and orthogroup classification was dowloaded from Ensembl [Zerbino et al. 2018]. Figure 12 summarizes distribution of replicates and sampled time points in both species.

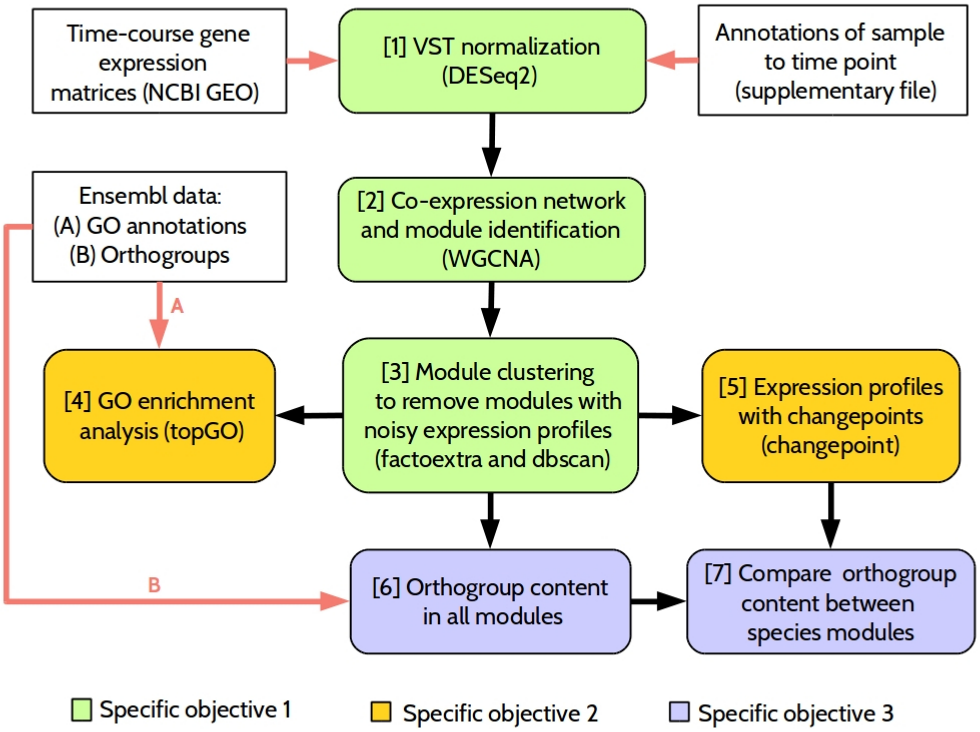

Time-course gene expression matrices (NCBI GEO)
[1] VST normalization (DESeq2)
Annotations of sample to time point (supplementary file)
Ensembl data: (A) GO annotations (B) Orthogroups
[2] Co-expression network and module identification (WGCNA)
A
[4] GO enrichment analysis (topGO)
[3] Module clustering to remove modules with noisy expression profiles (factoextra and dbscan)
[5] Expression profiles with changepoints (changepoint)
B
[6] Orthogroup content in all modules
[7] Compare orthogroup content between species modules
🟩 Specific objective 1
🟨 Specific objective 2
🟪 Specific objective 3

Figure 11. General pipeline of seven steps for co-expression analysis. White boxes with red arrows represent raw input data. Each color box is a step consisting of one script written in R. Black arrows indicate that the output is passed to the next step. Steps one to five are used for specific objective one, while steps six and seven correspond to specific objective two. Gene expression matrices were downloaded from NCBI GEO (project accession codes: GSE70185 and GSE60755), while stage annotation providing the time point (in minutes post-fertilization) associated to each sample is provided as a supplementary file in the original article [Levin et al. 2016].

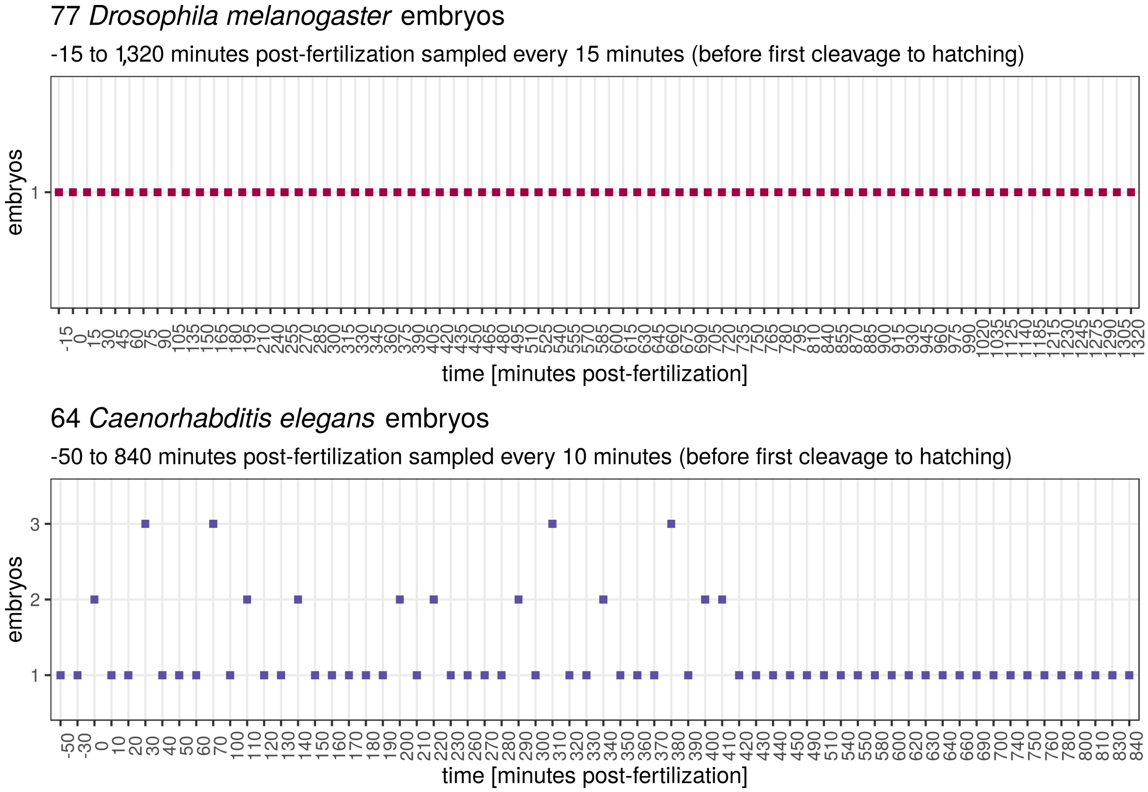


Figure 12. Distribution of samples for the dataset used in the analysis. A total of 77 *Drosophila melanogaster* single embryos were collected from 15 minutes before the first cleavage to 1,320 minutes post-fertilization which corresponds to hatching every 15 minutes. For *Caenorhabditis elegans* embryos were collected in duplicate or triplicate. Nine time points with two replicates: 0, 110, 140, 200, 220, 290, 340, 400, 410 minutes post-fertilization. Four time points with three replicates: 30, 70, 310, 380 minutes post-fertilization. Embryos were collected starting from 50 minutes before the first cleavage to hatching (which occurs approximately 800-840 minutes post fertilization). Transcriptomes of *Drosophila melanogaster* and and *Caenorhabditis elegans* contained 15,682 and 20,687 genes respectively. In both cases the total number of genes includes both coding and non-coding transcripts, transcriptomes were obtained using CEl-Seq [Hashimshony et al. 2012; Levin et al. 2016]. Expression matrices were downloaded from NCBI Gene Expression Omnibus (project accession code GSE70185), accession codes for *Drosophila melanogaster* and *Caenorhabditis elegans* datasets are GSE60471 and GSE60755 respectively, stage annotation for samples (sample mapped to minutes post-fertilization) is available as a supplementary file in the original article [Levin et al. 2016].

4.2 Normalization using VST

Raw time-course gene expression matrices contain information related to transcriptome assembly which should be removed before normalization, an automatic text editor script was implemented in bash and Python to perform this task. Samples mapped to time points (in minutes post-fertilization) were published as supplementary file [Levin et al. 2016], another text editor was implemented in Python to generate a tab-delimited file to incorporate this data into the analysis. Variance-stabilizing transformation (VST) is implemented within the R package DESeq2 and can be obtained using the function getVarianceStabilizedData and it is used for normalization of gene expression matrices. This method can be used in differential expression analysis for both microarray and RNA-Seq data [Love et al. 2014; Durbin et al. 2002]. VST estimates a constant variance for each gene across samples. The main reason to use VST is that it allows to easily identify genes that have zero variance across samples, which is important to spot genes that cannot be used for the construction of the gene co-expression network. Before generating the co-expression networks of both species, all genes that had zero variance across samples were removed from the expression matrices.

4.3 Co-expression network and module identification

Before explaining how this process works, it is important to introduce a few important concepts related to gene co-expresion networks. A gene co-expression network can be represented as an undirected graph, where nodes are genes and edges represent co-expression relationships between genes [Langfelder and Horvath 2008]. In addition to graph representation a network can be also represented as an adjacency matrix containing all co-expression relationships between all pairs of genes [Langfelder and Horvath 2008; Zhang and Horvath 2005]. Adjacency matrices have two important properties: (1) diagonal of ones by convention which represents co-expression of any gene with itself, (2) symmetry meaning that there is no distinction between co-expression measurements performed between the same pair of genes, this means that for any pair of genes only one correlation is necessary to represent co-expression between them [Zhang and Horvath 2005]. There are other types of networks were this property is not valid. For example, in a Boolean model of a gene regulatory network a transcription factor could activate a target gene that does not affect the expression of the transcription factor and hence two numbers are required to represent the influence of the transcription factor on the gene (+1), and the absence of influence of the target on the transcription factor (represented with 0).

In Figure 13 co-expression is stronger between a and b than between b and c. And also, the co-expression between b and c is the same if c is chosen before b, which is a consequence of matrix symmetry. But, how are these co-expression values assigned? it is possible to use correlation as a measure of co-expression [Zhang and Horvath 2005]. WGCNA package has three standard correlation measures: Pearson, Spearman and biweight midcorrelation [Langfelder and Horvath 2008]. However, biweight midcorrelation (abbreviated as bicor) has a better performance for low number of samples and also facilitates the detection of outlier samples in the data [Zhang and Horvath 2005]. Bicor for x and y representing the expression of two genes acrosss m samples, obtained from a normalized gene expression matrix is be defined as

$$bicor(x,y) = \sum_{a=1}^{m} \tilde{x}_a \, \tilde{y}_a \tag{1}$$

Where $\sum$ represents a summation symbol over $a = 1, 2, \ldots, m$ components of x and y. An important aspect of this type of correlation is that columns of gene expression matrix are first transformed using a weight function that depends on the median expression of each gene across samples, from this transformation the components $\tilde{x}_a \, \tilde{y}_a$ can be obtained [Langfelder and Horvath 2012; Zhang and Horvath 2005]. WGCNA assigns a bicor value to each possible gene pair and generates a co-expresion similarity matrix S, and then uses it to compute an adjacency matrix A which considers the co-expression network as an approximated scale-free network (a network having a low mean number of edges connected to each node and a few highly connected nodes called hubs), in order to do this it uses a parameter called soft-power β which is selected based on the following criterion: "only consider those parameter values that lead to a network satisfying scale-free topology at least approximately, e.g. signed $R^2 > 0.80$ " [Langfelder and Horvath, 2008; Zhang and Horvath 2005]. Using the adjacency matrix A, a topological overlap matrix is computed (TOM, represented with Ω) [Langfelder and Horvath 2008; Zhang and Horvath 2005]. TOM can be interpreted as a measure of "relative interconnectedness" between two nodes [Zhang and Horvath 2005]. It is computed using the connectivity of each node with all other nodes and arranged into a matrix of the same dimensions as the adjacency matrix [Zhang and Horvath 2005; Langfelder and Horvath 2008]. Finally a dissimilarity matrix D is computed using TOM, with a simple matrix operation

$$D = L - \Omega \tag{2}$$

L represents a matrix of ones and Ω is TOM [Zhang and Horvath 2005]. This is the matrix that is used to identify groups of co-expressed genes as will be discussed in the following section. A diagram showing the basic steps for gene co-expression network construction is shown in figure 13. In principle, an adjacency matrix is sufficient to represent a gene co-expression network. TOM-based dissimilarity matrix is required to identify co-expression modules using a hierarchical clustering method.

Modules are defined as “groups of genes whose expression profiles are highly correlated across the samples” [Zhang and Horvath 2005].Once TOM-based dissimilarity matrix (D) is obtained, WGCNA separates the entire gene co-expression network into clusters of highly co-expressed genes called co-expression modules [Langfelder and Horvath 2008; Zhang and Horvath 2005]. Two general steps required to obtain modules: (1) average linkage hierarchical clustering which uses dissimilarity matrix as input to generate an initial set of candidate modules, (2) branch cutting to identify a final set of modules [Murtagh 1983; Murtagh and Contreras 2012].

4.3.1 Average linkage hierarchical clustering

Also known as unweighted pair group method with arithmetic mean (UPGMA) is an agglomerative hierarchical clustering method [Murtagh and Contreras 2012]. Agglomerative means it clusters data assuming that all objects (genes) are initially separated and then proceeds to link (or agglomerate) them iteratively based on their dissimilarities. In each iteration a new linkage is generated from the pair of most similar previous objects, which can be thought as the union of two sets. Because each new linkage represents an object which is absent in the previous step, it is necessary to update dissimilarites in each iteration. For agglomerative hierarchical clustering the standard method for this task is called Lance-Williams dissimilarity update formula [Murtagh and Contreras 2012]. R includes a function called hclust which can perform average linkage hierarchical clustering and other hierarchical clustering methods. WGCNA uses this function to generate an initial tree based on dissimiliarity matrix [Langfelder et al. 2007; Zhang and Horvath 2005].

4.3.2 Branch cutting

Given a dendrogram (tree) obtained with hclust, WGCNA then estimates a significant set of co-expression modules by cutting the initial tree [Langfelder et al. 2007; Langfelder 2008]. Standard branch cutting method in WGCNA is Dynamic Tree Cut and consists of two steps: (1) fixed height branch cut and (2) adaptive tree cut. Fixed height branch cut prunes the tree at a specific height (0.99 is the default value) and generates an initial number of branches, in WGCNA it is implemented in the function cutreeStatic [Langfelder et al. 2007]. Initial branches are large and should be cut too to identify modules. Adaptive tree cut is used to solve this problem, it is based on the idea that within a sequence of heights in a tree values can increase or decrease. At some points in the sequence, values can start increasing or decreasing preserving this trend for a finite number of points, at break points were a significant change from decreasing to increasing trend is detected the algorithm cuts the tree again. This process is repeated in each initial branch until no more modules are found [Langfelder et al. 2007]. WGNA includes this method in the function cutreeDynamic [Langfelder and Horvath 2008].

By default, WGCNA assigns a color to each co-expression module, being grey the only exception which is used for genes that were not assigned to any module [Langfelder et al. 2007; Langfelder and Horvath 2008]. Figure 13 shows a diagram of the main steps required for module identification with WGCNA.

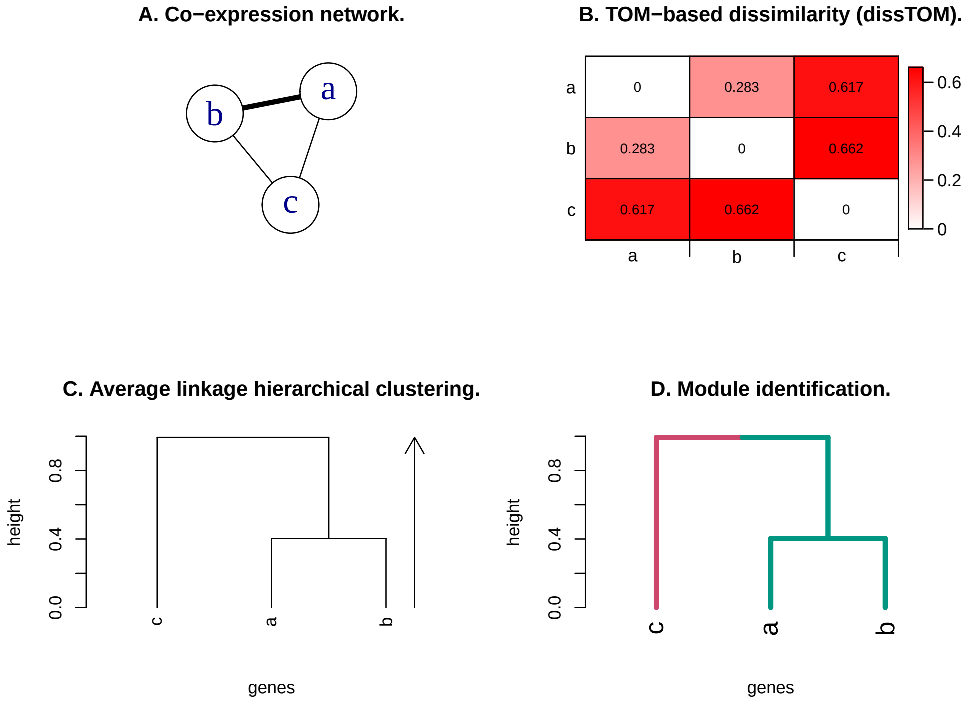

A. Co−expression network.
a
b
c
B. TOM−based dissimilarity (dissTOM).
a
b
c
0
0.283
0.617
0.283
0
0.662
0.617
0.662
0
0.6
0.4
0.2
0
C. Average linkage hierarchical clustering.
height
0.0
0.4
0.8
c
a
b
genes
D. Module identification.
height
0.0
0.4
0.8
c
a
b
genes

Figure 13. Module identification in WGCNA. Gene co-expression network (A) and its associated dissTOM matrix (B). Average linkage hierarchical clustering clusters genes in an agglomerative way using dissTOM; all genes start in their own cluster and then are linked iteratively until all genes are clustered, the arrow represents the direction in which this process occurs (C). Finally representative clusters of genes called modules are inferred from the initial tree using a dynamic tree cut algorithm [Langfelder et al. 2007].

4.3.3 Input parameters used for WGCNA analysis

WGCNA analysis requires two initial parameters: (1) soft-power β for the gene co-expression network that can be estimated using a tool already implemented in the package using the input expression matrix and (2) minimum module size (minimum number of genes) to restrict module size in module identification [Langfelder and Horvath 2008]. Estimated soft-powers were $\beta = 12$ and $\beta = 6$ for *Drosophila melanogaster* and *Caenorhabditis elegans* respectively. Minimum module size was set to 10 in both cases. This parameter can be set to any arbitrary integer value between one and the total number of genes in the expression matrix, but a low value was chosen to avoid mixing small modules into larger ones without any other biological criteria.

4.4 Module clustering

We observed that there were modules that had very low expression across samples, the most common pattern was that they only had a peak of expression in one sample and almost no expression in other samples. Hence we reasoned that this type of modules were most probably noise. Additionally the grey module that contains all genes that were not assigned to any module do not follow any pattern of expression. Other modules had more dynamic expression across samples. WGCNA uses eigengenes as a representative expression profile of a module, eigengenes are obtained from the expression matrix containing only the genes that are members of the module [Langfelder and Horvath 2007; Langfelder and Horvath 2008]. Figure 14 shows four examples of modules to illustrate these type of expression patterns.

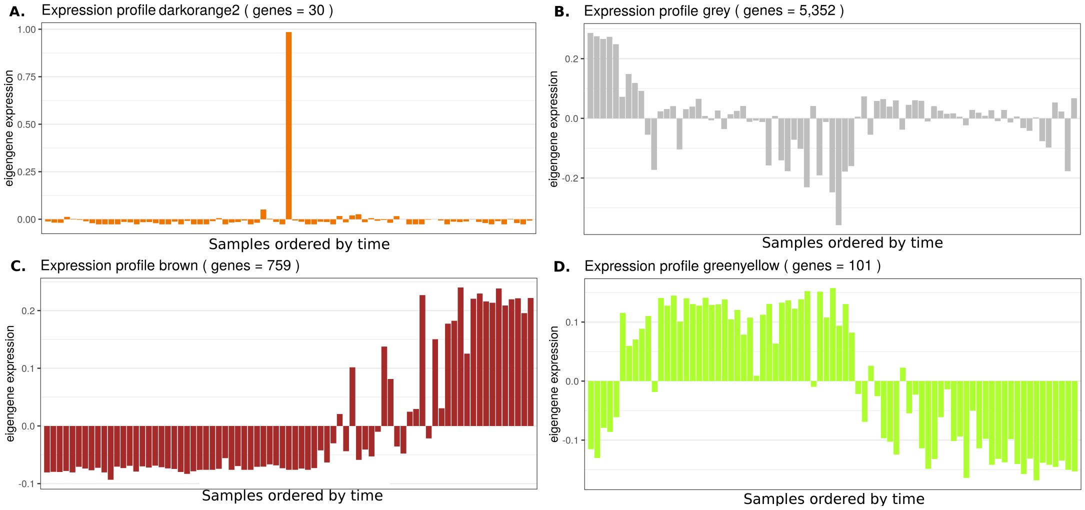


Figure 14. Examples of expression profiles of four modules. (A-B) Modules with noisy expression profiles. The darkorange2 is a module that has low expression across samples with only one peak in one sample, the grey module contains all genes that could not be assigned to any module and hence it does not have any clear pattern of expression. (C-D) Dynamic expression modules, brown module increases across samples, while greenyellow module starts increasing its expression at the beginning but starts decreasing across the last samples.

Because the total number of modules can be arbitrarily large, a R script to separate dynamic and noisy expression modules was implemented. First, it performs a PCA of all eigenge expression profiles for each species using the R package factoextra [Kassambara 2017]. As expected, modules with noisy expression profiles have low variance and contribute poorly to the first and second principal components, while modules that have dynamic expression profiles across samples contribute more to principal components as shown in figure 15.

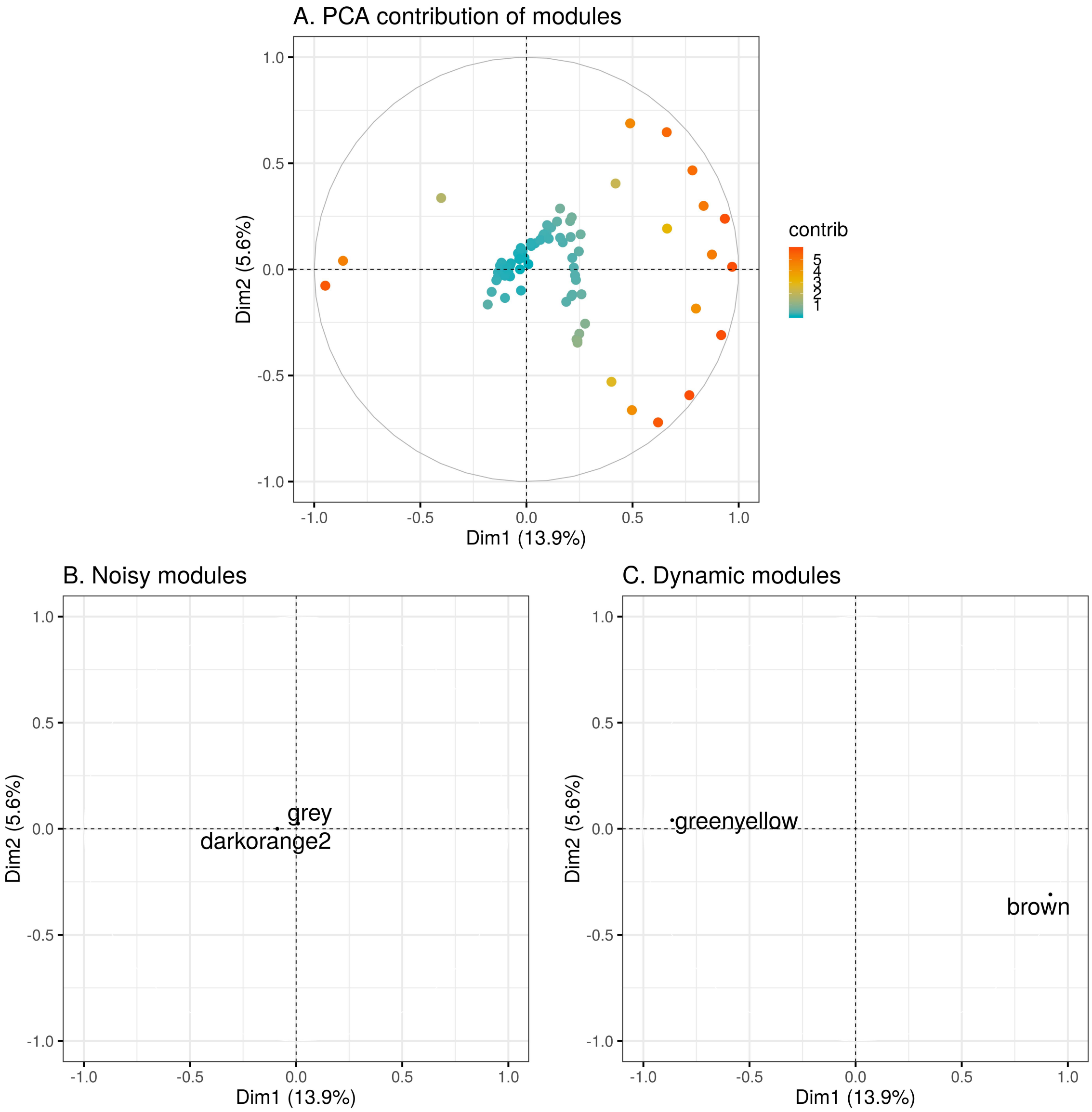


Figure 15. Example of PCA analysis of expression profiles. (A) Modules are distributed near the center or in the edges which means that they contribute less or more to the principal components respectively. (B) As expected modules with noisy expression profiles are near the center because of their low variance, only two examples are shown. (C) Modules that have dynamic expression profiles across samples tend to be far from the center.

Next, a DBSCAN (Density-based spatial clustering of applications with noise) analysis of the expression profiles in each species was made using the R package dbscan [Hahsler et al. 2018]. This type of clustering is useful to separate noisy signals from the rest of the data in distinct clusters, additionally allows the identification of clusters with different shapes [Kassambara 2017 chapter 19 p. 177-185]. Figure 16 shows an example of DBSCAN results, that contains the same example modules darkorange2, grey, brown and greenyellow described above.

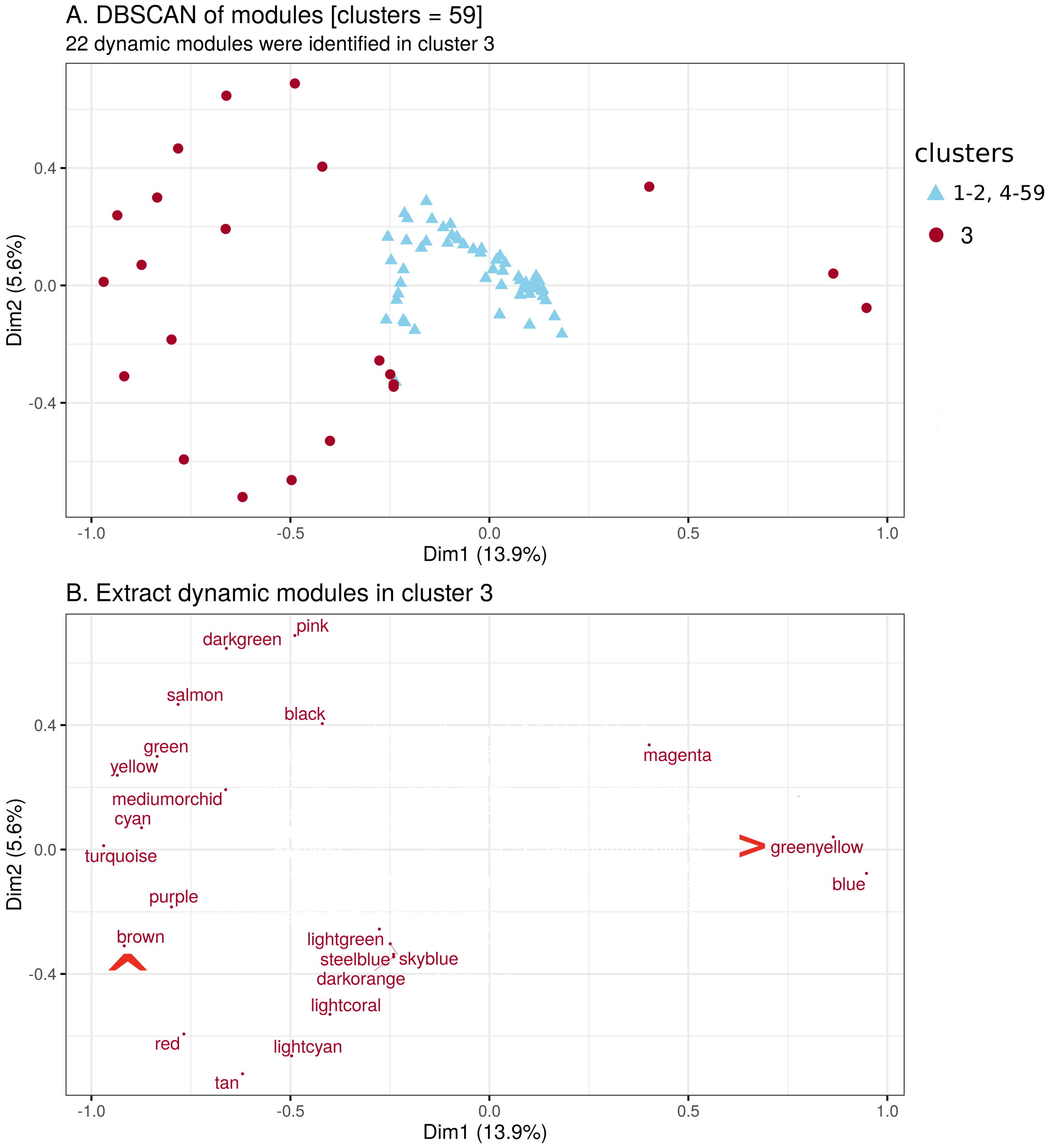


Figure 16. Example of DBSCAN analysis of expression profiles. (A) In this example there are a total of 81 modules that were assigned to 59 clusters, and cluster 3 contains the modules with dynamic expression. (B) Shows only the modules in cluster 3. Arrows indicate the position of the example modules brown and greenyellow.

4.5 GO enrichment analysis

For modules with dynamic expression we performed a GO enricment analysis for biological functions using the R package topGO [Alexa et al. 2006]. Reported p-values are for Fisher's exact test corrected for false discovery rate (FDR) of $p < 0.05$.

4.6 Expression profiles with changepoints

In order to characterize the expression profiles of each module with dynamic expression first we computed the arithmetic mean expression of all genes in the module. Then we estimated the time points in which the mean expression increases or decreases significantly using the R package changepoint [Killick et al. 2012], this package uses maximum-likelihood algorithms to detect changepoints in time-course data or similar data. Then we generated a plot representing each expression profile and its changepoints. Figure 17 shows an example of plot with 400 simulated data points and the estimated changepoints.

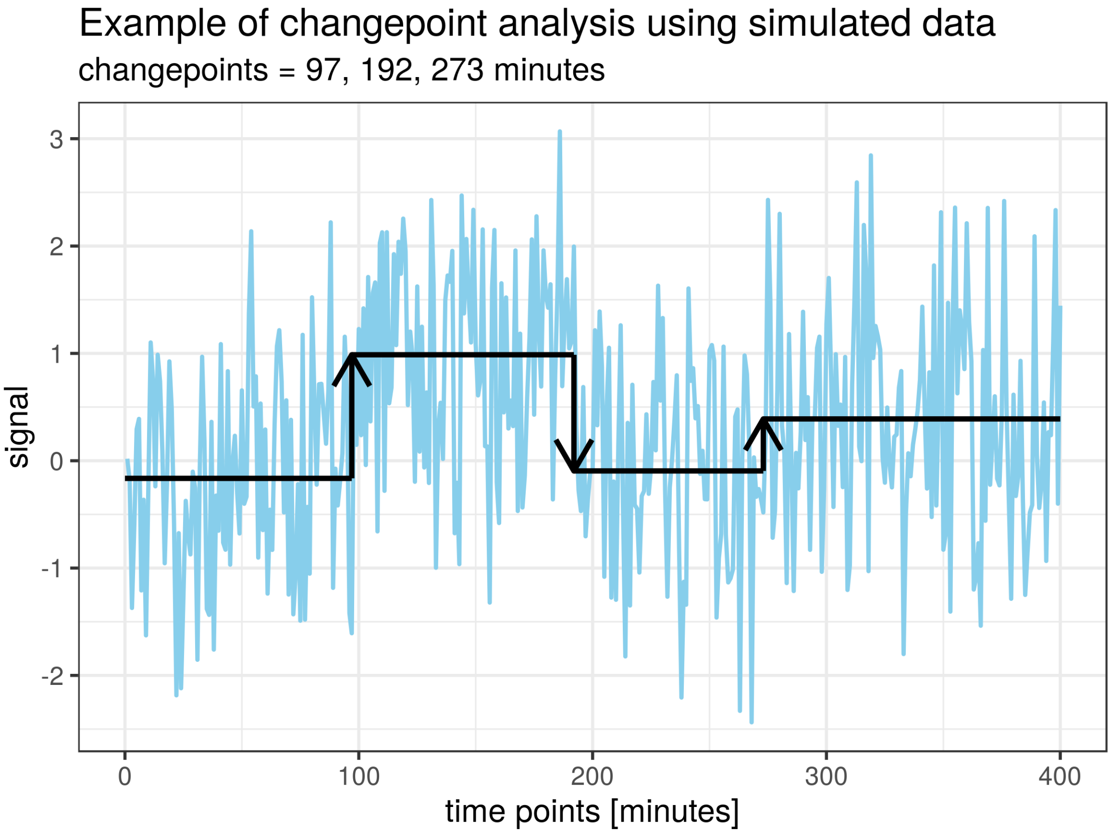


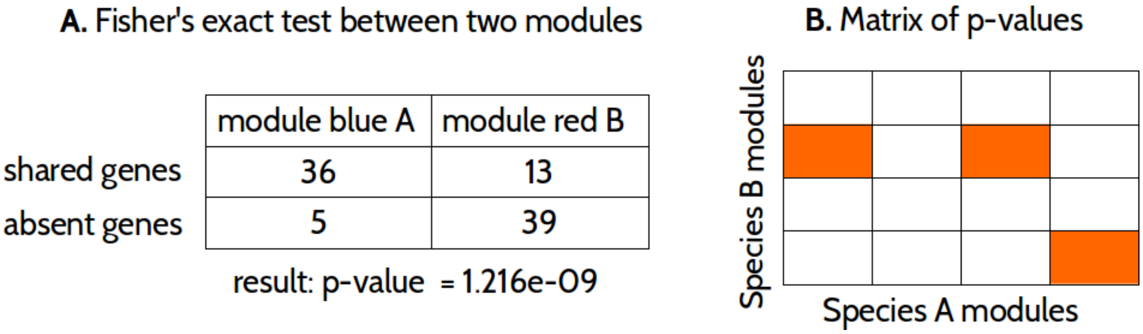
Figure 17. Example of changepoint analysis. Data consist of 400 simulated Gaussian data points. Three changepoints were estimated at 97, 192 and 273 minutes and are represented by vertical arrows. Horizontal lines represent time intervals in which the signal does not change significantly.

4.7 Orthogroup content in all modules

Orthologous families or orthogroups are groups of proteins between two or more species and that are clustered based on their phylogenetic relationships, there are various computational tools for orthogroup clustering the most commonly used are OrthoFinder and OrthoMCL [Emms and Kelly 2015; Li et al. 2003]. In this step we downloaded the orthogroup classification (gene families) from Ensembl database for *Caenorhabditis elegans* and *Drosophila melanogaster* using the online tool BioMart [Zerbino et al. 2018]. Genome versions for *Caenorhabditis elegans* and *Drosophila melanogaster* are Wcel235 and BDGP5 respectively in the original dataset (NCBI accession codes GSE60471 and GSE60755 respectively) [Levin et al. 2016] and genes in the same version were used for Ensembl data. Gene indentifiers in each module were converted to a table containing this identifier along with their gene names, gene family identifiers, and gene family description.

4.8 Compare orthogroup content between species modules

Two modules of different species can contain genes that are members of the same orthogroups. From these subset of orthologs present in both modules, it is still possible that in one species only a small fraction of these orthologs were detected in the module while in the other species module most of the members were detected. In order to evaluate if the number orthologs present in one module differs from the number of orthologs detected in the module of the other species a Fisher's exact test (adjusted to $p < 0.05$) [McDonald 2014] was performed as shown in figure 18.



A. Fisher's exact test between two modules

	module blue A	module red B
shared genes	36	13
absent genes	5	39

result: p-value = 1.216e-09

Figure 18. Fisher's exact test to compare orthogroup content between species modules. (A) Example of how the test is performed between two modules. Shared genes indicat the members of orthogroups detected in both modules. Absent genes are the genes that are members of these orthogroups that were not observed in modules. (B) The test is performed between all modules and the results are organized in a matrix containing all p-values, colored cells represent significant results. Because the total number of comparisons was large, p-values were corrected using the Bonferroni method implemented in the R function p.adjust.

The null hypothesis is that "that the relative proportions of one variable are independent of the second variable" [McDonald 2014]. In this case this means that the null hypothesis is that the fraction of orthologs detected in a module does not differ from the fraction of orthologs detected in a module of the other species. If the difference is statistically significant it means that most orthologs are expressed in one module and only a small fraction is detected in the module of the other species. The test itself does not tell why similar fractions of orthologs are observed in modules of different species. It only tells if the fraction of detected orthologs differs between modules.

5 Results

A summary of results for both species is presented here including a description of the co-expression modules that were detected. Complete expression profiles and GO enrichment results are provided as supplementary figures. Only general aspects of modules are described. In the final results for the comparisons of orthogroup content between species modules only two cases will be discussed in detail. Total number of genes was variable across modules in both species. Using data from Ensembl database downloaded using BioMart [Zerbino et al. 2018], the content of protein and non-coding genes was obtained in all modules. Gene or transcript types reported here follow the Ensembl classification and belong to one of the following biotypes [Zerbino et al. 2018](available at `https://www.ensembl.org/info/genome/genebuild/biotypes.html`):

1. Protein coding: "Protein coding: Gene that contains an open reading frame (ORF)".

2. ncRNA: "ncRNA: A non-coding gene".

3. lincRNA: "(long intergenic ncRNA): Transcripts that are long intergenic non-coding RNA locus with a length $> 200\,bp$. Requires lack of coding potential and may not be conserved between species".

4. snoRNA: "Small RNA molecules that are found in the cell nucleolus and are involved in the post-transcriptional modification of other RNAs".

5. miRNA/pre-miRNA: "A small RNA ($\sim 22\,bp$) that silences the expression of target mRNA".

6. snRNA: "Small RNA molecules that are found in the cell nucleus and are involved in the processing of pre messenger RNAs".

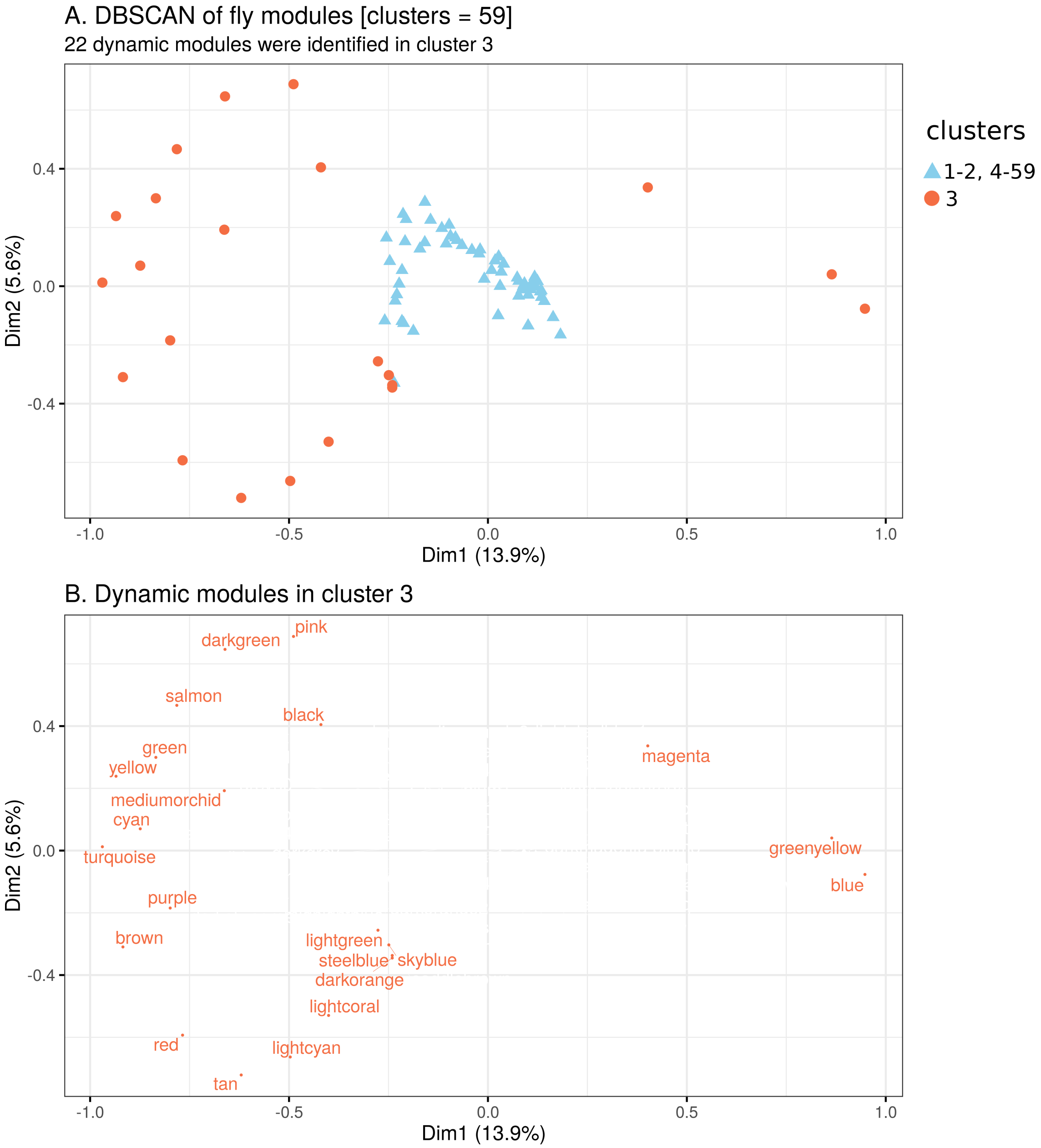
7. tRNA: "A transfer RNA, which acts as an adaptor molecule for translation of mRNA".

8. rRNA: "The RNA component of a ribosome".

9. Pseudogene: "A gene that has homology to known protein-coding genes but contain a frameshift and/or stop codon(s) which disrupts the ORF. Thought to have arisen through duplication followed by loss of function".

5.1 Co-expression modules in *Drosophila melanogaster*

The initial expression matrix contained 15,682 coding and non-coding genes sampled across 77 time points [Levin et al. 2016] and 13,842 genes were used for module identification after normalization. WGCNA analysis identified 81 co-expression modules. Only 22 modules with dynamic expression (same number as in *C. elegans*) were found using DBSCAN [Hahsler et al. 2018]. Figure 19 shows the DSCAN results for all modules in this species.



A. DBSCAN of fly modules [clusters = 59]

22 dynamic modules were identified in cluster 3

B. Dynamic modules in cluster 3

Figure 19. DBSCAN results for *Drosophila melanogaster* modules. (A) Clusters identified by DBSCAN. (B) Cluster 3 contains the 22 modules with dynamic expression across samples.

Gene content in all modules is summarized in figure 20. All modules contain a total of 6,836 genes. First panel shows the total number of genes in each module, middle panel represents the number of protein coding genes and the last panel is the content of non-coding RNAs detected in each module. Table 1 is a summary of total gene number, first changepoint (first time point in which the expression of a module increases), known stage containing this changepoint, and total number of GO terms (biological function) obtained using topGO [Alexa et al. 2006].

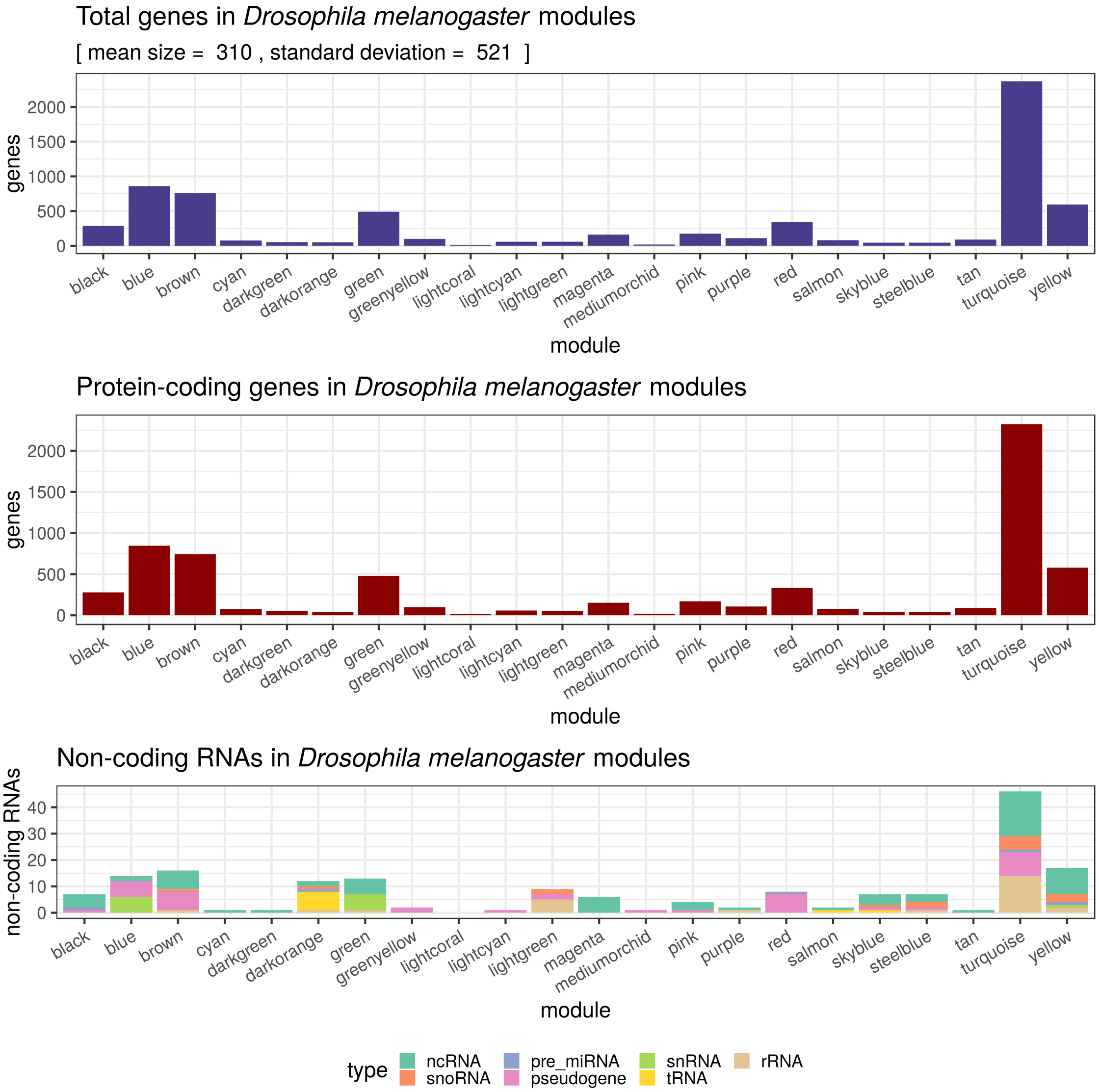


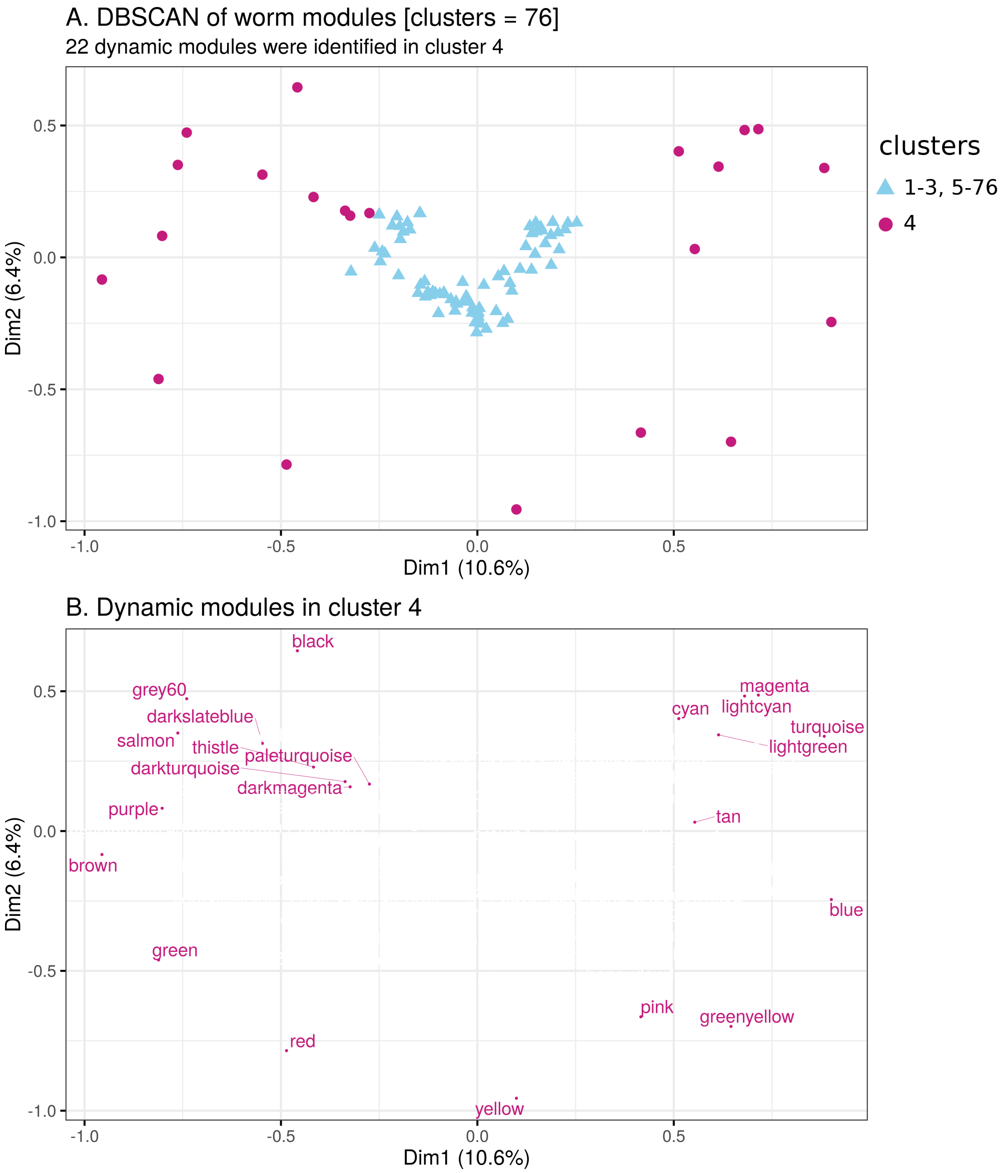
Figure 20. Gene content in *Drosophila melanogaster* modules. First panel is total gene number, second panel number of protein coding genes and last panel is the number of non-conding RNAs.

Module	Genes	GO terms	FCTP [mpf]	Stage
black	286	221	45	2, early cleavage, MZT
greenyellow	101	78	45	2, early cleavage, MZT
magenta	160	126	45	2, early cleavage, MZT
darkorange	50	22	195	6, gastrulation
green	491	355	240	9, germband elongation
yellow	595	388	345	11, parasegmentation
pink	173	85	390	11, parasegmentation
lightgreen	58	31	600	13, germ band retraction
darkgreen	51	17	630	14, dorsal closure
brown	759	476	645	14, dorsal closure
cyan	77	40	645	14, dorsal closure
mediumorchid	18	6	645	14, dorsal closure
red	342	192	645	14, dorsal closure
salmon	80	41	645	14, dorsal closure
skyblue	47	21	645	14, dorsal closure
steelblue	44	19	645	14, dorsal closure
lightcoral	14	10	780	15, cuticle deposition
lightcyan	60	37	795	15, cuticle deposition
tan	90	54	900	16, VNC shortening
turquoise*	2,370	1,699	45	2, early cleavage, MZT (?)
blue*	861	732	160	11, parasegmentation (?)
purple*	109	99	780	15, cuticle deposition (?)

Table 1. Co-expression modules in *Drosophila melanogaster*. Total GO terms indicate the total number of terms that were found for the biological process category in topGO [Alexa et al. 2006]. Modules are ordered according to the first changepoint (FCTP) in minutes post-fertilization [mpf] in which their expression increases. There are three exceptions indicated by asterisks. First two exceptions are modules turquoise and blue in which their first changepoint corresponds to a decrease in expression and a general trend to increase or decrease respectively. Module purple is the third exception, it increases its expression only once while any other module has more than one changepoint. Stages are indicated in the standard nomenclature of 17 stages along with an important event observed during each stage. Interval for stage 2 is 25-70 mpf. MZT indicates maternal-to-zygotic transition and occurs from 48 to 150 mpf. Gastrulation occurs during stages 6-7 approximately 180-195 mpf. Stage 9 is observed within 230-260 mpf and corresponds to slow germ band elongation. During stage 11 epidermal parasegmentation is observed and occurs from 320 to 440 mpf. Stage 13 occurs within 560-620 mpf, germ band retraction and central nervous system differentiation are observed. Dorsal closure is observed during stage 14 which starts at 620 mpf and ends at 680. Stage 15 is observed from 680 to 800 mpf and it is characterized by head involution, end of dorsal closure and cuticle deposition. During stage 16 within 800-900 mpf the shortening of the ventral nerve cord (VNC) is observed. Information for stages adapted from [Campos-Ortega and Hartenstein 1997; Hales et al. 2015; Lee et al. 2014; Palfy et al. 2017; Tadros and Lipshitz 2009].

5.2 Co-expression modules in *Caenorhabditis elegans*

In this species the original expression matrix contained 20,687 coding and non-coding genes sampled across 64 time points [Levin et al. 2016]. After normalization which removes genes with zero variance across samples 18,553 genes were used for further analysis. A total of 97 co-expression modules were identified using WGCNA [Langfelder and Horvath 2008]. From these original set only 22 modules with dynamic expression were identified using DBSCAN [Hahsler et al. 2018]. Figure 21 shows the DSCAN results for all modules.


A. DBSCAN of worm modules [clusters = 76]

22 dynamic modules were identified in cluster 4

B. Dynamic modules in cluster 4

Figure 21. DBSCAN results for *Caenorhabditis elegans* modules. (A) Clusters identified by DBSCAN. (B) Cluster 4 contains the 22 modules with dynamic expression across samples.

Figure 22 shows a summary of gene content across modules. Gene types are from Ensembl [Zerbino et al. 2018] as described above. The 22 modules with dynamic expression contain a total of 13,454 genes. First panel shows the total number of genes, second panel number of protein coding genes and the third panel is the content of non-coding RNAs detected in modules. Table 2 is a summary of total gene number, first changepoint (first time point in which the expression of a module increases), the known stage in which this first changepoint occurs, and total number of GO terms (biological function) obtained for each module.

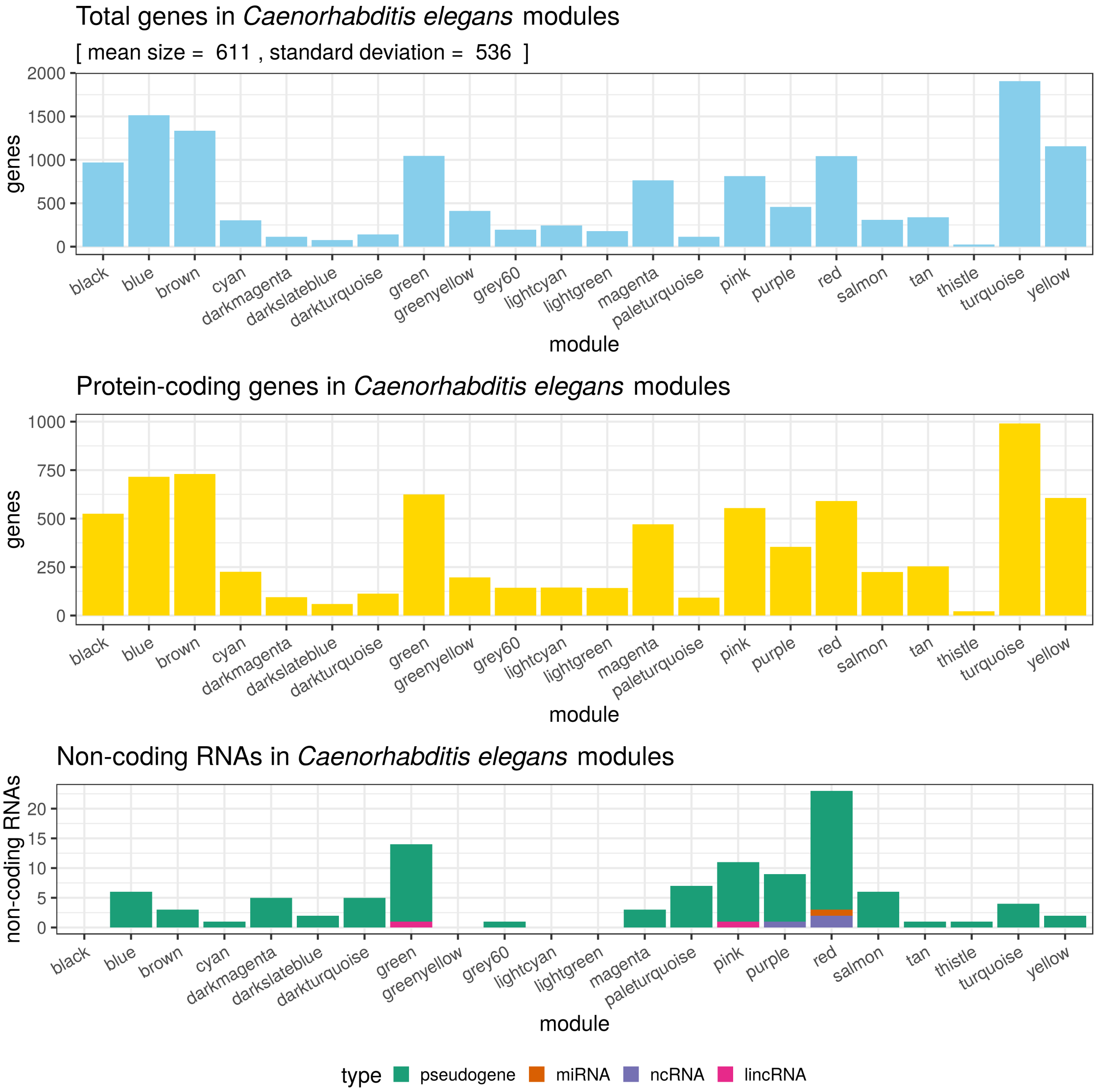


Figure 22. Gene content in *Caenorhabditis elegans* modules. First panel is total gene number, second panel number of protein coding genes and last panel is the number of non-conding RNAs.

Module	Genes	GO terms	FCPT [mpf]	Stage
darkslateblue	76	22	0	2-cell
grey60	196	74	0	2-cell
lightcyan	244	63	20	2-cell
darkmagenta	114	38	30	2-cell
darkturquoise	140	42	30	2-cell
paleturquoise	115	32	30	2-cell
purple	458	200	30	2-cell
salmon	308	111	30	2-cell
thistle	25	7	30	2-cell
green	1,046	389	70	4-cell, MZT
red	1,044	360	70	4-cell, MZT
yellow	1,156	400	70	4-cell, MZT
greenyellow	413	109	140	26-cell
blue	1513	440	160	Gastrulation
pink	812	278	180	Gastrulation
cyan	303	72	230	Gastrulation
turquoise	1906	520	290	Gastrulation
tan	338	83	340	Gastrulation-Bean
lightgreen	179	53	400	Bean, VE
magenta	763	214	400	Bean, VE
black*	970	321	60	4-cell, MZT (?)
brown*	1335	505	230	Gastrulation (?)

Table 2. Summary of co-expression modules of *Caenorhabditis elegans*. Total GO terms indicate the total number of terms that were found for the biological process category in topGO [Alexa et al. 2006]. Modules are ordered according to the first changepoint in minutes post-fertilization [mpf] in which their expression increases. Two exceptions are modules black and brown marked with an asterisk, their first changepoint was observed to be associated with a decrease in expression and this trend was observed across stages (see expression profiles in supplementary figures) for both modules. First cell division (2-cell) occurs between 0 to 50 minutes after fertilization. Four cell stage (4-cell) is observed between 50 to 100 mpf. MZT is the maternal-to-zygotic transition and occurs approximately between 70 to 90 mpf. The 26-cell stage is observed before gastrulation around 100-150 mpf. Gastrulation occurs from 150 to 330 mpf. Bean stage is the stage in which organogenesis begins, it starts at 360 mpf and ends at 400 mpf, during this stage there is another stage called ventral (VE) enclosure which is important in nematodes and it is observed around 365-375 mpf [Chisholm and Hsiao 2012; Levin et al. 2012]. Stages were obtained from [Altun and Hall 2019 WormAtlas; Lee et al. 2014; Tadros and Lipshitz 2009]

5.3 General results for orthogroup content between modules

Here, the results for module comparisons that searched for differences in orthogroup content are discussed. First a summary of results is presented, then the results with significant p-values for the Fisher's exact are discusssed with one example. Finally an example of two modules that contain important transcription factors and other genes is discussed. Table 3 summarizes the results for orthogroup content in both species.

Species	Modules	Co-expressed orthologs
C. elegans	22	1,721
D. melanogaster	22	2,002

Table 3. Summary of results for orthogroup content between modules. Co-expressed orthologs are members of 1,009 orthogroups (Ensembl gene families).

We used a curated set of 629 transcription factors from FlyBase (Gene Group: Transcription Factors available at `https://flybase.org/reports/FBgg0000745.html`) [Thurmond et al. 2019 FlyBase] to search for transcription factors in the *D. melanogaster* set of co-expressed orthologs. Only 56 of these known transcription factors were found in this set, corresponding approximately to 3% of the total number of co-expressed orthologs in this species and *9%* of the curated transcription factors. The estimated number of transcription factors in the *C. elegans* genome is 763 [Narasimhan et al. 2015]. A total of 63 *C. elegans* transcription factors were identified in the set of co-expressed orthologs, corresponding approximately to 4% of all co-expressed orthologs and to 8% of the known transcription factors. These transcription factors belong to only 50 orthogroups (5% of co-expressed orthogroups). Tables 5 and 4 summarizes the distribution of these transcription factors in *C. elegans* and *D. melanogaster* modules respectively.

Module	Stage	TFs
black	2	15
greenyellow	2	3
magenta	2	5
green	9	12
yellow	11	7
pink	11	1
salmon	14	1
turquoise*	2	6
blue*	11	6

Table 4. Distribution of orthologous factors in *D. melanogaster* modules. First row is the name of the module, the second row is the stage associated to the first changepoint in which the expression increased. There are two exceptions indicated by asterisks. Two exceptions are the modules turquoise and blue in which their first changepoint corresponds to a decrease in expression and a have a general trend to increase or decrease respectively. The third row contains the total number of orthologous transcription factors that were also detected in *C. elegans*.

Module	Stage	TFs
grey60	2-cell	2
paleturquoise	2-cell	3
purple	2-cell	4
green	4-cell	2
red	4-cell	4
yellow	4-cell	15
greenyellow	26-cell	3
blue	Gastrulation	6
pink	Gastrulation	10
turquoise	Gastrulation	3
tan	Gastrulation-Bean	1
magenta	Bean	1
black*	4-cell	4
brown*	Gastrulation	5

Table 5. Distribution of orthologous transcription factors in *C. elegans* modules. First row is the name of the module, the second row is the stage associated to the first changepoint in which the expression increased. Two exceptions are modules black and brown marked with an asterisks, their first changepoint was observed to be associated with a decrease in expression and this trend was observed across stages. The third row contains the total number of orthologous transcription factors that were also detected in *D. melanogaster*.

5.4 Results for the Fisher's exact test

When performing comparisons of orthogroup content, the total number of comparisons depend on the number of modules. Because the number of modules to compare is the same in both species, the total number of comparisons is $22^2 = 484$. It is possible that two modules do not have overlapping orthogroups and hence the test cannot be performed. This implies that only 219 statistical tests were actually performed. From this subset only 7 (3 % of all comparisons) were observed to be significant (p-value < 0.05, Bonferroni correction) for Fisher's exact test and are summarized in 6. All matrices that were used to perform the test are available in supplementary figures.

C. elegans module	Genes	*D. melanogaster* module	Genes	Orthogroups	p-value
yellow	13, 39	blue	36, 5	13	4.84e-07
salmon	13, 45	blue	47, 25	4	0.000484
grey60	3, 16	blue	24, 4	3	0.000968
yellow	87, 127	turquoise	103, 54	84	0.000968
black	27, 77	blue	66, 53	23	0.00484
purple	27, 57	blue	57, 23	14	0.0242
blue	88, 145	turquoise	100, 75	79	0.0484

Table 6. Results with significant p-values for Fisher's exact test. Genes indicates total orthologs in module, orthologs of the same orthogroup not detected in module. Orthogroups indicates the total number of orthogroups present in both modules.

Then in the seven comparisons with significant p-values the null hypothesis does not hold, and this means that the proportion of detected orthologs differs in these modules. Whereas for the other comparisons the observed proportions are independent and the null hypothesis cannot be rejected. Because this statistical test did not consider relationships between genes within orthogroups it cannot be used as a method to detect differences in ortholog content.

5.5 Modules with high content of orthologous transcription factors

C. elegans yellow module and *D. melanogaster* black module that were observed to have similar changepoint patterns. These modules have the largest intersection of transcription factors across all modules (9 in black module and 7 in yellow module that are members of 7 transcription factor families). The first module that will be described is the *D. melanogaster* black module which contains 286 genes. First it was observed that it contained the Hox cluster genes *Ubx, Abd-B, Antp, abd-A, Scr*, while the other genes of the cluster *pb* and *Dfd* were assigned to the grey module, and *lab* was not detected in any module. Additionally it contains other genes such as *Wnt4* and its receptor *fz2* with diverse roles during development, *elav, ey, sv, toy* (eye development), *Nrg, Sema-1a, Oli, tup, robo, grn, zfh1, Fas2, NetB* (nervous system development) [Thurmond et al. 2019 FlyBase]. Figure 23 shows 25 of the 221 GO terms for biological functions that were associated to this module. *C. elegans* yellow module is also enriched for developmental genes. Examples include *ceh-10, cfz-2, cwn-2, egl-46, gpn-1, hch-1, ina-1, kal-1*, *unc-39* (with roles in neuron migration, GO:0001764, p-value = 0.00032), *efn-4, mnm-2, ebax-1, pry-1, pxn-2, unc-130, spon-1, zag-1, zig-4, ddr-1* (axon guidance, GO:0007411, p-value=0.01601), *ceh-20, let-381, mab-5* (mesodermal cell fate specification, GO:0048337, p-value = 0.00392). However its GO enrichment has more content of genes that are not direct regulators of development such as genes associated to oxidation-reduction process (49 genes, GO:0055114, pvalue = 2.5e-06) or mitochondrial translation (9 genes, GO:0032543, p-value = 7.8e-06). Figure 24 shows the GO enrichment for biological functions in this module, only the first 25 of 383 terms are shown. We selected the *C. elegans* yellow module, because it was the one that had more intersecting genes with *D. melanogaster* black module. A total of 24 orthogroups were detected in these modules encompassing 26 and 24 genes in *D. melanogaster* and *C. elegans* respectively. Table 7 contains the genes that are shared between *C. elegans* and *D. melanogaster* modules along wiht their Ensembl gene family (or orthogroup) identifier, and a brief description of their function in both species, is this function is reported as conserved it is also indicated, function were verified in the GO enrichment as well as FlyBase [Thurmond et al. 2019 FlyBase] and WormBase [Lee et al. 2017 WormBase].

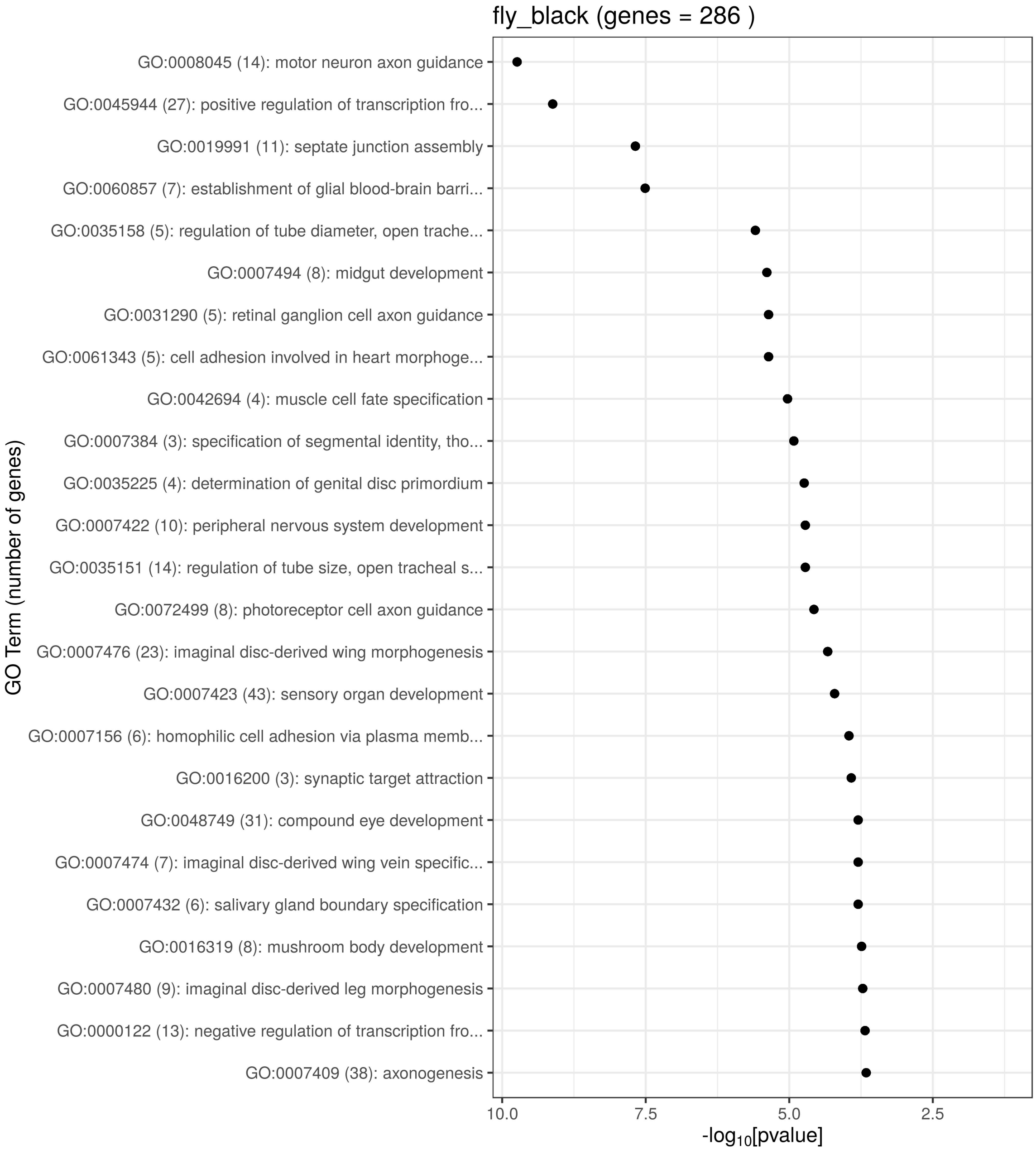


Figure 23. GO enrichment analysis of biological functions for *D. melanogaster* black module containing 286 genes. Only 25 of a total of 221 terms are shown.

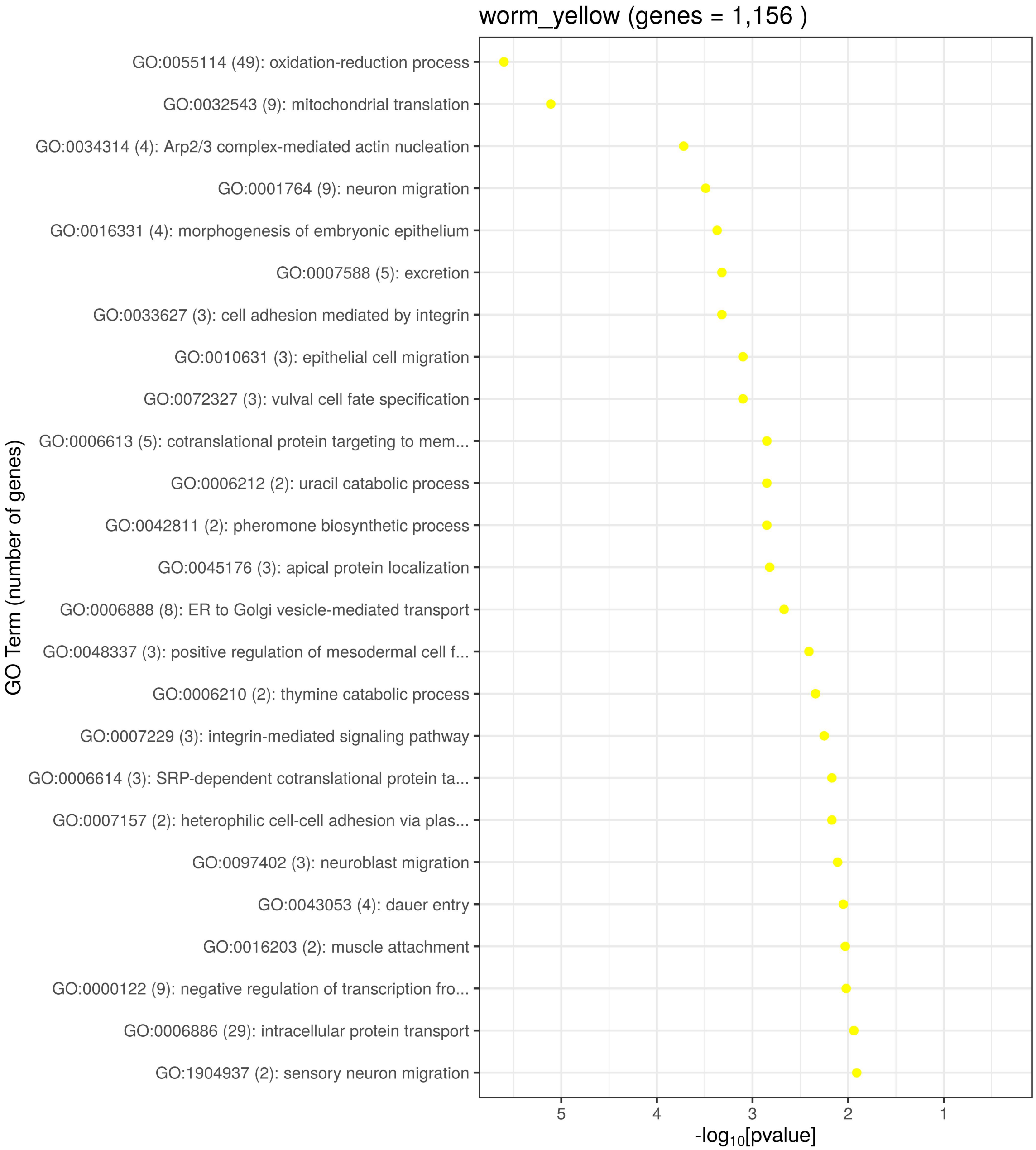


Figure 24. GO enrichment analysis of biological functions for *C. elegans* yellow module containing 1,156 genes (coding and non-coding). Only 25 of a total of 383 terms are shown.

Ensembl gene family	*D. melanogaster*	*C. elegans*	Functions
PTHR11211_SF2	*mirr, caup*	*irx-1*	Homeobox TFs, (1) eye development and others, (2) various
PTHR11559_SF157	*Gli*	*ges-1*	Conserved carboxylesterases
PTHR13803	*Sec24CD*	*sec-24.1*	Conserved subunit of coat protein complex II (COPII) in vesicles
PTHR11309_SF82	*fz2*	*cfz-2*	Conserved *frizzled2* receptor of Wnt
PTHR24049	*uif*	*crb-1*	Transmembrane protein. (1) Notch signaling, (2) epithelial structure
PTHR11389_SF345	*scrt, Kah*	*ces-1*	Conserved C2H2 zinc-finger TFs, neural commitment
PTHR16064	CG16952	C39F7.5	Uncharacterized
PTHR24329_SF265	*Poxn*	*pax-1*	Paired-box TFs, (1) sensory organ specification, (2) uncharacterized
PTHR10390_SF17	*Optix*	*ceh-32*	SIX TFs, (1) eye development, (2) head morphogenesis
PTHR24204	*tup*	*lim-7*	Conserved LIM-homeobox TFs, neuron differentiation
PTHR10807_SF54	*mtm*	*mtm-1*	Conserved Phosphatidylinositol-3-phosphatases
PTHR12309_SF5	*Sec61gamma*	*emo-1*	Conserved subunit of Sec61-gamma, protein transport
PTHR10822	*dally*	*gpn-1*	Conserved proteoglycan membrane receptors of growth factors
PTHR24326_SF81	*Scr*	mab-5	Homeobox TFs, (1) anterior thorax identity, (2) neuron cell fate
TF614188	*Sec61beta*	Y38F2AR.9	Conserved subunit of Sec61-beta, protein transport
PTHR10827	*scf*	*calu-2*	Conserved calcium-binding homologs, regulation of DNA topoisomerase
PTHR12924	*l(1)G0320*	*trap-1*	Conserved translocon-associated proteins, protein transport
PTHR24391	*zfh1*	*zag-1*	Conserved C2H2 zinc-finger TFs, motor neuron axon guidance
PTHR12990	*Manf*	*manf-1*	Homologues of mesencephalic astrocyte-derived neurotrophic factors
PTHR10529_SF210	*mgl*	F14B4.1	Homologues of epidermal growth factors (EGF) receptors
PTHR12861	*SsRbeta*	*trap-2*	Conserved translocon-associated proteins, protein transport
PTHR10656	*mab-21*	*mab-21*	Homologues of MAB21L1 (nervous system development in vertebrates)
PTHR12587	*Liprin-alpha*	*syd-2*	Conserved, axon guidance and synapse formation
PTHR12877	CG43658	*osg-1*	Conserved RhoGEFs (Guanine Nucleotide Exchange Factors)

Table 7. Orthogroups (Ensembl ID) and the associated orthologs detected in *D. melanogaster* black module and *C. elegans* yellow module. Function were verified in the GO enrichment as well as FlyBase[Thurmond et al. 2019 FlyBase] and WormBase [Lee et al. 2017 WormBase] and only a short summary is shown. Functions descriptions in which there is evidence for conserved function are mentioned, other wise (1) means function in *D. melanogaster* and (2) a different function in *C. elegans*. Homologues means that are associated to protein family with known functions in vertebrates (*Mus musculus* and *Homo sapiens*).

In order to have a general view of all known functions for shared genes between these modules, a biological function GO enrichment analysis for the shared genes was performed for each species set (26 genes in *D. melanogaster* black module and 24 in *C. elegans* yellow module). The following two figures show the results for this GO enrichment analysis in both species.

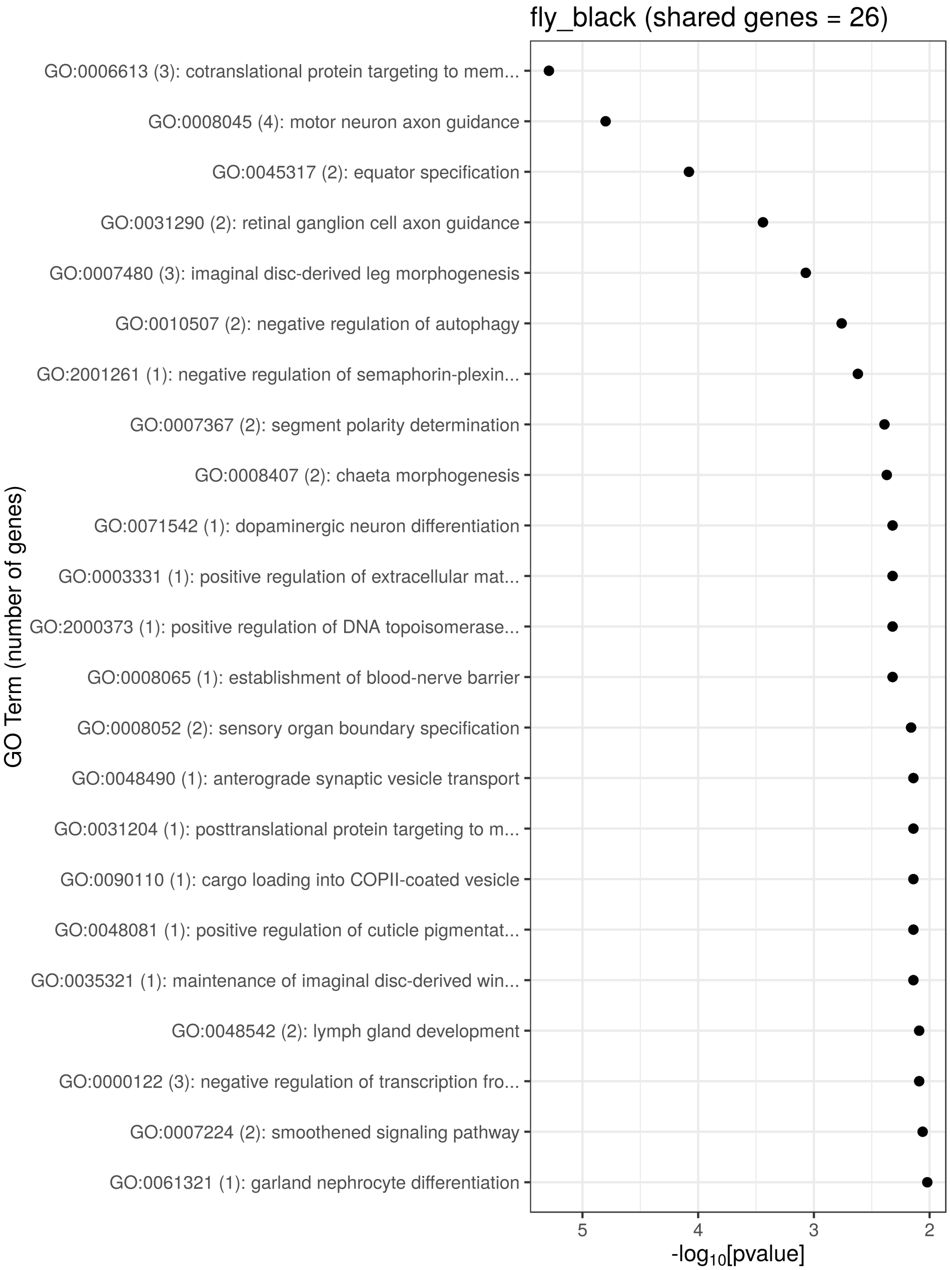


Figure 25. GO enrichment results for biological functions of genes in *D. melanogaster* black module that intersected with *C. elegans* yellow module. For the 26 shared genes a total of 23 terms were associated. First term GO:0006613 is cotranslational protein targeting to membrane (p-value = 5.1e-06) and includes the genes *Sec61beta*, *SsRbeta*, *l(1)G0320*.

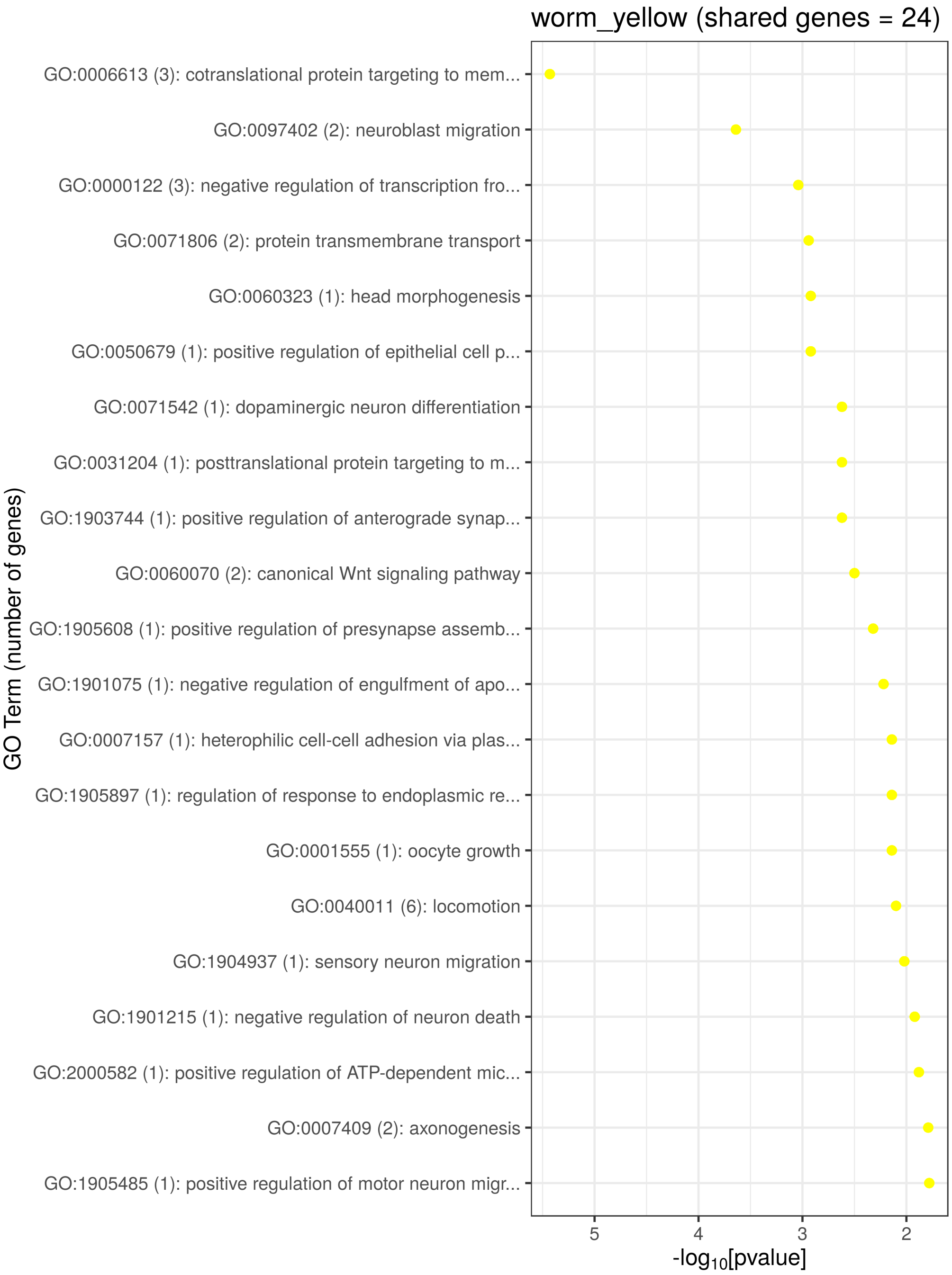


Figure 26. GO enrichment results for biological functions of genes in *C. elegans* yellow module that intersected with *D. melanogaster* black module. For the 24 shared genes a total of 21 terms were associated. First term GO:0006613 is co-translational protein targeting to membrane (p-value = 3.7e-06) and includes the genes *Y38F2AR.9*, *trap-2*, *trap-1* (orthologs of *Sec61beta*, *SsRbeta*, *l(1)G0320* respectively).

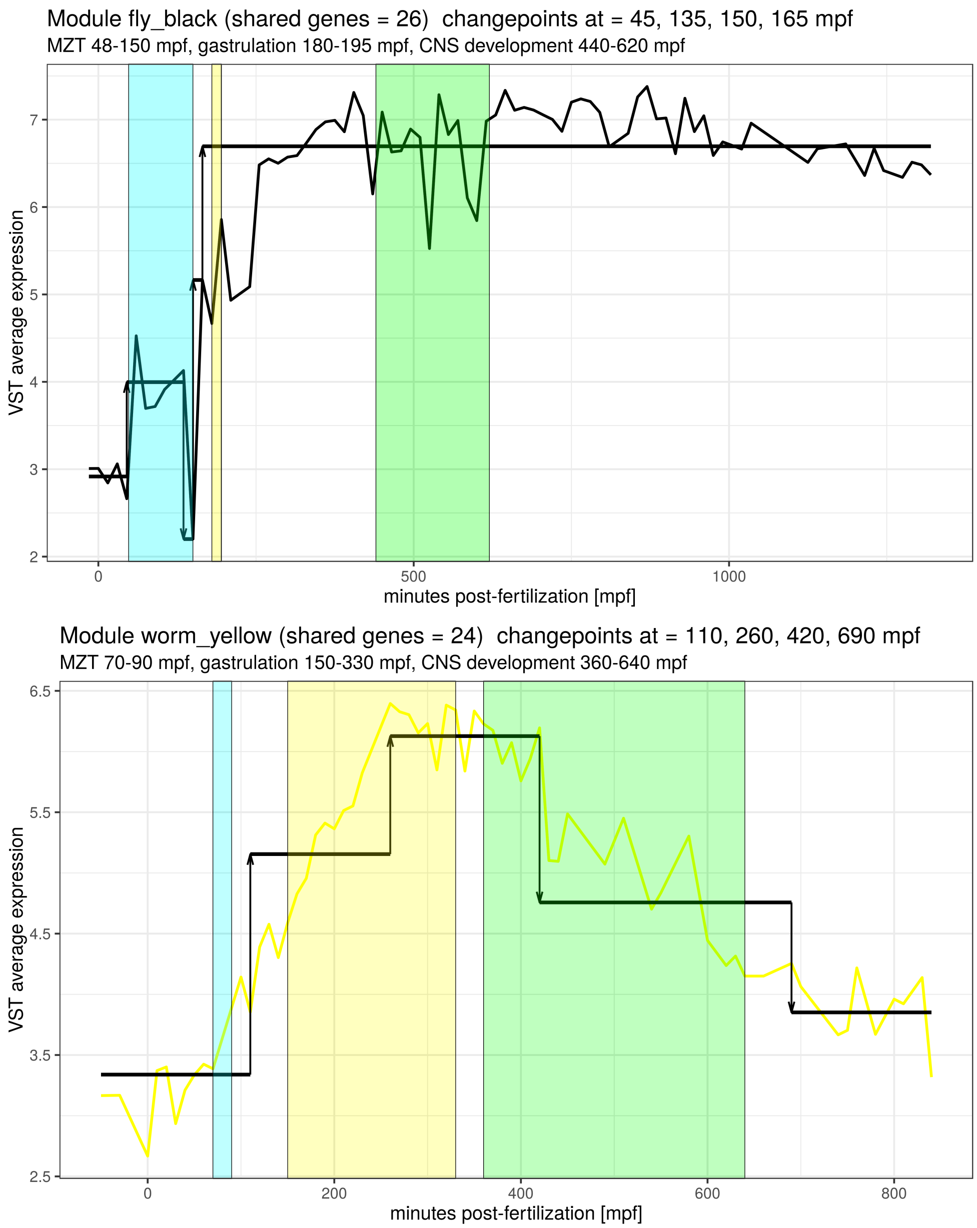


Figure 27. Expression profiles of orthologs detected in *D. melanogaster* black module and *C. elegans* yellow module. MZT maternal-to-zygotic transition, CNS central nervous system. In *D. melanogaster* nervous system development starts during the germband retraction stages 11-12 (approximately 440 to 620 minutes post-fertilization) [Campos-Ortega and Hartenstein 1997]. *C. elegans* nervous system development is slightly different, while most lineages are specified very early before or during gastrulation, axon formation and ventral nerve cord development begins approximately at bean stage (360-400 minutes post fertilization) and continues during the elongation phase (400-640 minutes post-fertilization) [Altun and Hall 2019 WormAtlas]. In both species other events such as terminal neuron differentiation in the peripherical nervous system (e.g. sensory organs or gonad motor neurons) occur during larval stages.

Next we searched for similarities in the expression profiles for the shared genes between these modules. Figure 27 shows the two expression profiles for each module indicating only the mean expression profiles of shared genes along with their estimated changepoints. In each plot three intervals are indicated, the first is for the maternal-to-zygotic transition, the second for gastrulation and the last one for the interval in which central nervous system development begins. The only important pattern that is observed is that in both cases expressions increases during gastrulation. However in *D. melanogaster* this high expression is mantained across later stages while in *C. elegans* it decreases steadily until hatching.

By checking the table of the 24 orthogroups detected in both modules it was clear that they could be subdivided into five categories depending of the type of protein that their member genes encode: (1) transcription factors (7, e.g. *tup/lim-7*), (2) ligand/receptors (7, e.g. *fz2/cfz-2*), (3) highly-conserved proteins involved in basic cellular processes (6, e.g. *SsRbeta/trap-2*), (4) enzymes (3, e.g. *Gli/ges-1*), and (5) only one uncharacterized group (CG16952/C39F7.5). Most of these genes can have various roles as shown by the GO enrichment results. However an interesting trend that was observed is that the transcription factors are known to participate in nervous system development. For example *zfh1/zag-1* activate genes involved in motor neuron axon guidance and they might have unknown targets [Chisholm et al. 2016; Clark and Chiu 2003; Zarin et al. 2014; Zarin and Labrador 2019]. *D. melanogaster tup* which regulates motor neuron identity [Thurmond et al. 2019 FlyBase], while mutants of its ortholog *lim-7* result in L1 larval lethality in *C. elegans*, it is unclear if it has a more specific function in nervous system development [Lee et al. 2017 WormBase]. An important observation is that one of the many functions of the transcription factor *vvl* is to activate the neuron-specific gene encoding dopa decarboxylase [Thurmond et al. 2019 FlyBase], and this gene was also detected in *D. melanogaster* black module, its ortholog *ceh-6* is expressed in neurons and other ectodermal tissues but its specific functions are also unclear [Lee et al. 2017 WormBase], it was detected in *C. elegans* yellow module. While these two genes are orthologs they were not present in the Ensembl data used for the analysis. Other transcription factors that participate in axon guidance in *D. melanogaster* were detected in black module: *Oli*, *HGTX*, and *grn*. Components of signaling pathways involved in axon guidance were also observed in this module including *robo1*, *NetB*, *Sema1a*, *Wnt4/fz2*. However only *cfz-2* which is the ortholog of *fz2* was observed in *C. elegans* yellow module. Other interesting genes are *mab-21* orthologues which have various functions in nervous system development in echinoderms and vertebrates [Israel et al. 2016; Zerbino et al. 2018 Ensembl] as well *Manf/manf-1* [Zerbino et al. 2018 Ensembl]. While neuron specification and early axon growth occur during later stages in both species, specifically during germband retraction stages 11-12 (440-620 minutes post-fertilization) [Campos-Ortega and Hartenstein 1997] and bean stage/elongation phase in *C.elegans* (360-640 minutes post-fertilization) [Altun and Hall 2019 WormAtlas]. There were not many similarities between the expression profiles of shared genes between modules. In *D. melanogaster* the expression tends to be more stable after gastrulation and during germband retraction. This is different in *C. elegans*, in which the expression is high at the beginning of the bean stage but tends to decrease during the elongation phase. The estimated changepoints in *D. melanogaster* black module (45, 135, 150, 165 minutes post-fertilization) did not differ when considering all genes (286) or the genes intersecting with *C. elegans* yellow module (26). This was different for the *C. elegans* yellow module that has a total of 1,156 genes case the estimated changepoints were 70, 200, 420, and 690 minutes-post fertilization, but when these were estimated only for the subset of genes shared with the other module, the results were 110, 260, 420 and 690 minutes post-fertilization. In summary the expression profiles are different, but this does not neccessarily mean that orthologs co-expressed in the two modules are not involved in similar developmental processes. Based on the GO enrichment results is more probably that genes such as the transcription factors and components of signaling pathways are involved in several processes that can occur simultaneously in both species. While some of these processes are well characterized such as axon guidance, genes involved in this process that were detected in these modules increase their expression very early during the maternal-to-zygotic transition in *D. melanogaster* (45 minutes post-fertilization) and between the maternal-to-zygotic transition and gastrulation in *C. elegans* (110 minutes post-fertilization). These results suggest that they might have early activity and then participate in neurogenesis and axon guidance in later stages.

Finally we checked if the non-coding genes that were detected in these modules have any known functions. In *C. elegans* yellow module only two pseudogenes were detected, the first is *cyp-25A5* (cytochrome P450 family), and F40H6.6, there is no clear evidence for known developmental functions of these pseudogenes at least from WormBase [Lee et al. 2017 WormBase]. Five long non-coding RNAs (lncRNAs) were detected in *D. melanogaster* black module and are described in table 8 below. Two of these lncRNAs have known functions and the functions of the other three are unknown, all the information was retrieved from FlyBase [Thurmond et al. 2019 FlyBase].

Gene ID	Name	Chromosome	Length [kbp]	Function
FBgn0001234	*Hsromega*	3R	25.71	Increases its transcription in response to heat shock
FBgn0019660	*roX2*	X	5.368	Male-specific, dosage compensation
FBgn0263019	CR43314	2L	22.89	Unknown
FBgn0260722	CR42549	3R	7.703	Unknown
FBgn0050009	CR30009	2R	6.513	Unknown

Table 8. Long non-coding RNAs detected in *D. melanogaster* black module. All data were retrieved from FlyBase [Thurmond et al. 2019 FlyBase].

The expression dynamics of CR30009 has been recently characterized. It colocalizes with the glial marker *repo* (reversed polarity, PRD-homeobox TF, glial terminal differentiation) in brain and the ventral nerve cord during stages 9-12 (230-580 mpf) and 13 (580-620 mpf) [McCorkindale et al. 2019]. Figure 28 shows the colocalization of this lncRNA with the glial marker *repo* during stages 11-12 and 13 using

RNA-FISH. At least one of the neuron markers used in this study was detected in black module *elav*: (embryonic lethal abnormal vision, RNA-binding protein, neuron differentiation). The glial marker*repo* was detected in the *D. melanogaster* green module. An important remark is that the first changepoint of green module occurs during stage 9 (240 mpf), followed by other increase in expression during stage 12 (570 mpf), another during stage 13 (615 mpf), and a last increase at stage 14 (660 mpf). These results are compatible with the suggestion made by the authors that *repo* expression is independent of CR30009 expression, which could be begin earlier (during the maternal-to-zygotic transition approximately 45-165 mpf according to the changepoint results) [McCorkindale et al. 2019].

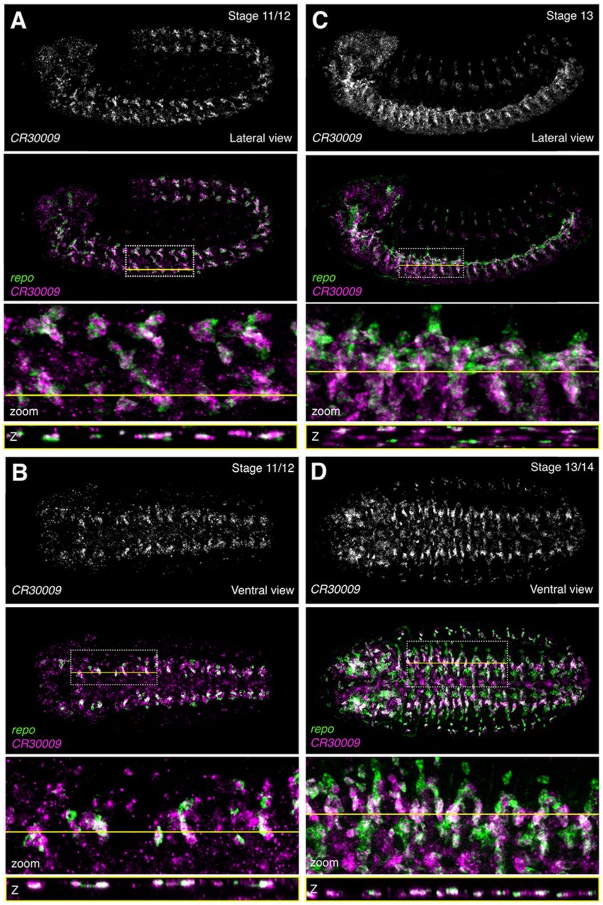


Figure 28. Colocalization of the lncRNA CR30009 with the glial marker *repo* in *D. melanogaster* embryos. CR30009 is marked with magenta and *repo* with green. All micrographs are oriented with the anterior side to the left and posterior to the right. Boxes indicate the region selected for zoom. CR30009 was detected in black module (first increase in expression around 45 minutes post-fertilization), and *repo* is expressed later in green module during stage 9 (240 minutes post-fertilization). Adapted from [McCorkindale et al. 2019].

6 Discussion

In this project we analyzed the co-expression of orthologous genes during the embryogenesis of *D. melanogaster* and *C. elegans*, without quantifying the differences between their expression profiles. Despite the large phylogenetic distance between these protostome species (587-543 Mya) [Rota-Stabelli et al. 2013], we found that a subset of orthologs was co-expressed in both species. From these subset we identified seven transcription factor orthogroups that are co-expresed in both species in a pattern that resembles the proposed inverse-hourglass model for transcriptome similarity across phyla [Levin et al. 2016]. This suggests that these transcription factors were components of the developmental regulatory program in the last common ancestor of these protostome species as reported in previous studies [Degnan et al. 2009; Erwin 2009; Friedrich 2015].

Overall we found that the percentage of co-expressed orthologs was low compared to the total number of genes with dynamic expression (29% in *D. melanogaster* and 13 % in *C. elegans*). This implies that the majority of genes with dynamic expression in both species (71 % in *D. melanogaster* and 87 % in *C. elegans*) are not classified as members of the same orthologous families, at least according to Ensembl. As a comparison, from the 3,019 single-copy orthologs compared by Kalinka et al. 2010 in six *Drosophila* species, 1,188 (39 %) were reported to be highly correlated in all species. Evidence from other species such as nematodes [Levin et al. 2012; Macchietto et al. 2017] and sea urchins [Gildor and Ben-Tabou de-Leon 2015; Israel et. al 2016] suggest that the low number of co-expressed orthologs is expected because of a large divergence time between these species, which is approximately 587-543 Mya [Rota-Stabelli et al. 2013].

Interestingly 50 transcription factor families are co-expressed in both organisms despite this large phylogenetic distance. Previous evidence suggests that orthologous transcription factors are co-expressed in early or late stages outside the phylotypic stage when comparing species of different phyla, and only genes of the homeobox family are co-expressed during this stage [Levin et al. 2016]. We observed this pattern of expression in the *D. melanogaster* black module and *C. elegans* yellow module which contain two groups of homeobox transcription factors, the first group includes *Scr* and *mab-5* which are members of the Hox cluster in *D. melanogaster* and *C. elegans* respectively. The second homeobox group detected in these modules includes *mirr, caup* in *D. melanogaster* and *irx-1* in *C. elegans.* Both modules increase their expression during the maternal-to-zygotic transition and are dynamically expressed during the phylotypic stage of each species: germband stages in *D. melanogaster* (stages 9-13, 230-620 minutes post-fertilization) and ventral enclosure in *C. elegans* (365-375 minutes post fertilization) as shown in figure 27.

Other zinc-finger, paired-box (Pax), sine oculis homeobox (SIX), and LIM-homeobox transcription factors are also members of this co-expression modules (see table 7). This suggests that the deployment of these transcription factors during the maternal-to-zygotic transition and their co-expression across all stages of embryogenesis was present at least in the common ancestor of ecdysozoans and that they have an expression profile compatible with the inverse-hourglass model. Additionally they also have similar expression during the phylotypic stages as previously reported for homeobox transcription factors [Levin et al. 2016].

6.1 Known developmental processes observed in modules

Important known patterns of developmental gene expression were observed in the co-expression modules. For example the *D. melanogaster* greenyellow module increases its expression near the maternal-to-zygotic transition at 45 mpf while the reported interval for this transition is 48 to 150 mpf, it contains also the gene *zelda* which is a regulator of the transition in this species [Lee et al. 2014; Palfy et al. 2017], and at least one gene *kruppel* which is an early-zygotic gene [De-Renzis et al. 2007]. Another example in *C. elegans* is related to *tbx-35* and *end-3* which are transcription factors that regulate mesoderm (MS lineage) and endoderm (E lineage) specification respectively [Maduro 2010; Owraghi et al. 2010] were assigned to the in the darkslateblue and grey60 modules respectively, these modules were observed to increase their expression at the 2-cell stage just before the division of the EMS progenitor [Rose and Gonczy 2014]. However, the transcription factors *med-1* and *med-2* that activate these transcription factors in the EMS lineage [Maduro 2010] were not detected in any module. Other important genes with conserved functions were not classified in co-expression modules. The most important example are the components of the Chordin/Tolloid/BMP pathway that specifies the identity of the dorsal-ventral regions in *D. melanogaster* and vertebrates. Genes involved in this pathway include *cv-2*, *dpp*, *tld*, *tsg*, *sog*, and *scw* [Bier and De Robertis 2015; De Robertis 2008], all of these genes were assigned to the grey module (could not be assigned to a co-expression module with dynamic expression). These examples illustrate the fact that even genes with conserved functions were not always assigned to co-expression modules, this is inevitable and implied that we had to analyze only a subset of co-expressed orthologs in each species. We also found non-coding genes in co-expression modules that could be associated to cell types using single-cell transcriptomic data.

6.2 Changepoints are associated to known embryonic stages

In both species we observed that the time points in which mean expression of co-expression modules increases (first changepoint) can be associated to specific stages. First changepoints in *C. elegans* map to only five stages: 2-cell, 4-cell, 26-cell, gastrulation, and bean stage (ventral enclosure). A similar result was observed in *D. melanogaster* in which these time points map to only eight stages: 2, 6-7 (gastrulation), 9 (germband elongation), 11 (parasegmentation), 13 (germband retraction), 14 (dorsal closure), 15, and 16 (ventral nerve shortening). This means that heterochronic shifts [Israel et al. 2016] in ortholog expression might occur frequently between these stages, at least for this dataset. The matrix for orthogroups detected in both species in which orthogroups can be detected between modules associated to very different stages in each species (orthogroup content between modules, supplementary figures 10.3), suggest that this is be the case. It would be necessary to quantify the extent of this heterochronic shifts and perform the analysis in other species to verify that the same pattern of association to stages is observed. Moreover this might facilitate to test hypothesis about heterochronic shifts in which the function of orthologs is conserved or lost. For the estimation of changepoints it will be useful to evaluate how to estimate an error interval for each changepoint considering that it is based in a maximum-likelihood algorithm [Killick et al. 2012], this might reduce errors in the association of co-expression modules to known stages of embryogenesis.

6.3 Scope and limitations of the analysis

Because each transcriptome was obtained from a whole embryo, we cannot say much of the tissues or cell types in which the genes in a module are expressed, unless information of known cell-type or tissue specific genes is used. Analysis of single-cell transcriptomic data could be more informative to understand gene regulation during development [Briggs et al. 2018; Sebé-Pedrós et al. 2018] and could be useful to determine if the non-coding genes detected in modules are cell-type specific. The variation of gene expression levels between species is also difficult to explain only from the results we generated because cell proportions might affect the global observed expression levels [Pantalacci et al. 2017]. Heterochronic shifts, or the event in which orthologs are expressed in different stages could indicate changes in function or differences in the timing of a developmental process [Israel et al. 2016]. We did not quantified this type of shifts, but it would be possible to design a pipeline to detect the occurence of these shifts based on the associaton of modules to specific stages in each species. By examining the *D. melanogaster* black module *C. elegans* yellow module we were able to identify genes that might have similar functions in nervous system development such as *Manf/manf-1* and *mab-21* orthologs. Orthologues of these genes participate in nervous system development in vertebrates (as reported in Ensembl for *Mus musculus*) [Zerbino et al. 2018 Ensembl], and *mab-21* is expressed in ectoderm in echinoderms [Israel et. al 2016]. Despite the large divergence time between *D. melanogaster* and *C. elegans* (587-543 Mya) [Rota-Stabelli et al. 2013] we found 1,009 orthogroups that are co-expressed during the embryogenesis of these species. From these co-expressed orthogroups we identified seven transcription groups detected in *D. melanogaster* black module and *C. elegans* yellow module that were observed to have an expression pattern that is compatible with the inverse-hourglass model for transcriptome similarity across phyla [Levin et al. 2016].

7 Conclusions

1. Despite the large divergence time between *D. melanogaster* and *C. elegans* (587-543 Mya) we found conserved co-expression of orthogroups between both species, including several transcription factor orthogroups.

2. Co-expression of orthologs during embryogenesis is low compared with the total number of genes with dynamic expression in both species.

3. Heterochronic shifts in ortholog expression between stages are common between these species.

4. The largest group of co-expressed orthologous transcription factors increases its expression during the maternal-to-zygotic transition in both species suggesting this stage is under tight regulation in both species.

8 Future directions

1. Performing an analysis using single-cell transcriptomic data of whole embryos to estimate cell state transitions would be useful to understand the sequential organization of co-expression modules.

2. A better method to detect differences and similarities in ortholog content between modules could be implemented. This method should consider the relationships among genes in each orthogroup.

3. The dataset generated in this project could be used to further quantify heterochronic shifts in gene expression between species.

4. Search for interactions between genes in the same module could be useful to understand regulatory processes in each stage.

5. Additional species could be included in the analysis to verify if the observed co-expression patterns occur at different phylogenetic distances.

6. Understanding why non-coding genes are also co-expressed across embryogenesis and if they are cell-type specific could be important to determine if they may have functions during animal embryogenesis.

9 References

1. Aguinaldo, A. M. A. et al. Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387, 489–493 (1997).

2. Alexa, A., Rahnenfuhrer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 22, 1600–1607 (2006).

3. Altun, Z. F. & Hall, D. H. Hermaphrodite Introduction. Handbook of *C. elegans* Anatomy. In WormAtlas. (2019). Available at: https://www.wormatlas.org/hermaphrodite/hermaphroditehomepage.htm. (Accessed: 19th August 2019).

4. Babonis, L. S. & Martindale, M. Q. Phylogenetic evidence for the modular evolution of metazoan signalling pathways. Phil. Trans. R. Soc. B 372, 20150477 (2017).

5. Baroux, C., Autran, D., Gillmor, C. S., Grimanelli, D. & Grossniklaus, U. The Maternal to Zygotic Transition in Animals and Plants. Cold Spring Harbor Symposia on Quantitative Biology 73, 89–100 (2008).

6. Bier, E. & De Robertis, E. M. BMP gradients: A paradigm for morphogen-mediated developmental patterning. Science 348, aaa5838–aaa5838 (2015).

7. Briggs, J. A. et al. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. Science 360, eaar5780 (2018).

8. Brummel, T., Ching, A., Seroude, L., Simon, A. F. & Benzer, S. *Drosophila* lifespan enhancement by exogenous bacteria. Proceedings of the National Academy of Sciences 101, 12974–12979 (2004).

9. Campos-Ortega, J. A. & Hartenstein, V. The Embryonic Development of Drosophila Melanogaster. (Springer Berlin / Heidelberg, 2013).

10. Chisholm, A. D. & Hsiao, T. I. The *Caenorhabditis elegans* epidermis as a model skin. I: development, patterning, and growth: *C. elegans* skin I: development and pattern. WIREs Dev Biol 1, 861–878 (2012).

11. Chisholm, A. D., Hutter, H., Jin, Y. & Wadsworth, W. G. The Genetics of Axon Guidance and Axon Regeneration in *Caenorhabditis elegans*. Genetics 204, 849–882 (2016).

12. Clark, S. G. and Chiu, C. *C. elegans* ZAG-1, a Zn-finger-homeodomain protein, regulates axonal development and neuronal differentiation. Development 130, 3781–3794 (2003).

13. Comte, A., Roux, J. & Robinson-Rechavi, M. Molecular signaling in zebrafish development and the vertebrate phylotypic period: Molecular signaling and the phylotypic period. Evolution & Development 12, 144–156 (2010).

14. De Robertis, E. M. Evo-Devo: Variations on Ancestral Themes. Cell 132, 185–195 (2008).

15. Degnan, B. M., Vervoort, M., Larroux, C. & Richards, G. S. Early evolution of metazoan transcription factors. Current Opinion in Genetics & Development 19, 591–599 (2009).

16. Domazet-Lošo, T. & Tautz, D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. Nature 468, 815–818 (2010).

17. Drost, H.-G., Janitza, P., Grosse, I. & Quint, M. Cross-kingdom comparison of the developmental hourglass. Current Opinion in Genetics & Development 45, 69–75 (2017).

18. Durbin, B. P., Hardin, J. S., Hawkins, D. M. & Rocke, D. M. A variance-stabilizing transformation for gene-expression microarray data. Bioinformatics 18, S105–S110 (2002).

19. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16, 157 (2015).

20. Erwin, D. H. Early origin of the bilaterian developmental toolkit. Phil. Trans. R. Soc. B 364, 2253–2261 (2009).

21. Erwin, D. H. Early metazoan life: divergence, environment and ecology. Phil. Trans. R. Soc. B 370, 20150036 (2015).

22. Friedrich, M. Evo-Devo gene toolkit update: at least seven Pax transcription factor subfamilies in the last common ancestor of bilaterian animals: Bilaterian Pax gene subfamilies. Evolution & Development 17, 255–257 (2015).

23. Genikhovich, G. & Technau, U. On the evolution of bilaterality. Development 144, 3392–3404 (2017).

24. Gilbert, S. F. & Barresi, M. J. F. Developmental biology. (Sinauer Associates, Inc, 2016).

25. Gildor, T. & Ben-Tabou de-Leon, S. Comparative Study of Regulatory Circuits in Two Sea Urchin Species Reveals Tight Control of Timing and High Conservation of Expression Dynamics. PLoS Genet 11, e1005435 (2015).

26. Giribet, G. & Edgecombe, G. D. Current Understanding of Ecdysozoa and its Internal Phylogenetic Relationships. Integrative and Comparative Biology 57, 455–466 (2017).

27. Hahsler, M., Piekenbrock, M. & Doran, D. dbscan: Fast Density-based Clustering with R. 28

28. Hales, K. G., Korey, C. A., Larracuente, A. M. & Roberts, D. M. Genetics on the Fly: A Primer on the Drosophila Model System. Genetics 201, 815–842 (2015).

29. Hall, B. K. Evolutionary Developmental Biology. (Springer Netherlands, 1999).

30. Hashimshony, T., Feder, M., Levin, M., Hall, B. K. & Yanai, I. Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. Nature 519, 219–222 (2015).

31. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. Cell Reports 2, 666–673 (2012).

32. Herman, M. Hermaphrodite cell-fate specification. WormBook (2006). doi:10.1895/wormbook.1.39.1

33. Heyn, P. et al. The Earliest Transcribed Zygotic Genes Are Short, Newly Evolved, and Different across Species. Cell Reports 6, 285–292 (2014).

34. Horner, V. L. & Wolfner, M. F. Transitioning from egg to embryo: Triggers and mechanisms of egg activation. Dev. Dyn. 237, 527–544 (2008).

35. Irie, N. & Kuratani, S. The developmental hourglass model: a predictor of the basic body plan? Development 141, 4649–4655 (2014).

36. Irie, N. & Kuratani, S. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. Nat Commun 2, 248 (2011).

37. Israel, J. W. et al. Comparative Developmental Transcriptomics Reveals Rewiring of a Highly Conserved Gene Regulatory Network during a Major Life History Switch in the Sea Urchin Genus *Heliocidaris*. PLoS Biol 14, e1002391 (2016).

38. Jukam, D., Shariati, S. A. M. & Skotheim, J. M. Zygotic Genome Activation in Vertebrates. Developmental Cell 42, 316–332 (2017).

39. Kalinka, A. T. & Tomancak, P. The evolution of early animal embryos: conservation or divergence? Trends in Ecology & Evolution 27, 385–393 (2012).

40. Kalinka, A. T. et al. Gene expression divergence recapitulates the developmental hourglass model. Nature 468, 811–814 (2010).

41. Kaneuchi, T. et al. Calcium waves occur as Drosophila oocytes activate. Proc Natl Acad Sci USA 112, 791–796 (2015).

42. Kassambara, A. Practical guide to cluster analysis in R: unsupervised machine learning. (STHDA, 2017).

43. Killick, R., Fearnhead, P. & Eckley, I. A. Optimal Detection of Changepoints With a Linear Computational Cost. Journal of the American Statistical Association 107, 1590–1598 (2012).

44. Langfelder, P. & Horvath, S. Eigengene networks for studying the relationships between co-expression modules. BMC Syst Biol 1, 54 (2007).

45. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559 (2008).

46. Langfelder, P. & Horvath, S. Fast R Functions for Robust Correlations and Hierarchical Clustering. J. Stat. Soft. 46, (2012).

47. Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is My Network Module Preserved and Reproducible? PLoS Comput Biol 7, e1001057 (2011).

48. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 24, 719–720 (2008).

49. Langley, A. R., Smith, J. C., Stemple, D. L. & Harvey, S. A. New insights into the maternal to zygotic transition. Development 141, 3834–3841 (2014).

50. Lee, M. T., Bonneau, A. R. & Giraldez, A. J. Zygotic Genome Activation During the Maternal-to-Zygotic Transition. Annu. Rev. Cell Dev. Biol. 30, 581–613 (2014).

51. Lee, R. Y. N. et al. WormBase 2017: molting into a new stage. Nucleic Acids Res 46, D869–D874 (2018).

52. Lemons, D. Genomic Evolution of Hox Gene Clusters. Science 313, 1918–1922 (2006).

53. Levin, M. et al. The mid-developmental transition and the evolution of animal body plans. Nature 531, 637–641 (2016).

54. Levin, M., Hashimshony, T., Wagner, F. & Yanai, I. Developmental Milestones Punctuate Gene Expression in the *Caenorhabditis* Embryo. Developmental Cell 22, 1101–1108 (2012).

55. Li, L. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. Genome Research 13, 2178–2189 (2003).

56. Linford, N. J., Bilgir, C., Ro, J. & Pletcher, S. D. Measurement of Lifespan in Drosophila melanogaster. JoVE 50068 (2013). doi:10.3791/50068

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

58. Macchietto, M. et al. Comparative Transcriptomics of Steinernema and *Caenorhabditis* Single Embryos Reveals Orthologous Gene Expression Convergence during Late Embryogenesis. Genome Biology and Evolution 9, 2681–2696 (2017).

59. Maduro, M. F. Cell fate specification in the *C. elegans* embryo. Dev. Dyn. (2010). doi:10.1002/dvdy.22233

60. Martindale, M. Q. The evolution of metazoan axial properties. Nat Rev Genet 6, 917–927 (2005).

61. Martínez, D. E., Bridge, D., Masuda-Nakagawa, L. M. & Cartwright, P. Cnidarian homeoboxes and the zootype. Nature 393, 748–749 (1998).

62. McCorkindale, A. L. et al. A gene expression atlas of embryonic neurogenesis in *Drosophila* reveals complex spatiotemporal regulation of lncRNAs. Development 146, dev175265 (2019).

63. McDonald, J. H. Handbook of biological statistics. (Sparky House Publishing, 2014).

64. Murtagh, F. A Survey of Recent Advances in Hierarchical Clustering Algorithms. The Computer Journal 26, 354–359 (1983).

65. Murtagh, F. & Contreras, P. Algorithms for hierarchical clustering: an overview: Algorithms for hierarchical clustering. WIREs Data Mining Knowl Discov 2, 86–97 (2012).

66. Narasimhan, K. et al. Mapping and analysis of*Caenorhabditis elegans* transcription factor sequence specificities. eLife 4, e06967 (2015).

67. Owens, N. D. L. et al. Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. Cell Reports 14, 632–647 (2016).

68. Owraghi, M., Broitman-Maduro, G., Luu, T., Roberson, H. & Maduro, M. F. Roles of the Wnt effector POP-1/TCF in the C. elegans endomesoderm specification gene network. Developmental Biology 340, 209–221 (2010).

69. Pálfy, M., Joseph, S. R. & Vastenhouw, N. L. The timing of zygotic genome activation. Current Opinion in Genetics & Development 43, 53–60 (2017).

70. Pantalacci, S. et al. Transcriptomic signatures shaped by cell proportions shed light on comparative developmental biology. Genome Biol 18, 29 (2017).

71. Piasecka, B., Lichocki, P., Moretti, S., Bergmann, S. & Robinson-Rechavi, M. The Hourglass and the Early Conservation Models—Co-Existing Patterns of Developmental Constraints in Vertebrates. PLoS Genet 9, e1003476 (2013).

72. Richardson, M. K. et al. There is no highly conserved embryonic stage in the vertebrates: implications for current theories of evolution and development. Anatomy and Embryology 196, 91–106 (1997).

73. Rose, L. & Gonczy, P. Polarity establishment, asymmetric division and segregation of fate determinants in early C. elegans embryos. WormBook 1–43 (2014). doi:10.1895/wormbook.1.30.2

74. Rota-Stabelli, O., Daley, A. C. & Pisani, D. Molecular Timetrees Reveal a Cambrian Colonization of Land and a New Scenario for Ecdysozoan Evolution. Current Biology 23, 392–398 (2013).

75. Roux, J. & Robinson-Rechavi, M. Developmental Constraints on Vertebrate Genome Evolution. PLoS Genet 4, e1000311 (2008).

76. Runft, L. L., Jaffe, L. A. & Mehlmann, L. M. Egg Activation at Fertilization: Where It All Begins. Developmental Biology 245, 237–254 (2002).

77. Ryan, J. F. & Baxevanis, A. D. Hox, Wnt, and the evolution of the primary body axis: insights from the early-divergent phyla. Biol Direct 2, 37 (2007).

78. Schep, A. N. & Adryan, B. A Comparative Analysis of Transcription Factor Expression during Metazoan Embryonic Development. PLoS ONE 8, e66826 (2013).

79. Schierwater, B. & Desalle, R. Current problems with the zootype and the early evolution of Hox genes. J. Exp. Zool. 291, 169–174 (2001).

80. Schumann, I., Kenny, N., Hui, J., Hering, L. & Mayer, G. Halloween genes in panarthropods and the evolution of the early moulting pathway in Ecdysozoa. R. Soc. open sci. 5, 180888 (2018).

81. Sebé-Pedrós, A. et al. Early metazoan cell type diversity and the evolution of multicellular gene regulation. Nat Ecol Evol 2, 1176–1188 (2018).

82. Slack, J. M., Holland, P. W. & Graham, C. F. The zootype and the phylotypic stage. Nature 361, 490–492 (1993).

83. Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts. Development 136, 3033–3042 (2009).

84. Takayama, J. & Onami, S. The Sperm TRP-3 Channel Mediates the Onset of a $Ca2+$ Wave in the Fertilized C. elegans Oocyte. Cell Reports 15, 625–637 (2016).

85. Telford, M. J., Budd, G. E. & Philippe, H. Phylogenomic Insights into Animal Evolution. Current Biology 25, R876–R887 (2015).

86. Thurmond, J. et al. FlyBase 2.0: the next generation. Nucleic Acids Res 47, D759–D765 (2019).

87. Wang, D. et al. Origin of ecdysis: fossil evidence from 535-million-year-old scalidophoran worms. Proc. R. Soc. B 286, 20190791 (2019).

88. Wolpert, L. Gastrulation and the evolution of development. Dev. Suppl. 7–13 (1992).

89. Yanai, I. Development and Evolution through the Lens of Global Gene Regulation. Trends in Genetics 34, 11–20 (2018).

90. Yanai, I., Peshkin, L., Jorgensen, P. & Kirschner, M. W. Mapping Gene Expression in Two Xenopus Species: Evolutionary Constraints and Developmental Flexibility. Developmental Cell 20, 483–496 (2011).

91. Zarin, A. A. et al. A Transcription Factor Network Coordinates Attraction, Repulsion, and Adhesion Combinatorially to Control Motor Axon Pathway Selection. Neuron 81, 1297–1311 (2014).

92. Zarin, A.-A. & Labrador, J.-P. Motor axon guidance in Drosophila. Seminars in Cell & Developmental Biology 85, 36–47 (2019).

93. Zerbino, D. R. et al. Ensembl 2018. Nucleic Acids Research 46, D754–D761 (2018).

94. Zhang, B. & Horvath, S. A General Framework for Weighted Gene Co-Expression Network Analysis. Statistical Applications in Genetics and Molecular Biology 4, (2005).

10 Supplementary figures and tables

1. Supplementary figures for co-expression modules. Expression profiles for each module are shown along with their biological function GO enrichment results for the first ten terms. Figures are ordered based on the stage in which their first changepoint was detected. Modules that had a different expression pattern in which their first changepoint was associated with a decrease in expression and that have different trends in expression across stages were classified in other group and are called modules with unique expression profiles. Grey module and its biological functions GO enrichment is presented at the end of all other expression groups. The final figures represent the number of coding and non-coding genes with unknown functions or poor characterization data according to Ensembl database, gene types are the same as in the results section [Zerbino et al. 2018]. *C. elegans* stages were obtained from [Altun and Hall 2019 WormAtlas; Lee et al. 2014; Tadros and Lipshitz 2009]. Stages for *D. melanogaster* are adapted from [Campos-Ortega and Hartenstein 1997; Hales et al. 2015; Lee et al. 2014; Palfy et al. 2017; Tadros and Lipshitz 2009]. MZT is the maternal-to-zygotic transition and occurs approximately from 48 to 150 minutes post-fertilization in *D. melanogaster* and 70 to 90 minutes-post fertilization in *C. elegans* if a set of modules increase their expression during it is indicated as MZT after the name of the stage. Minutes post-fertilization is abbreviated as mpf.

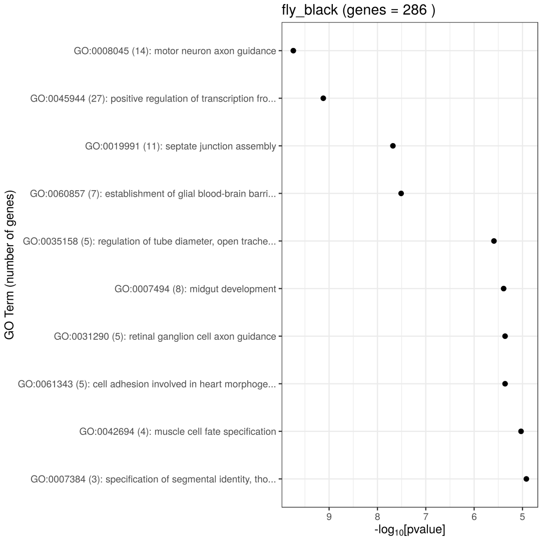
2. Supplementary figures for orthogroup content between modules. First matrix shows the total number of orthogroups that intersected between *C. elegans* and *D. melanogaster* modules. Second matrix is the total number of intersected genes between modules, in each cell the top number is the total number of genes that intersected for the *D. melanogaster* module and the number below represents the same for the *C. elegans module.* Third matrix represents the total number of members in the orthogroups that were not detected in modules. The last matrix is the Fisher's exact test that uses the counts in the shared genes and absent genes matrices to evaluate if the content of genes in each module is significant compared to the actual number of members in the shared orthogroups.

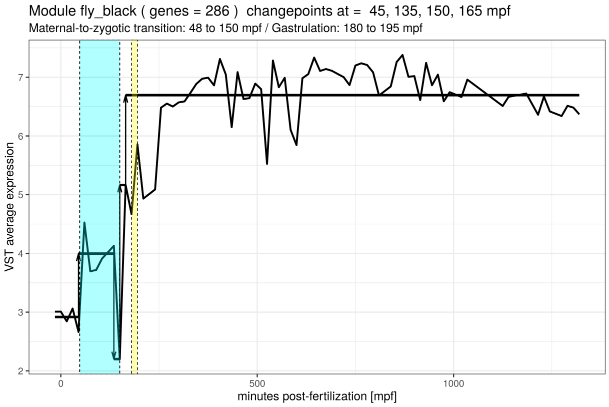
3. Supplementary figures and tables for transcription factor content between modules. The total number of transcription factor families that are shared between modules is shown in the first matrix. Second matrix are the total number of transcription factors shared between modules, in each cell the first number corresponds to the number of transcription factors in *D. melanogaster* and the number below is the total number of intersected transcription factors in C. elegans. Only one dataset of 629 transcription factors of *D. melanogaster* was used in the analysis, and was downloaded from FlyBase (available at `https://flybase.org/reports/FBgg0000745.html`). There are two tables at the end the first is the list of shared transcription factors in *C. elegans* modules, the second table represents the same but for *D. melanogaster* modules. In summary only 368 transcription factors were detected in *D. melanogaster* modules, and only 56 intersected with 63 transcription factors of *C. elegans*, these are members of 50 Ensembl gene families out of a total 1,009 families that intersected between modules. Shared transcription factors means orthologous transcription factors that are members of the same Ensembl gene family present in the other species modules. Asterisk in a module indicates that the its expression profile was different from the other modules as described in the figures of expression profiles.

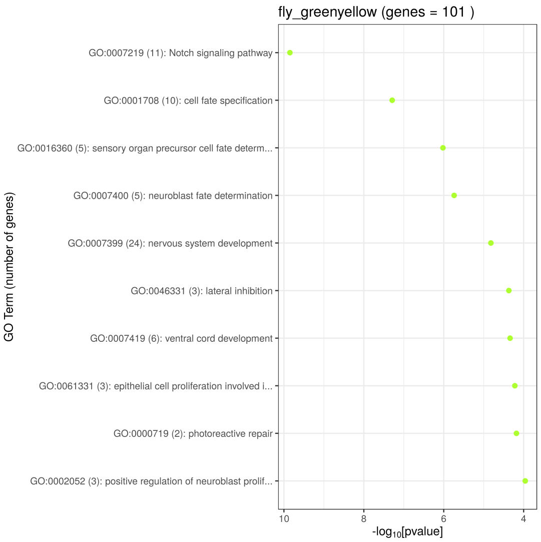
4. A repository containing all the R scripts used in this project can be found at: `https://bitbucket.org/alxndrdiaz/ccadt_final/src/master/`

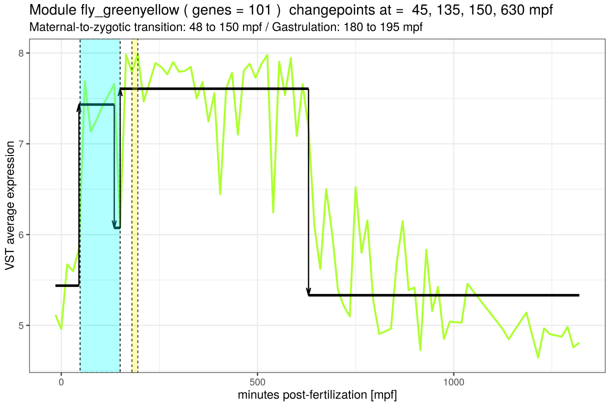
10.1 Supplementary figures for D. melanogaster modules

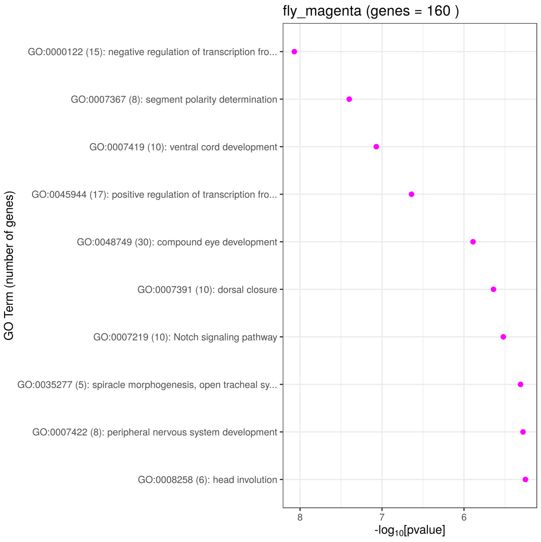
10.1.1 Modules associated to stage 2 (MZT) [25-70 mpf]

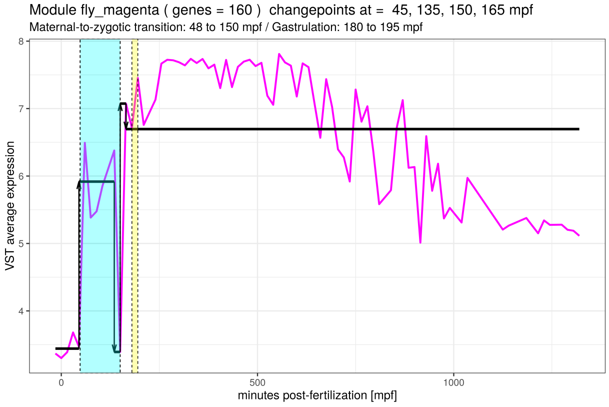




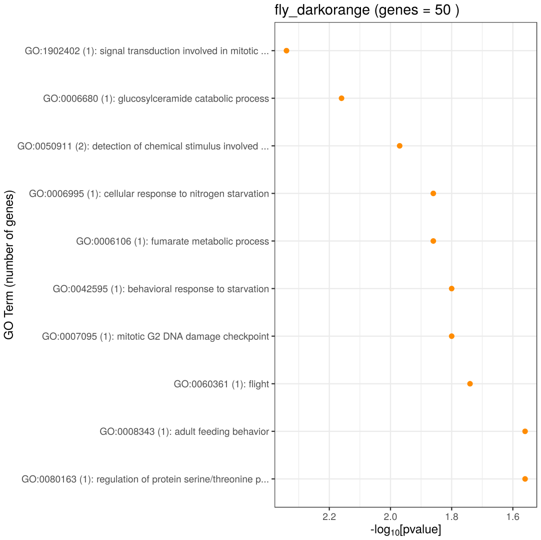

fly_greenyellow (genes = 101)
GO Term (number of genes)
GO:0007219 (11): Notch signaling pathway
GO:0001708 (10): cell fate specification
GO:0016360 (5): sensory organ precursor cell fate determ...
GO:0007400 (5): neuroblast fate determination
GO:0007399 (24): nervous system development
GO:0046331 (3): lateral inhibition
GO:0007419 (6): ventral cord development
GO:0061331 (3): epithelial cell proliferation involved i...
GO:0000719 (2): photoreactive repair
GO:0002052 (3): positive regulation of neuroblast prolif...
10
8
6
4
-log10[pvalue]

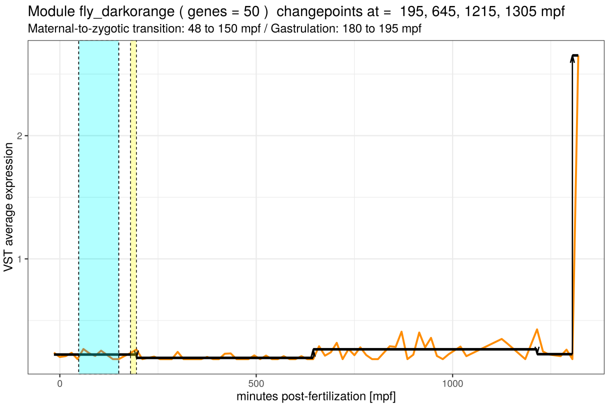

Module fly_greenyellow (genes = 101) changepoints at = 45, 135, 150, 630 mpf
Maternal-to-zygotic transition: 48 to 150 mpf / Gastrulation: 180 to 195 mpf
VST average expression
8
7
6
5
0
500
1000
minutes post-fertilization [mpf]


fly_magenta (genes = 160)
GO Term (number of genes)
GO:0000122 (15): negative regulation of transcription fro...
GO:0007367 (8): segment polarity determination
GO:0007419 (10): ventral cord development
GO:0045944 (17): positive regulation of transcription fro...
GO:0048749 (30): compound eye development
GO:0007391 (10): dorsal closure
GO:0007219 (10): Notch signaling pathway
GO:0035277 (5): spiracle morphogenesis, open tracheal sy...
GO:0007422 (8): peripheral nervous system development
GO:0008258 (6): head involution
8
7
6
-log10[pvalue]

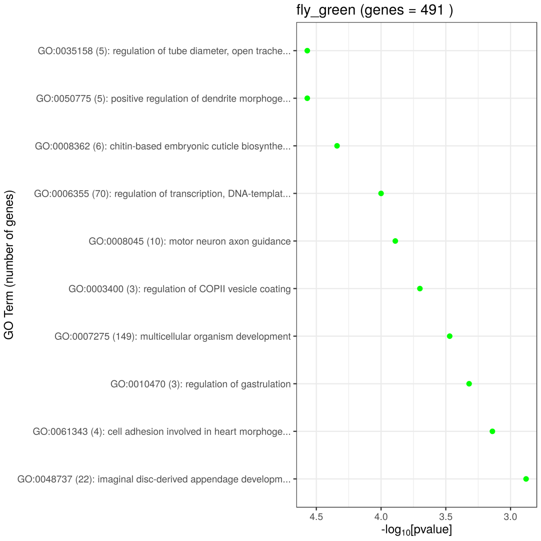

Module fly_magenta (genes = 160) changepoints at = 45, 135, 150, 165 mpf
Maternal-to-zygotic transition: 48 to 150 mpf / Gastrulation: 180 to 195 mpf
VST average expression
8
7
6
5
4
0
500
1000
minutes post-fertilization [mpf]

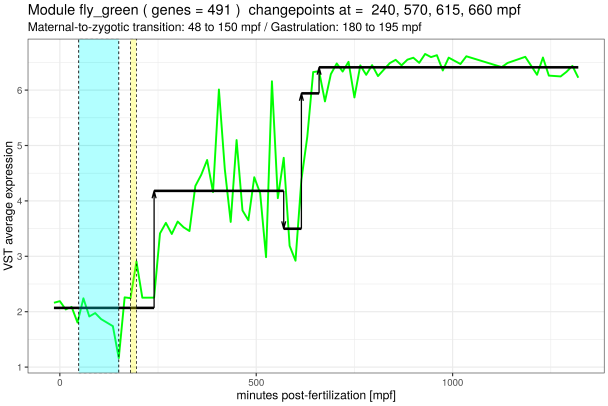
10.1.2 Module associated to stage 6 (gastrulation) [180-195 mpf]



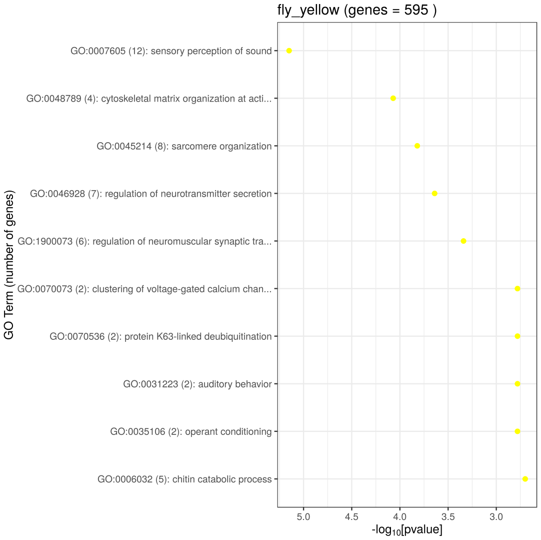


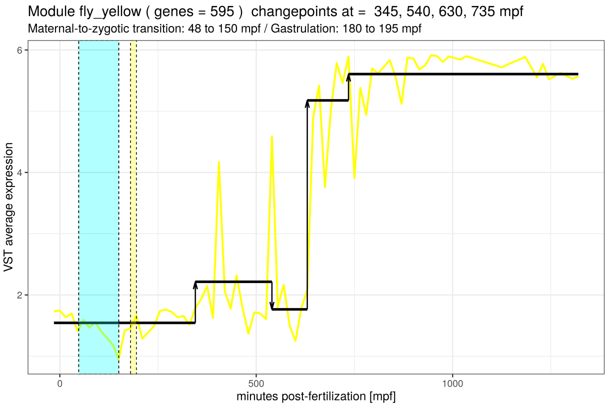
10.1.3 Module associated to stage 9 [230-260 mpf]

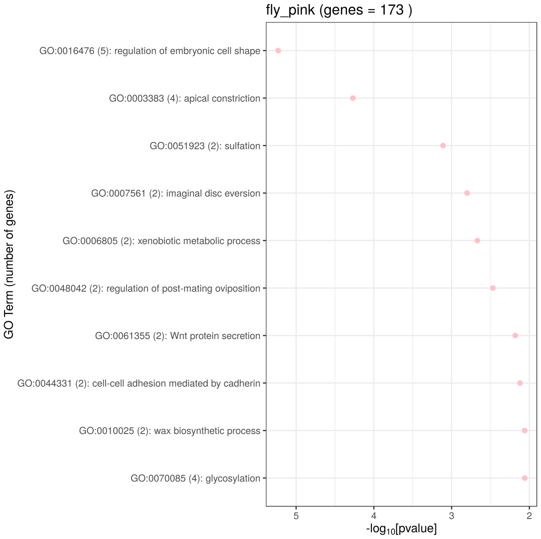


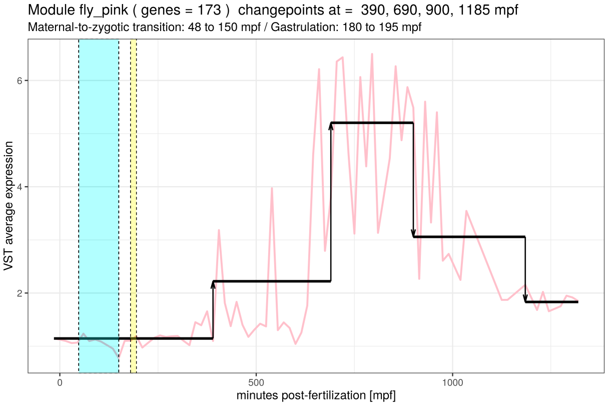


10.1.4 Modules associated to stage 11 [320-440 mpf]

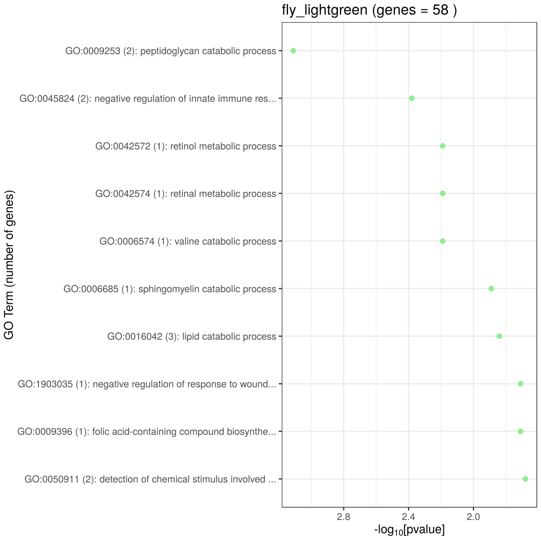


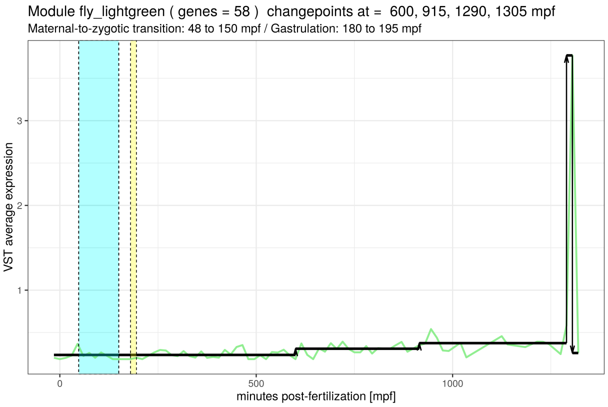




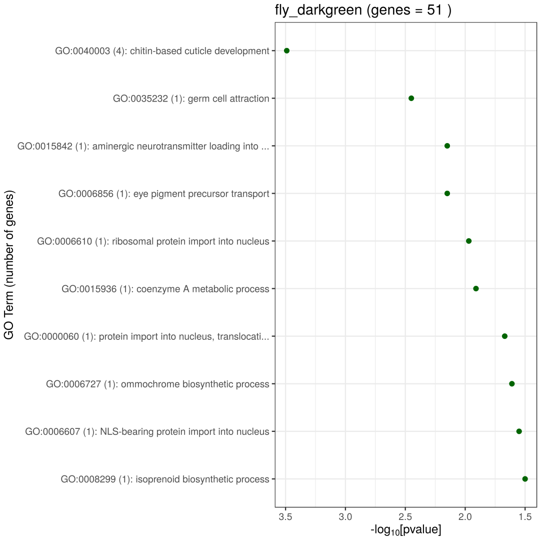


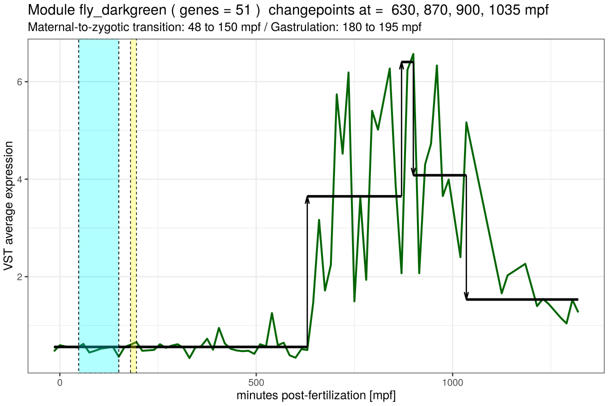
10.1.5 Module associated to stage 13 [560-620 mpf]

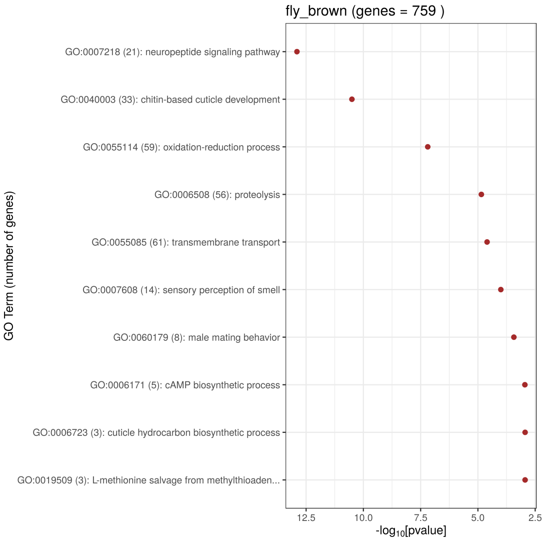


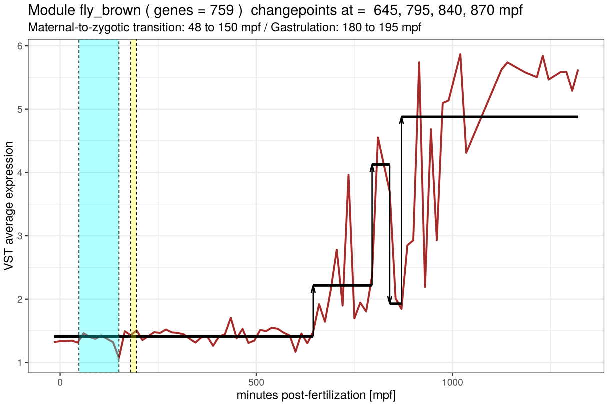


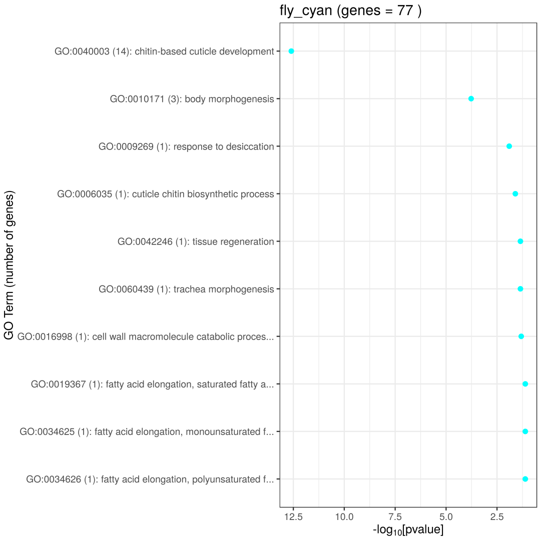
10.1.6 Modules associated to stage 14 [620-680 mpf]

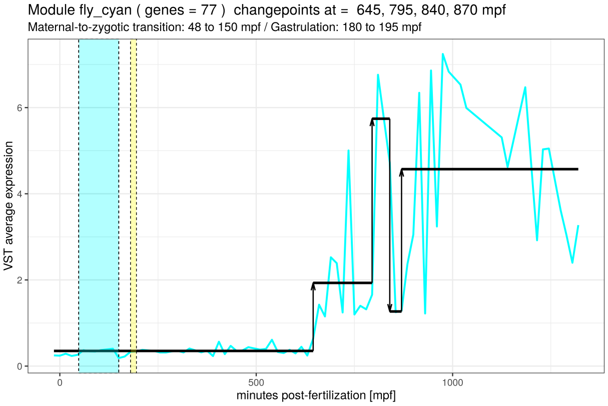


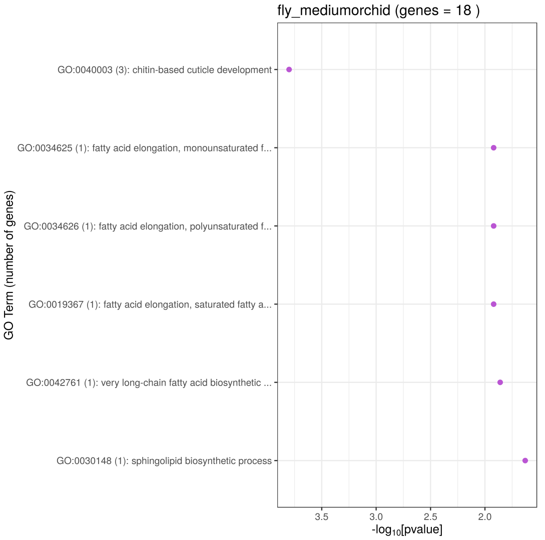


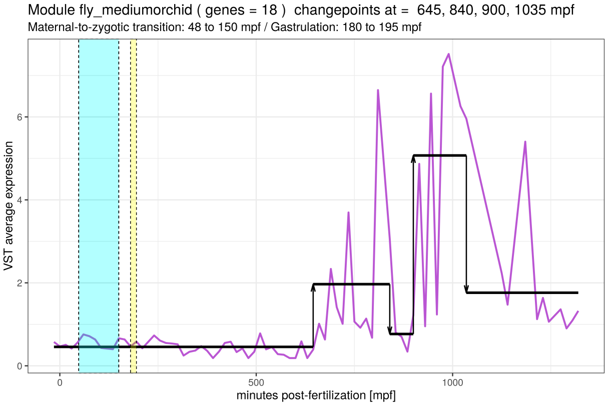

fly_brown (genes = 759)
GO:0007218 (21): neuropeptide signaling pathway
GO:0040003 (33): chitin-based cuticle development
GO:0055114 (59): oxidation-reduction process
GO:0006508 (56): proteolysis
GO:0055085 (61): transmembrane transport
GO:0007608 (14): sensory perception of smell
GO:0060179 (8): male mating behavior
GO:0006171 (5): cAMP biosynthetic process
GO:0006723 (3): cuticle hydrocarbon biosynthetic process
GO:0019509 (3): L-methionine salvage from methylthioaden...
GO Term (number of genes)
12.5
10.0
7.5
5.0
2.5
-log$_{10}$[pvalue]

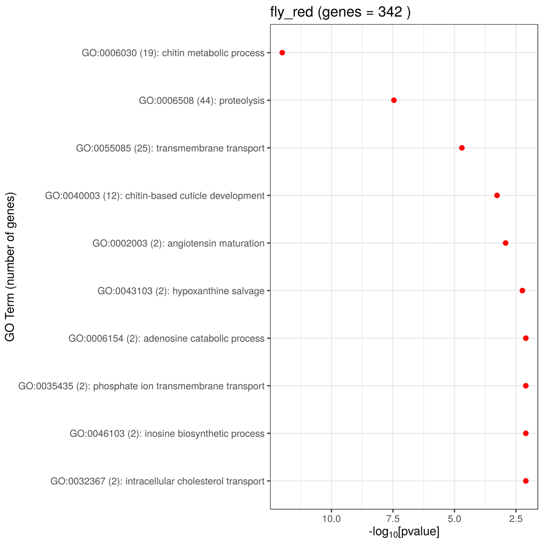

Module fly_brown (genes = 759) changepoints at = 645, 795, 840, 870 mpf
Maternal-to-zygotic transition: 48 to 150 mpf / Gastrulation: 180 to 195 mpf
VST average expression
6
5
4
3
2
1
0
500
1000
minutes post-fertilization [mpf]

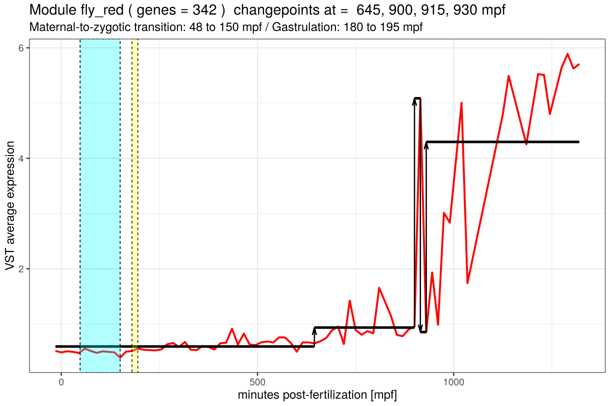

fly_cyan (genes = 77)
GO:0040003 (14): chitin-based cuticle development
GO:0010171 (3): body morphogenesis
GO:0009269 (1): response to desiccation
GO:0006035 (1): cuticle chitin biosynthetic process
GO:0042246 (1): tissue regeneration
GO:0060439 (1): trachea morphogenesis
GO:0016998 (1): cell wall macromolecule catabolic proces...
GO:0019367 (1): fatty acid elongation, saturated fatty a...
GO:0034625 (1): fatty acid elongation, monounsaturated f...
GO:0034626 (1): fatty acid elongation, polyunsaturated f...
GO Term (number of genes)
12.5
10.0
7.5
5.0
2.5
-log$_{10}$[pvalue]

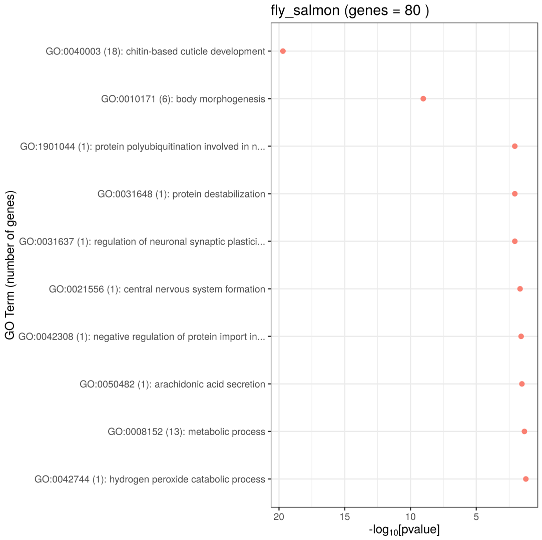

Module fly_cyan (genes = 77) changepoints at = 645, 795, 840, 870 mpf
Maternal-to-zygotic transition: 48 to 150 mpf / Gastrulation: 180 to 195 mpf
VST average expression
6
4
2
0
0
500
1000
minutes post-fertilization [mpf]

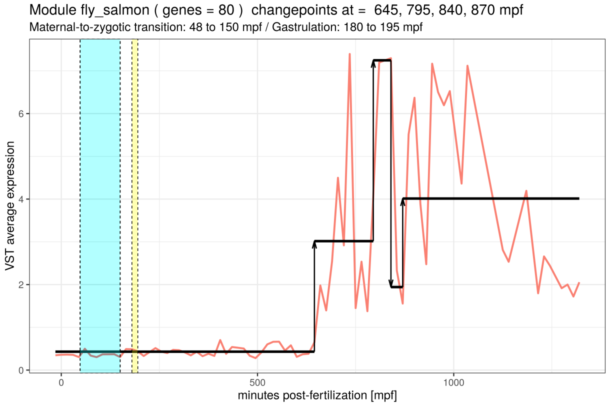

fly_mediumorchid (genes = 18)
GO:0040003 (3): chitin-based cuticle development
GO:0034625 (1): fatty acid elongation, monounsaturated f...
GO:0034626 (1): fatty acid elongation, polyunsaturated f...
GO:0019367 (1): fatty acid elongation, saturated fatty a...
GO:0042761 (1): very long-chain fatty acid biosynthetic ...
GO:0030148 (1): sphingolipid biosynthetic process
GO Term (number of genes)
3.5
3.0
2.5
2.0
-log$_{10}$[pvalue]

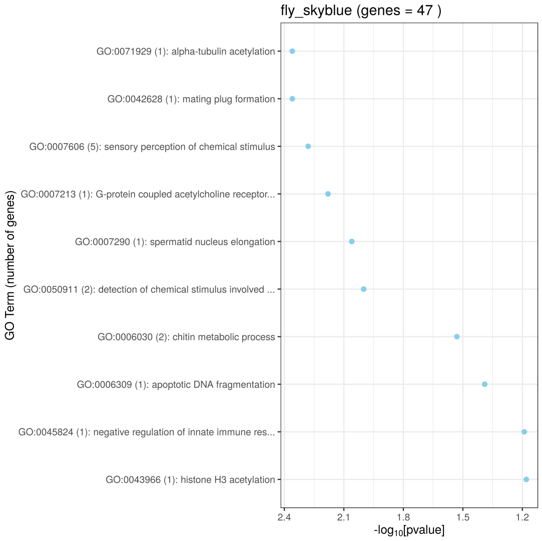

Module fly_mediumorchid (genes = 18) changepoints at = 645, 840, 900, 1035 mpf
Maternal-to-zygotic transition: 48 to 150 mpf / Gastrulation: 180 to 195 mpf
VST average expression
6
4
2
0
0
500
1000
minutes post-fertilization [mpf]

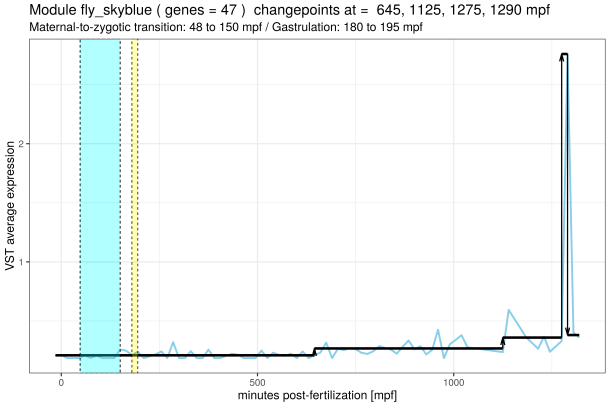

fly_red (genes = 342)
GO Term (number of genes)
GO:0006030 (19): chitin metabolic process
GO:0006508 (44): proteolysis
GO:0055085 (25): transmembrane transport
GO:0040003 (12): chitin-based cuticle development
GO:0002003 (2): angiotensin maturation
GO:0043103 (2): hypoxanthine salvage
GO:0006154 (2): adenosine catabolic process
GO:0035435 (2): phosphate ion transmembrane transport
GO:0046103 (2): inosine biosynthetic process
GO:0032367 (2): intracellular cholesterol transport
10.0
7.5
5.0
2.5
-log$_{10}$[pvalue]

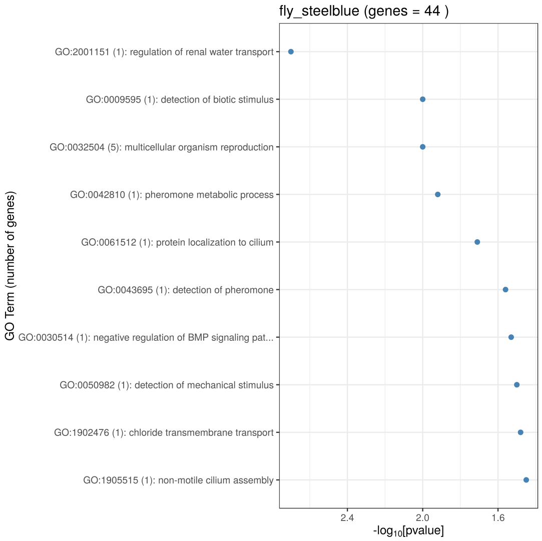

Module fly_red (genes = 342) changepoints at = 645, 900, 915, 930 mpf
Maternal-to-zygotic transition: 48 to 150 mpf / Gastrulation: 180 to 195 mpf
VST average expression
6
4
2
0
500
1000
minutes post-fertilization [mpf]

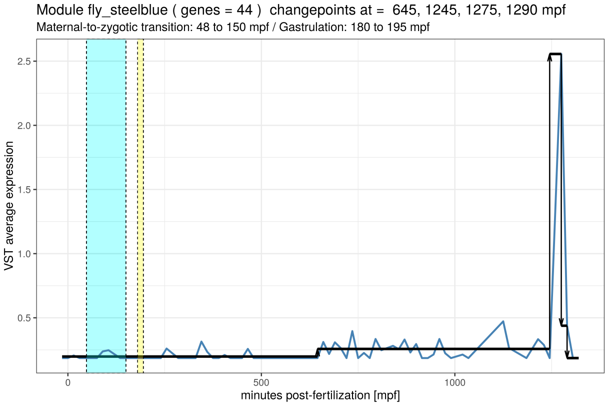

fly_salmon (genes = 80)
GO Term (number of genes)
GO:0040003 (18): chitin-based cuticle development
GO:0010171 (6): body morphogenesis
GO:1901044 (1): protein polyubiquitination involved in n...
GO:0031648 (1): protein destabilization
GO:0031637 (1): regulation of neuronal synaptic plastici...
GO:0021556 (1): central nervous system formation
GO:0042308 (1): negative regulation of protein import in...
GO:0050482 (1): arachidonic acid secretion
GO:0008152 (13): metabolic process
GO:0042744 (1): hydrogen peroxide catabolic process
20
15
10
5
-log$_{10}$[pvalue]


Module fly_salmon (genes = 80) changepoints at = 645, 795, 840, 870 mpf
Maternal-to-zygotic transition: 48 to 150 mpf / Gastrulation: 180 to 195 mpf
VST average expression
6
4
2
0
0
500
1000
minutes post-fertilization [mpf]

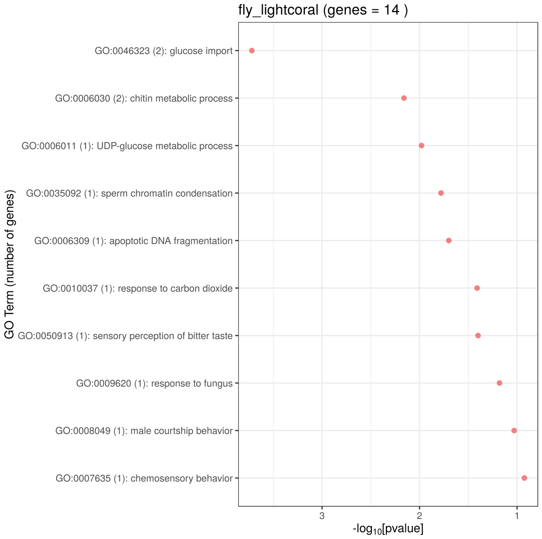

fly_skyblue (genes = 47)
GO Term (number of genes)
GO:0071929 (1): alpha-tubulin acetylation
GO:0042628 (1): mating plug formation
GO:0007606 (5): sensory perception of chemical stimulus
GO:0007213 (1): G-protein coupled acetylcholine receptor...
GO:0007290 (1): spermatid nucleus elongation
GO:0050911 (2): detection of chemical stimulus involved ...
GO:0006030 (2): chitin metabolic process
GO:0006309 (1): apoptotic DNA fragmentation
GO:0045824 (1): negative regulation of innate immune res...
GO:0043966 (1): histone H3 acetylation
2.4
2.1
1.8
1.5
1.2
-log$_{10}$[pvalue]

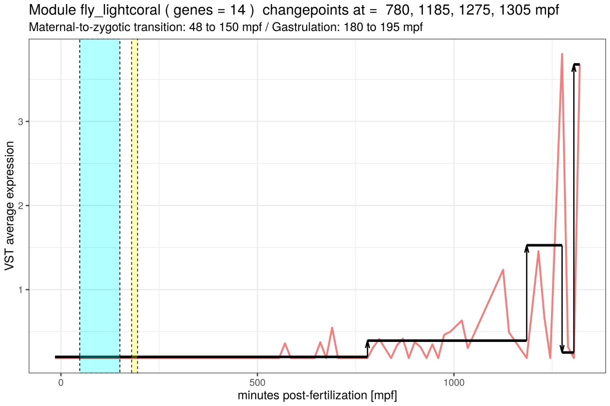

Module fly_skyblue (genes = 47) changepoints at = 645, 1125, 1275, 1290 mpf
Maternal-to-zygotic transition: 48 to 150 mpf / Gastrulation: 180 to 195 mpf
VST average expression
2
1
0
500
1000
minutes post-fertilization [mpf]

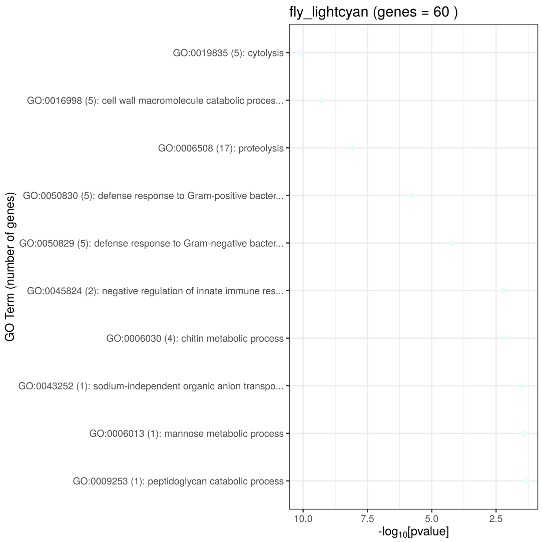


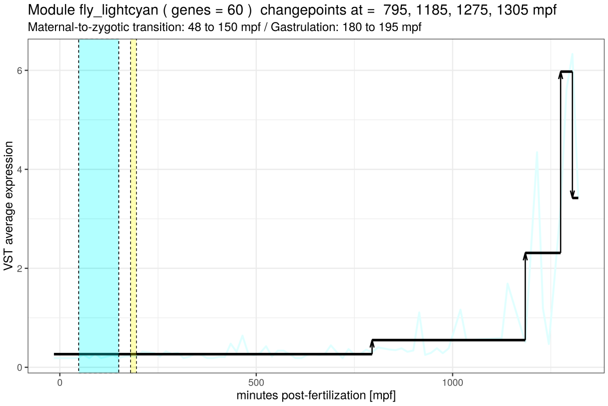


10.1.7 Modules associated to stage 15 [680-800 mpf]

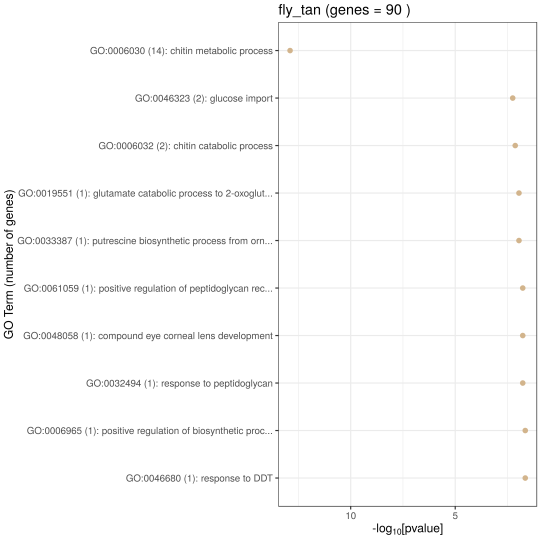


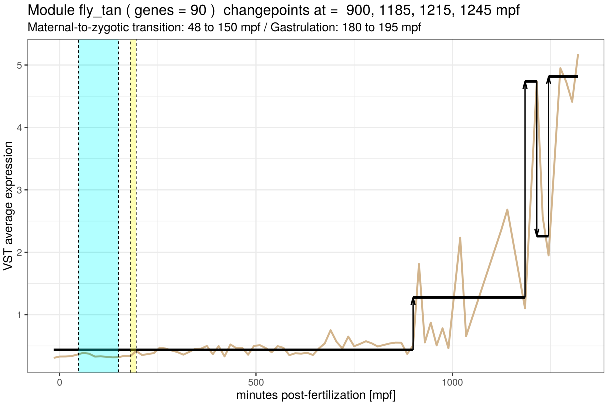




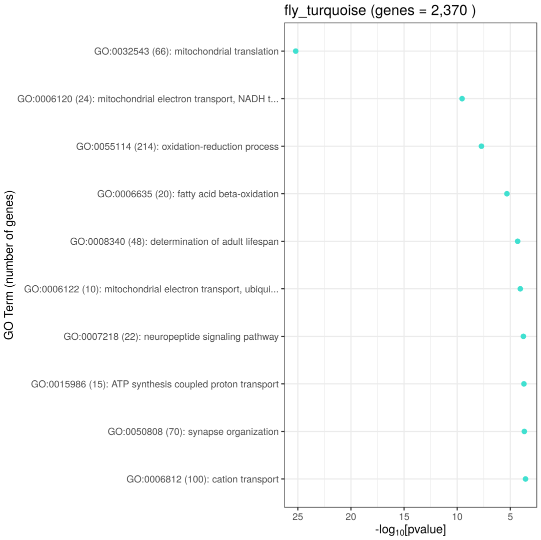


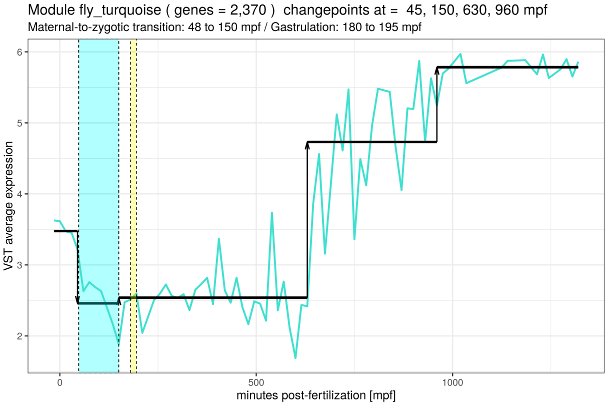
10.1.8 Modules associated to stage 16 [800-900 mpf]

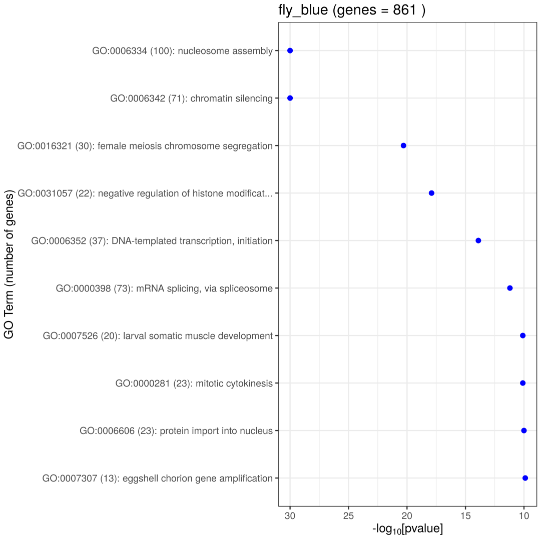


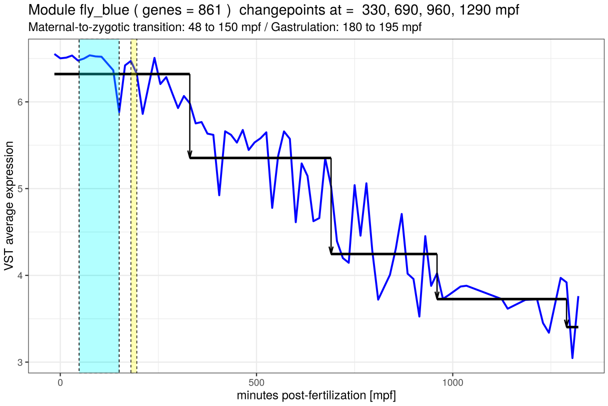


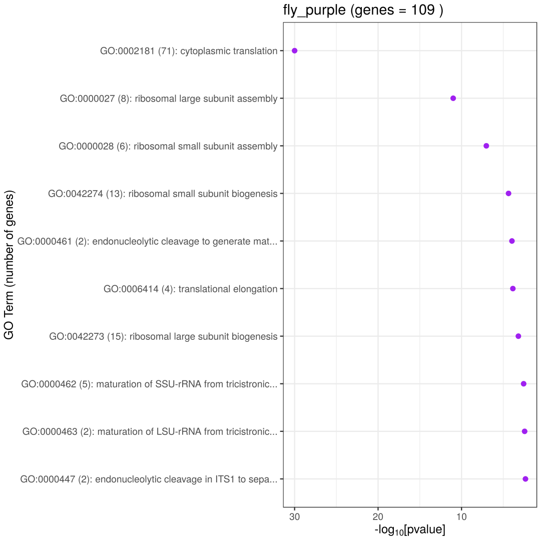
10.1.9 Modules with unique expression profiles

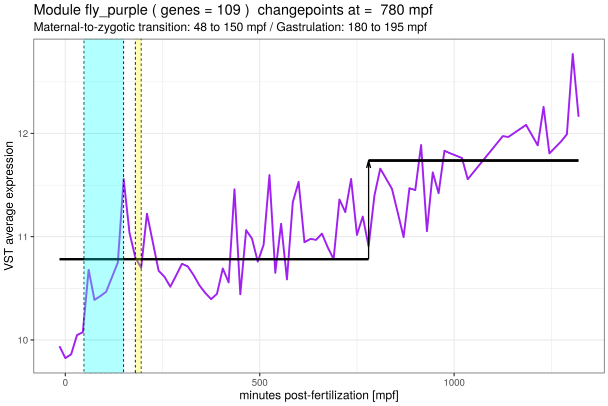




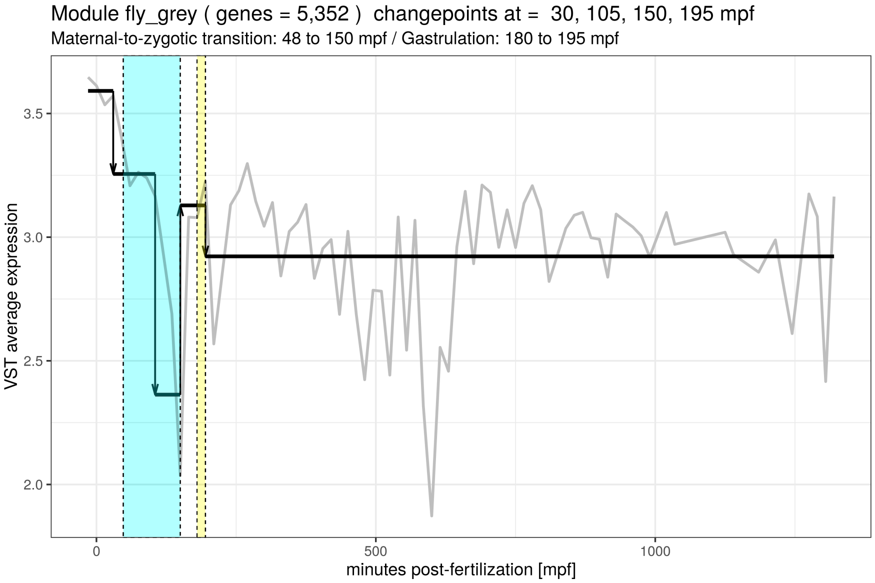


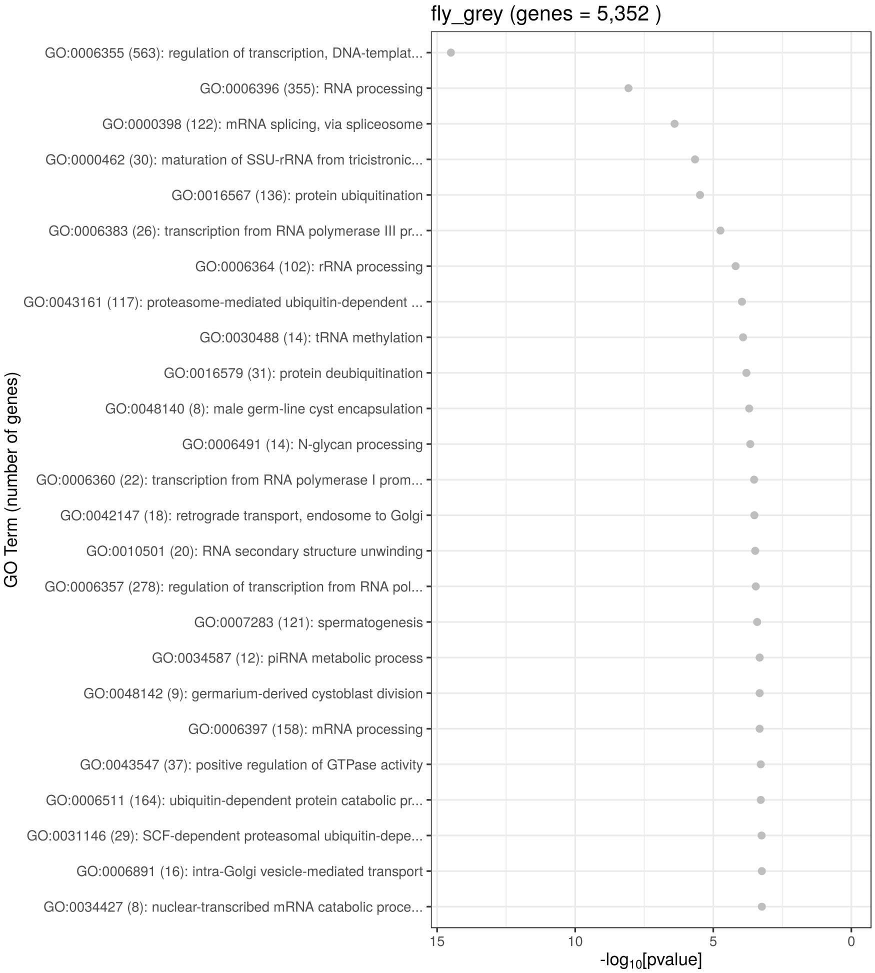



fly_purple (genes = 109)
GO Term (number of genes)
GO:0002181 (71): cytoplasmic translation
GO:0000027 (8): ribosomal large subunit assembly
GO:0000028 (6): ribosomal small subunit assembly
GO:0042274 (13): ribosomal small subunit biogenesis
GO:0000461 (2): endonucleolytic cleavage to generate mat...
GO:0006414 (4): translational elongation
GO:0042273 (15): ribosomal large subunit biogenesis
GO:0000462 (5): maturation of SSU-rRNA from tricistronic...
GO:0000463 (2): maturation of LSU-rRNA from tricistronic...
GO:0000447 (2): endonucleolytic cleavage in ITS1 to sepa...
30
20
10
-log10[pvalue]

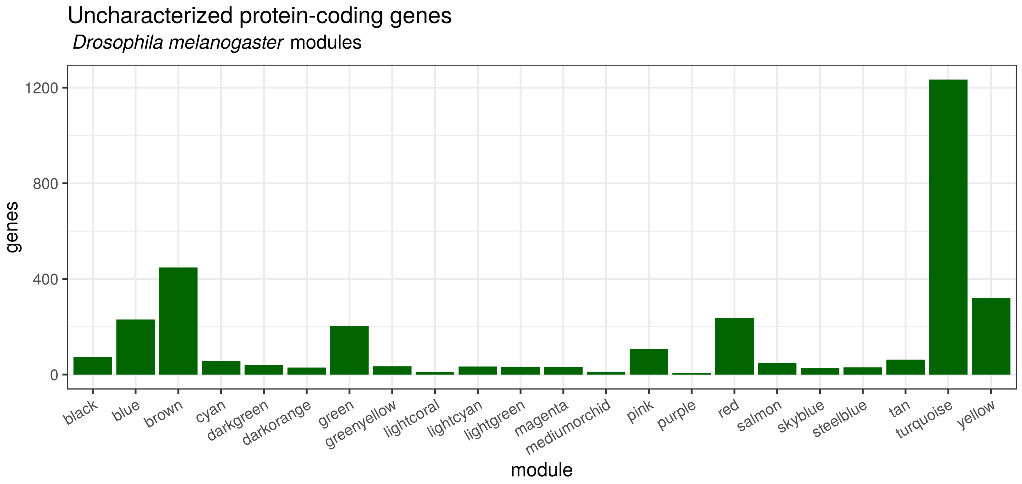

Module fly_purple (genes = 109) changepoints at = 780 mpf
Maternal-to-zygotic transition: 48 to 150 mpf / Gastrulation: 180 to 195 mpf
VST average expression
12
11
10
0
500
1000
minutes post-fertilization [mpf]

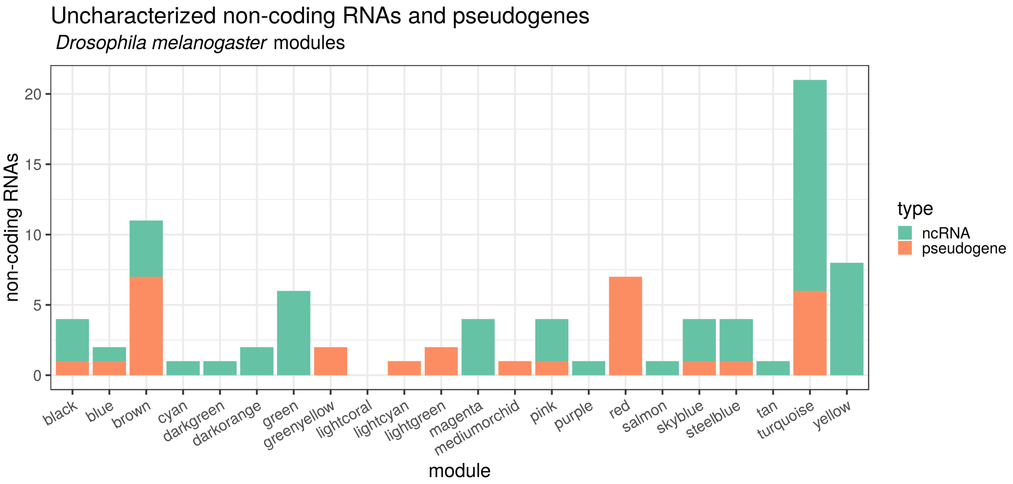
10.1.10 Grey module





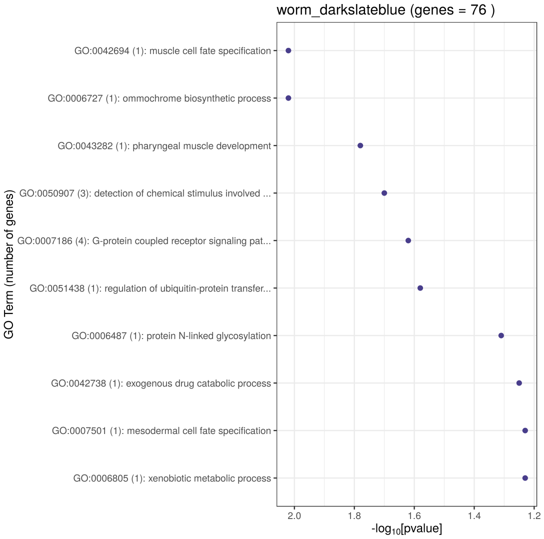
10.1.11 Uncharacterized genes in modules

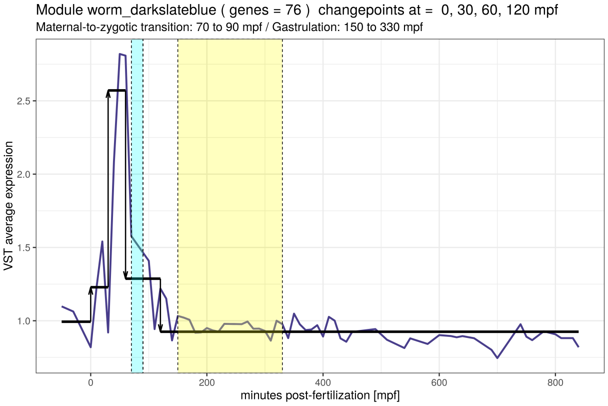

Uncharacterized protein-coding genes
Drosophila melanogaster modules

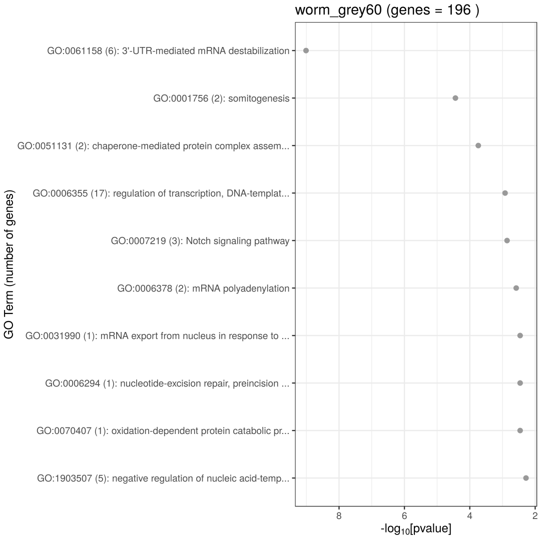

Uncharacterized non-coding RNAs and pseudogenes
Drosophila melanogaster modules

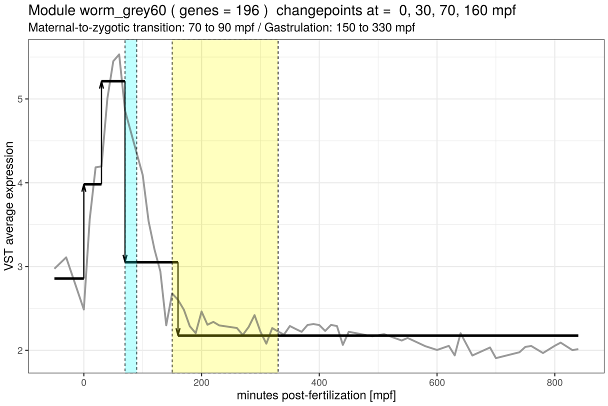
10.2 Supplementary figures for C. elegans modules

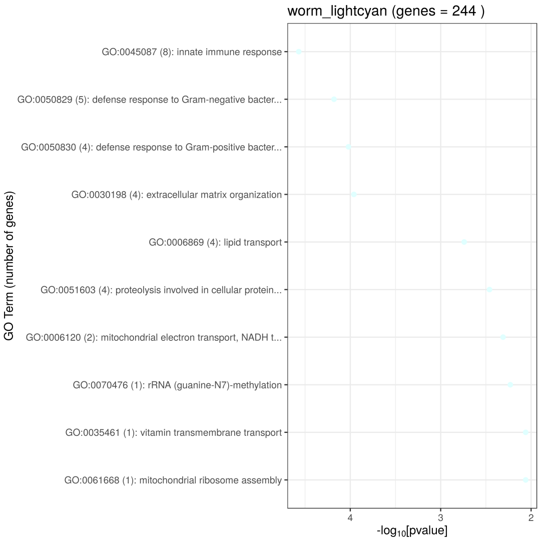
10.2.1 Modules associated to 2-cell stage [0-50 mpf]



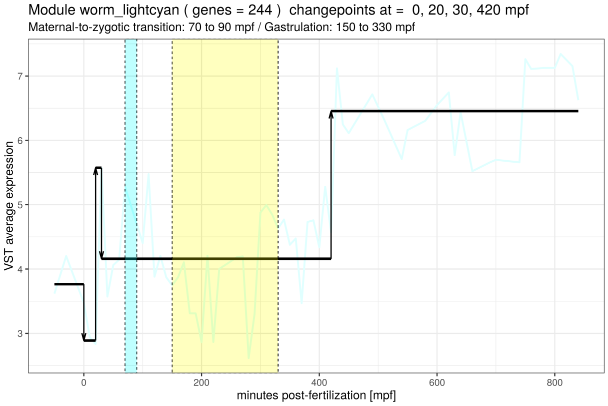






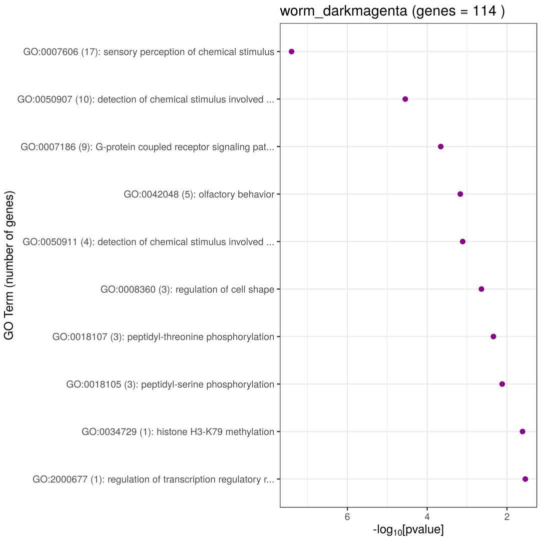


worm_lightcyan (genes = 244)

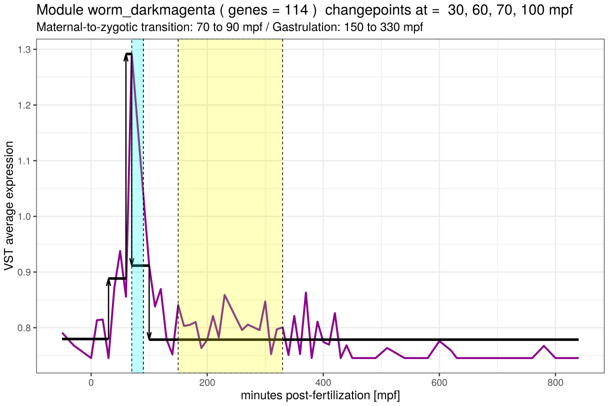


Module worm_lightcyan (genes = 244) changepoints at = 0, 20, 30, 420 mpf

Maternal-to-zygotic transition: 70 to 90 mpf / Gastrulation: 150 to 330 mpf

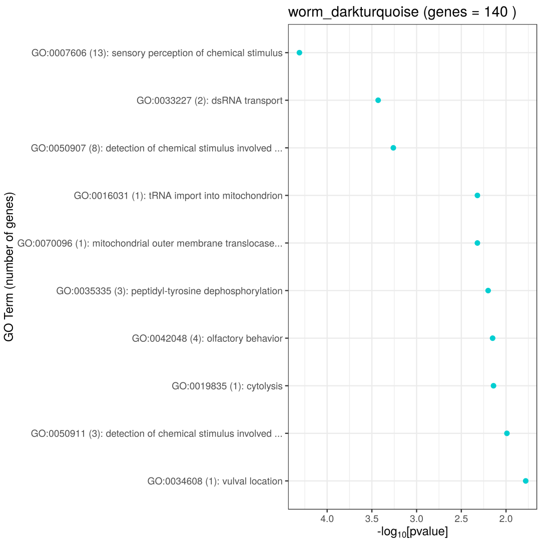


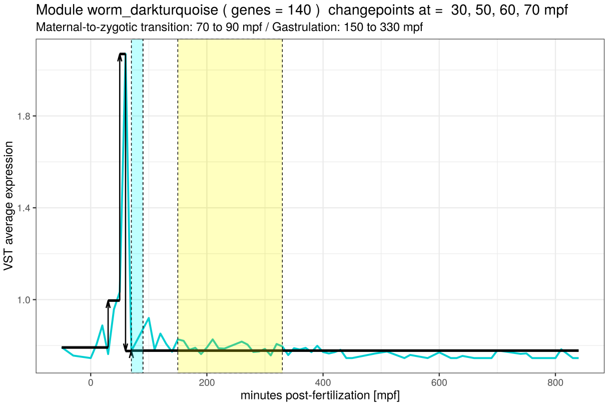
worm_darkmagenta (genes = 114)

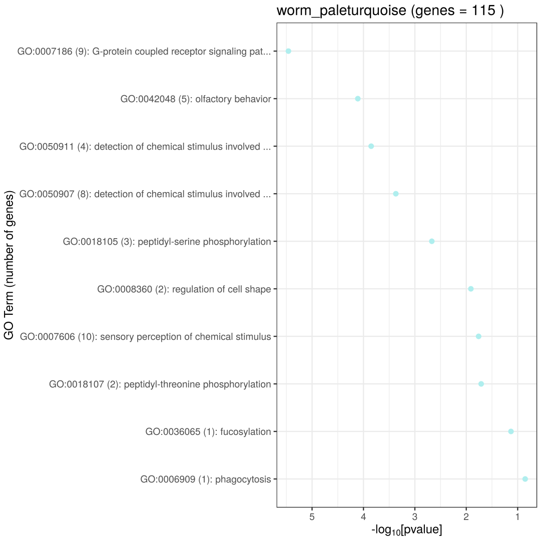


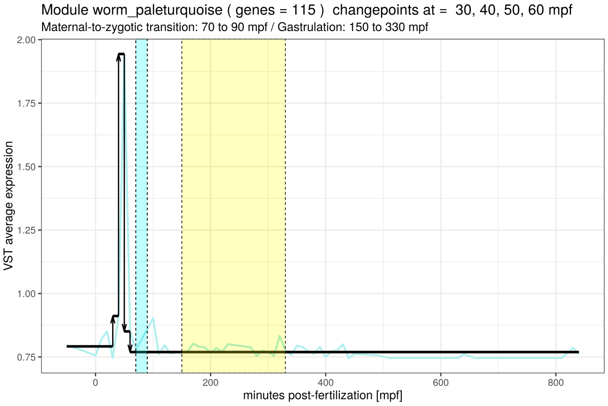
Module worm_darkmagenta (genes = 114) changepoints at = 30, 60, 70, 100 mpf

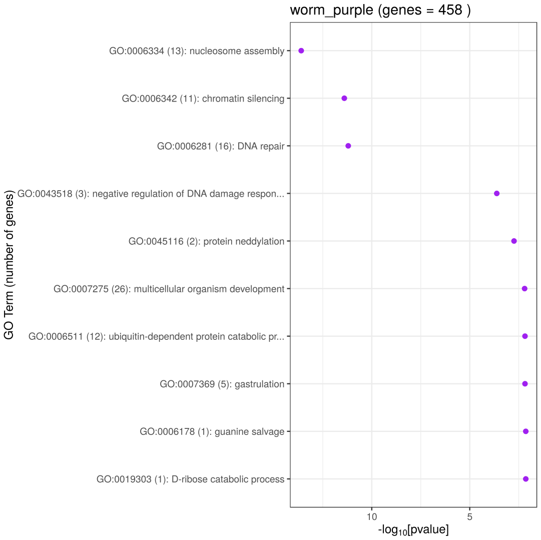
Maternal-to-zygotic transition: 70 to 90 mpf / Gastrulation: 150 to 330 mpf

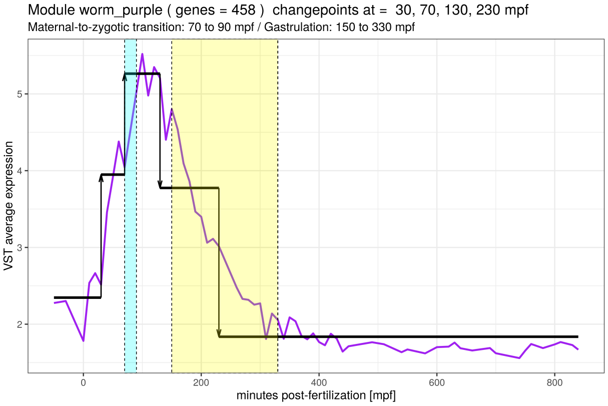

worm_darkturquoise (genes = 140)
GO Term (number of genes)
GO:0007606 (13): sensory perception of chemical stimulus
GO:0033227 (2): dsRNA transport
GO:0050907 (8): detection of chemical stimulus involved ...
GO:0016031 (1): tRNA import into mitochondrion
GO:0070096 (1): mitochondrial outer membrane translocase...
GO:0035335 (3): peptidyl-tyrosine dephosphorylation
GO:0042048 (4): olfactory behavior
GO:0019835 (1): cytolysis
GO:0050911 (3): detection of chemical stimulus involved ...
GO:0034608 (1): vulval location
4.0
3.5
3.0
2.5
2.0
-log10[pvalue]

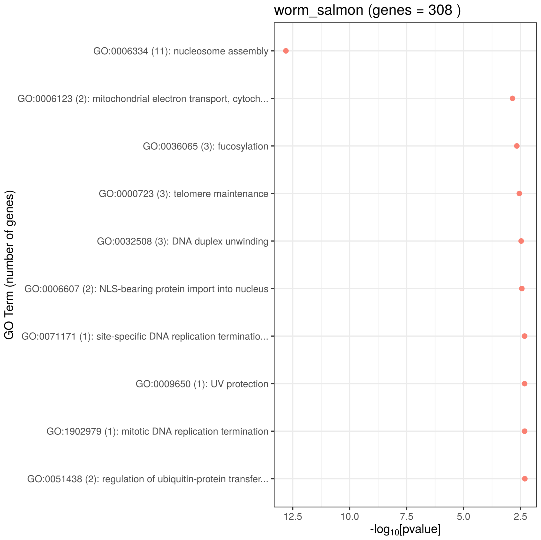

Module worm_darkturquoise (genes = 140) changepoints at = 30, 50, 60, 70 mpf
Maternal-to-zygotic transition: 70 to 90 mpf / Gastrulation: 150 to 330 mpf
VST average expression
1.8
1.4
1.0
0
200
400
600
800
minutes post-fertilization [mpf]

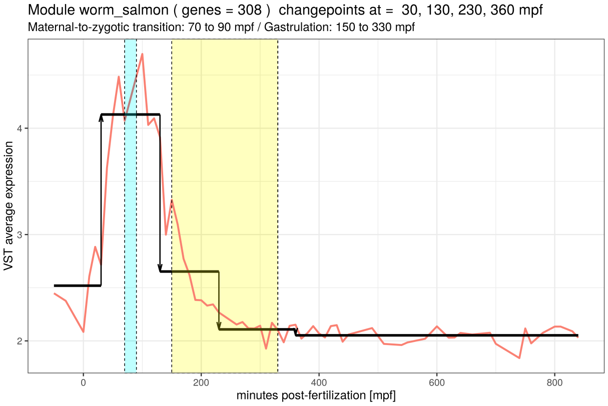

worm_paleturquoise (genes = 115)
GO Term (number of genes)
GO:0007186 (9): G-protein coupled receptor signaling pat...
GO:0042048 (5): olfactory behavior
GO:0050911 (4): detection of chemical stimulus involved ...
GO:0050907 (8): detection of chemical stimulus involved ...
GO:0018105 (3): peptidyl-serine phosphorylation
GO:0008360 (2): regulation of cell shape
GO:0007606 (10): sensory perception of chemical stimulus
GO:0018107 (2): peptidyl-threonine phosphorylation
GO:0036065 (1): fucosylation
GO:0006909 (1): phagocytosis
5
4
3
2
1
-log10[pvalue]

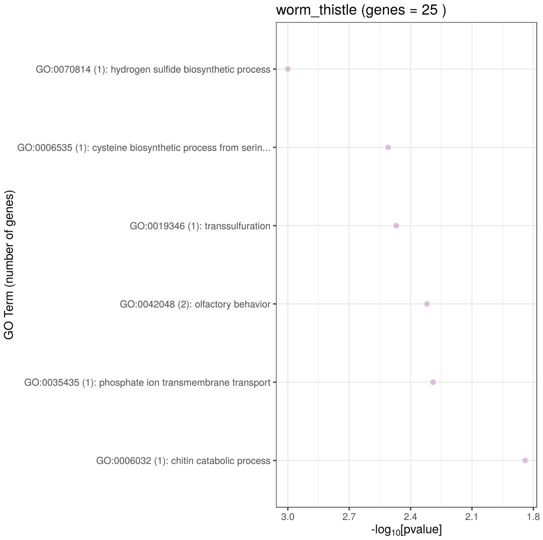

Module worm_paleturquoise (genes = 115) changepoints at = 30, 40, 50, 60 mpf
Maternal-to-zygotic transition: 70 to 90 mpf / Gastrulation: 150 to 330 mpf
VST average expression
2.00
1.75
1.50
1.25
1.00
0.75
0
200
400
600
800
minutes post-fertilization [mpf]

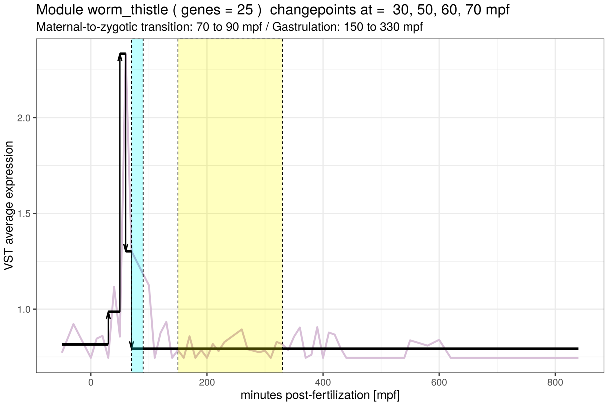

worm_purple (genes = 458)


Module worm_purple (genes = 458) changepoints at = 30, 70, 130, 230 mpf
Maternal-to-zygotic transition: 70 to 90 mpf / Gastrulation: 150 to 330 mpf

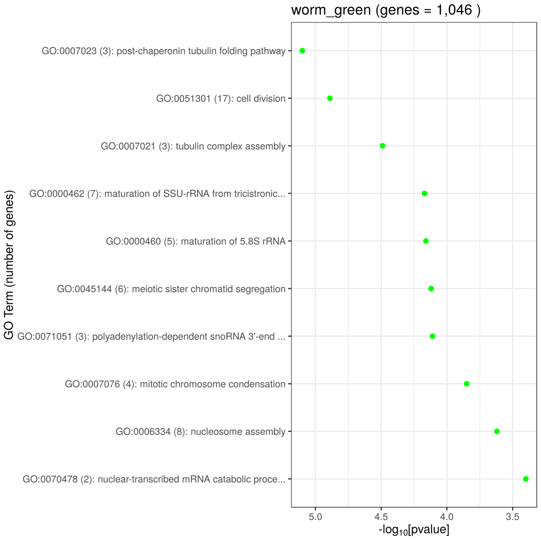

worm_salmon (genes = 308)

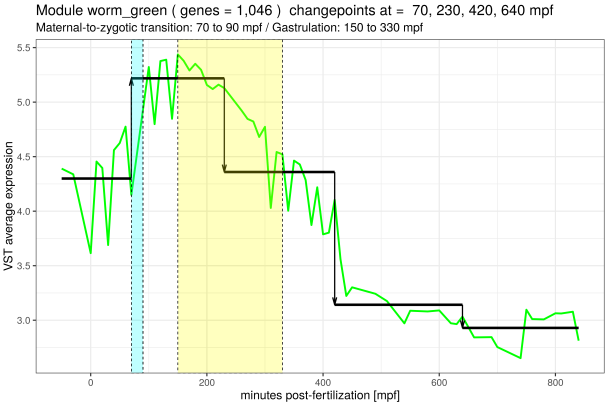

Module worm_salmon (genes = 308) changepoints at = 30, 130, 230, 360 mpf
Maternal-to-zygotic transition: 70 to 90 mpf / Gastrulation: 150 to 330 mpf

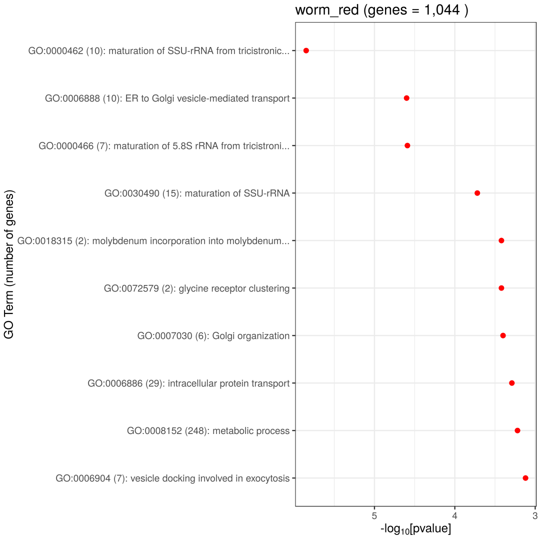

worm_thistle (genes = 25)

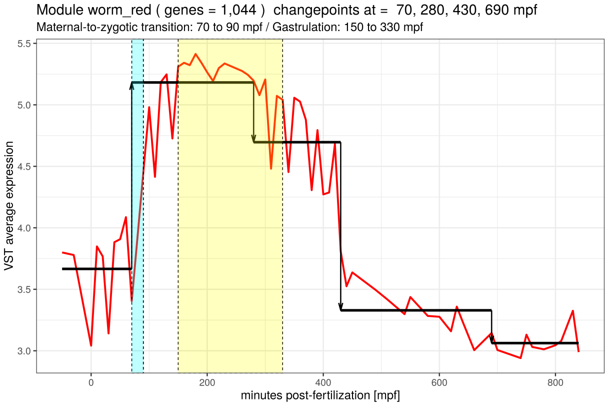

Module worm_thistle (genes = 25) changepoints at = 30, 50, 60, 70 mpf
Maternal-to-zygotic transition: 70 to 90 mpf / Gastrulation: 150 to 330 mpf

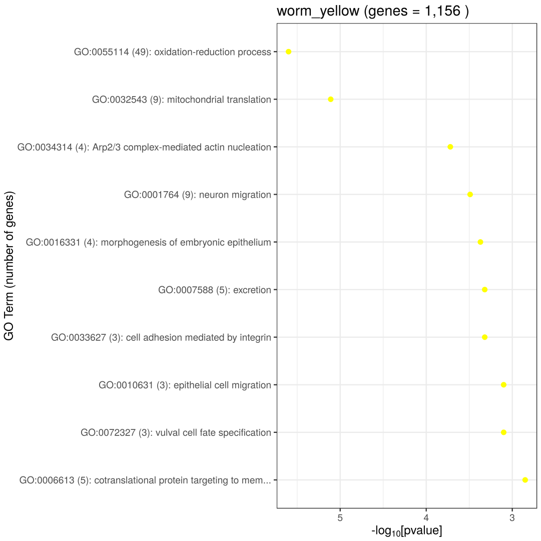
10.2.2 Modules associated to 4-cell stage (MZT) [50-100 mpf]

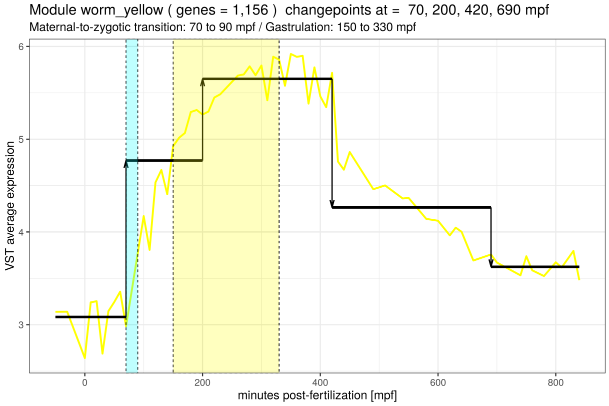




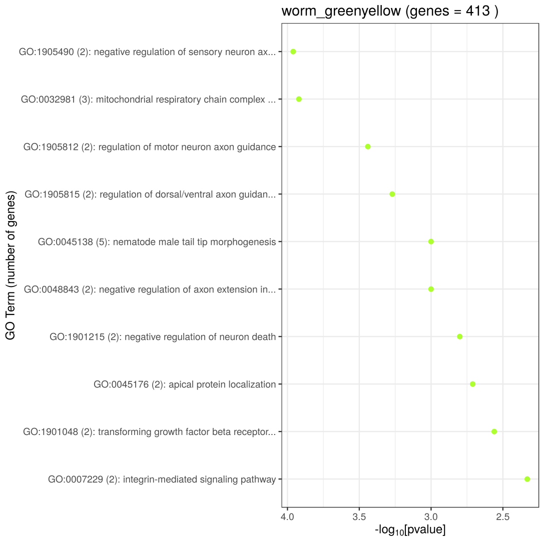


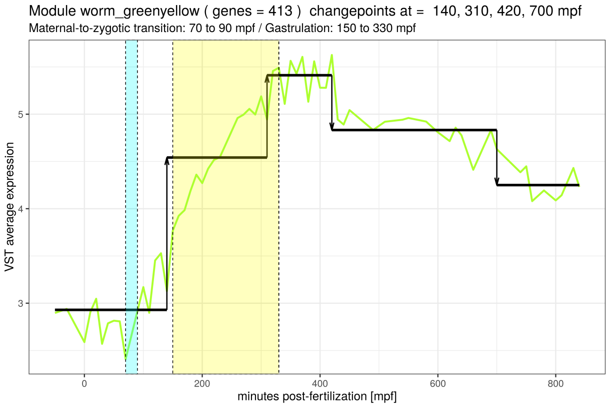




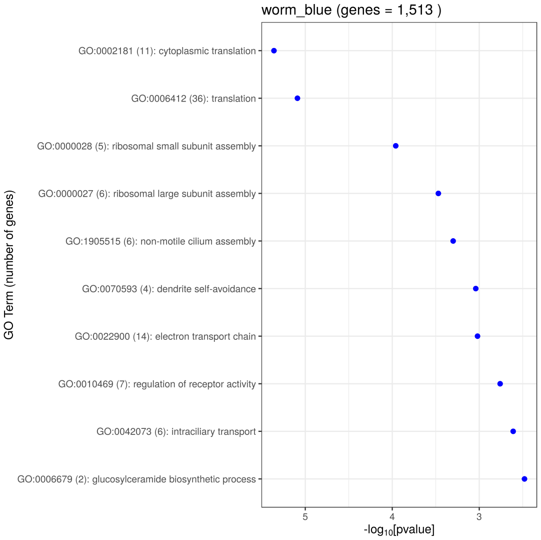


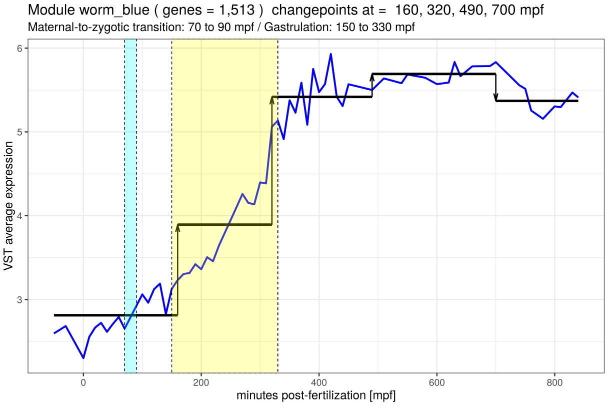
10.2.3 Module associated to 26-cell stage [100-150 mpf], MZT

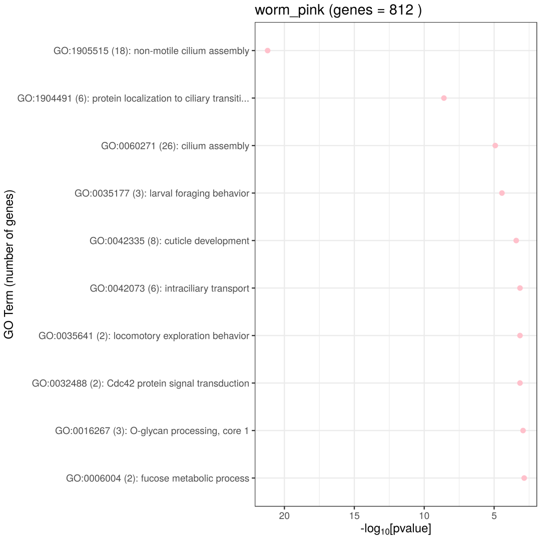


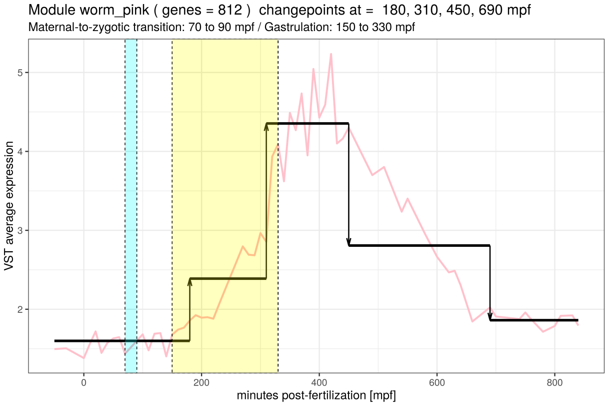


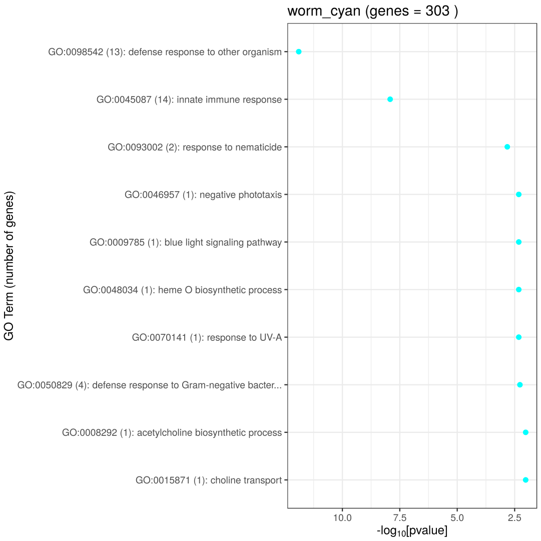
10.2.4 Modules associated to gastrulation [50-330 mpf]

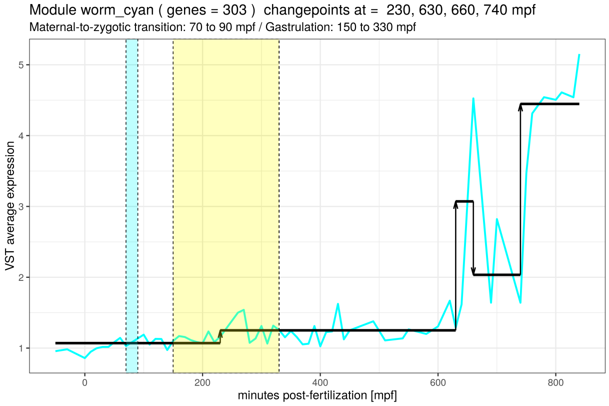


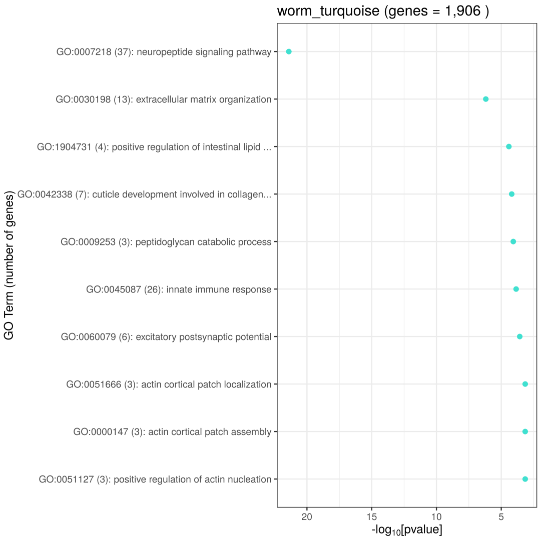


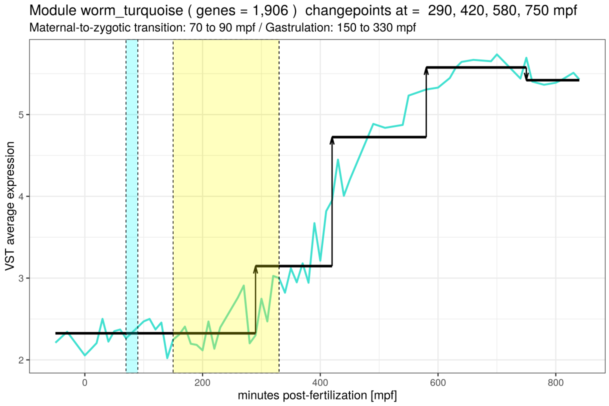




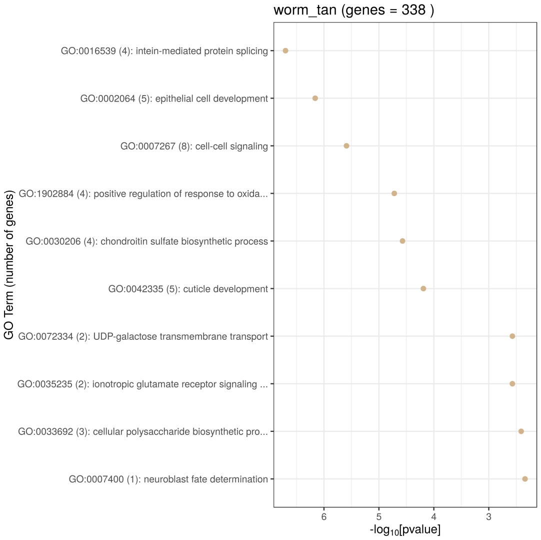

worm_cyan (genes = 303)
GO Term (number of genes)
GO:0098542 (13): defense response to other organism
GO:0045087 (14): innate immune response
GO:0093002 (2): response to nematicide
GO:0046957 (1): negative phototaxis
GO:0009785 (1): blue light signaling pathway
GO:0048034 (1): heme O biosynthetic process
GO:0070141 (1): response to UV-A
GO:0050829 (4): defense response to Gram-negative bacter...
GO:0008292 (1): acetylcholine biosynthetic process
GO:0015871 (1): choline transport
10.0
7.5
5.0
2.5
$-\log_{10}$[pvalue]

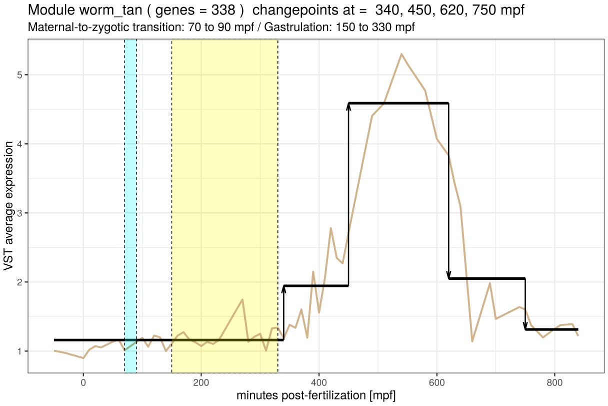

Module worm_cyan (genes = 303) changepoints at = 230, 630, 660, 740 mpf
Maternal-to-zygotic transition: 70 to 90 mpf / Gastrulation: 150 to 330 mpf
VST average expression
5
4
3
2
1
0
200
400
600
800
minutes post-fertilization [mpf]

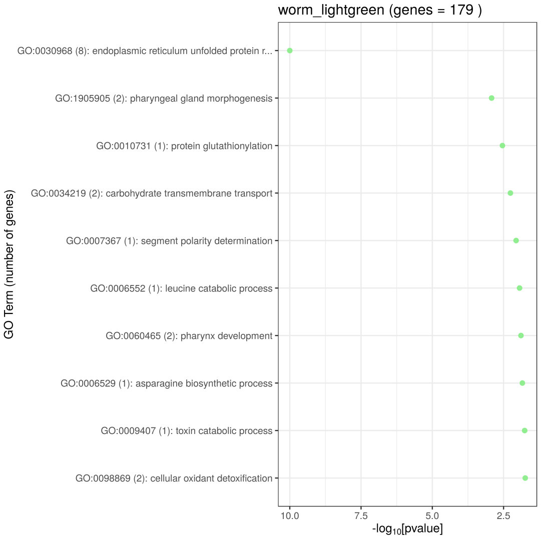

worm_turquoise (genes = 1,906)
GO Term (number of genes)
GO:0007218 (37): neuropeptide signaling pathway
GO:0030198 (13): extracellular matrix organization
GO:1904731 (4): positive regulation of intestinal lipid ...
GO:0042338 (7): cuticle development involved in collagen...
GO:0009253 (3): peptidoglycan catabolic process
GO:0045087 (26): innate immune response
GO:0060079 (6): excitatory postsynaptic potential
GO:0051666 (3): actin cortical patch localization
GO:0000147 (3): actin cortical patch assembly
GO:0051127 (3): positive regulation of actin nucleation
20
15
10
5
$-\log_{10}$[pvalue]

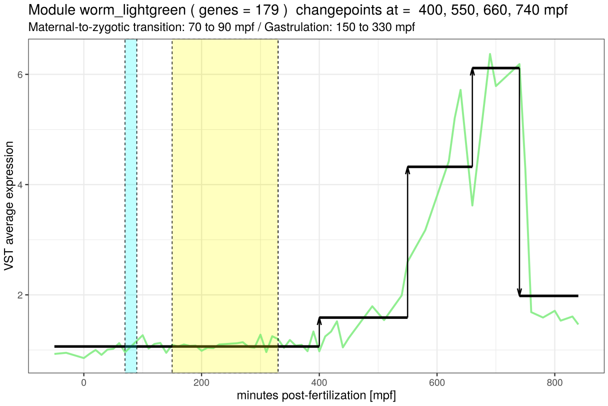

Module worm_turquoise (genes = 1,906) changepoints at = 290, 420, 580, 750 mpf
Maternal-to-zygotic transition: 70 to 90 mpf / Gastrulation: 150 to 330 mpf
VST average expression
5
4
3
2
0
200
400
600
800
minutes post-fertilization [mpf]

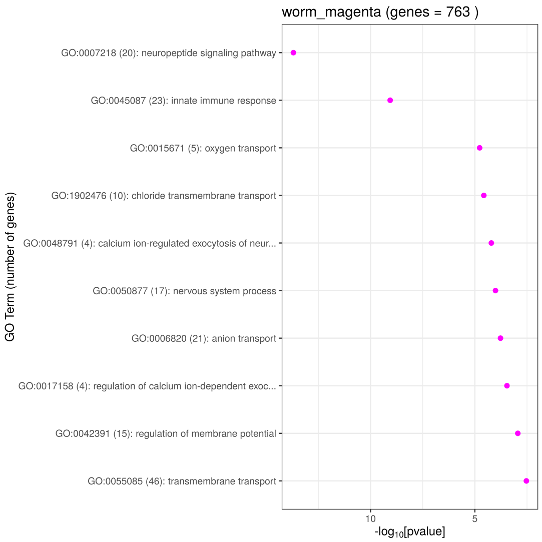
10.2.5 Modules associated to bean stage (ventral enclosure) [360-400 mpf]

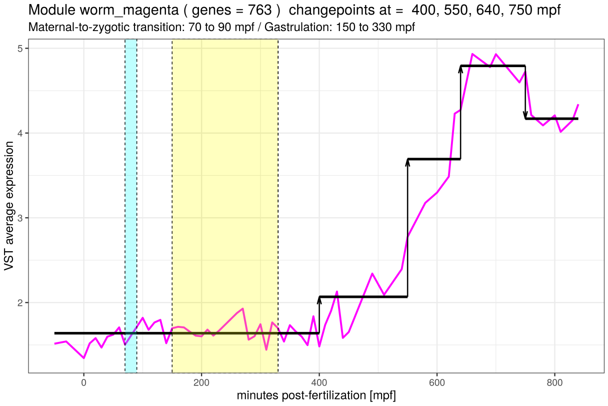




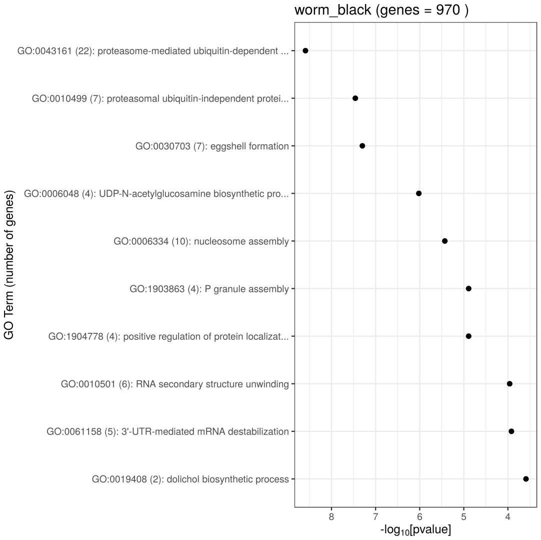


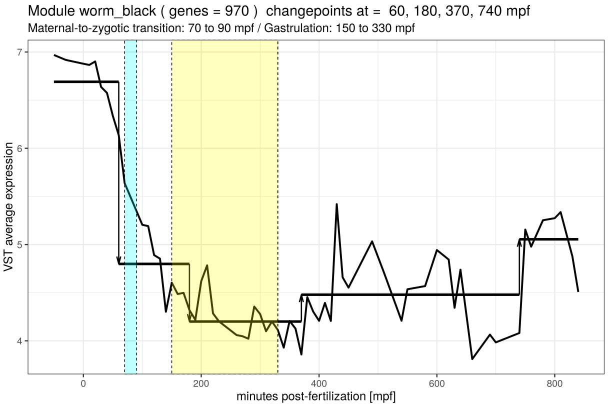


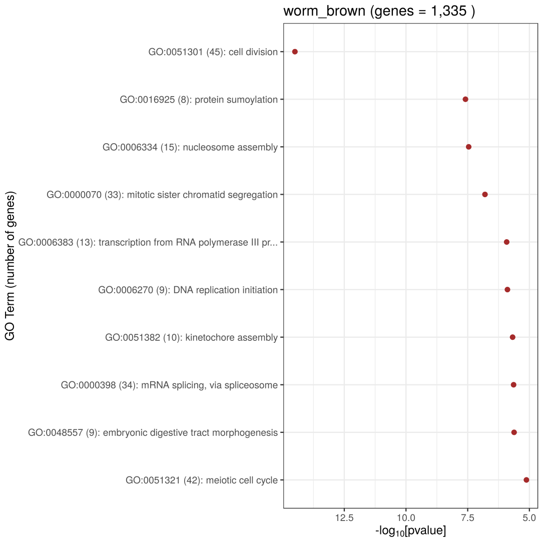


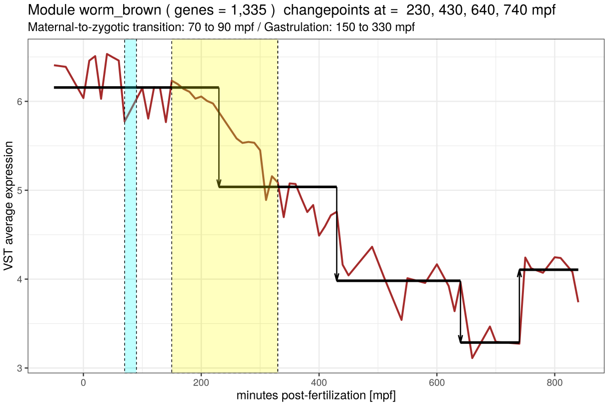


10.2.6 Modules with unique expression profiles

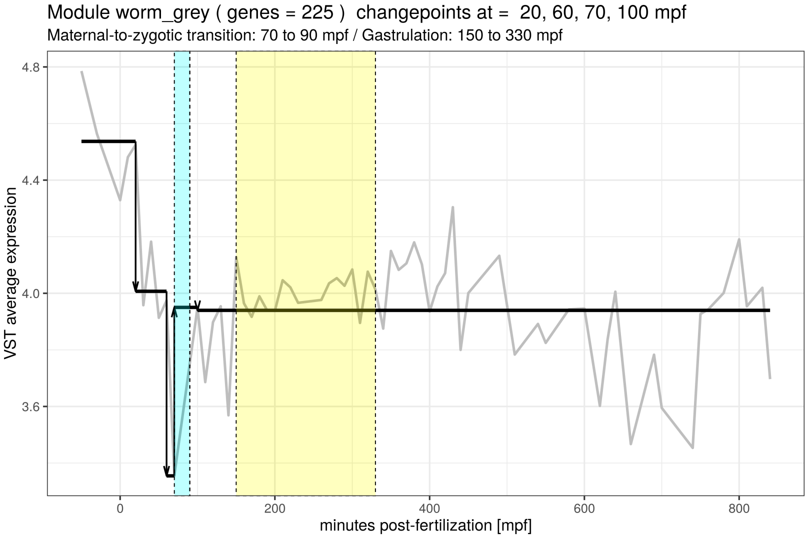


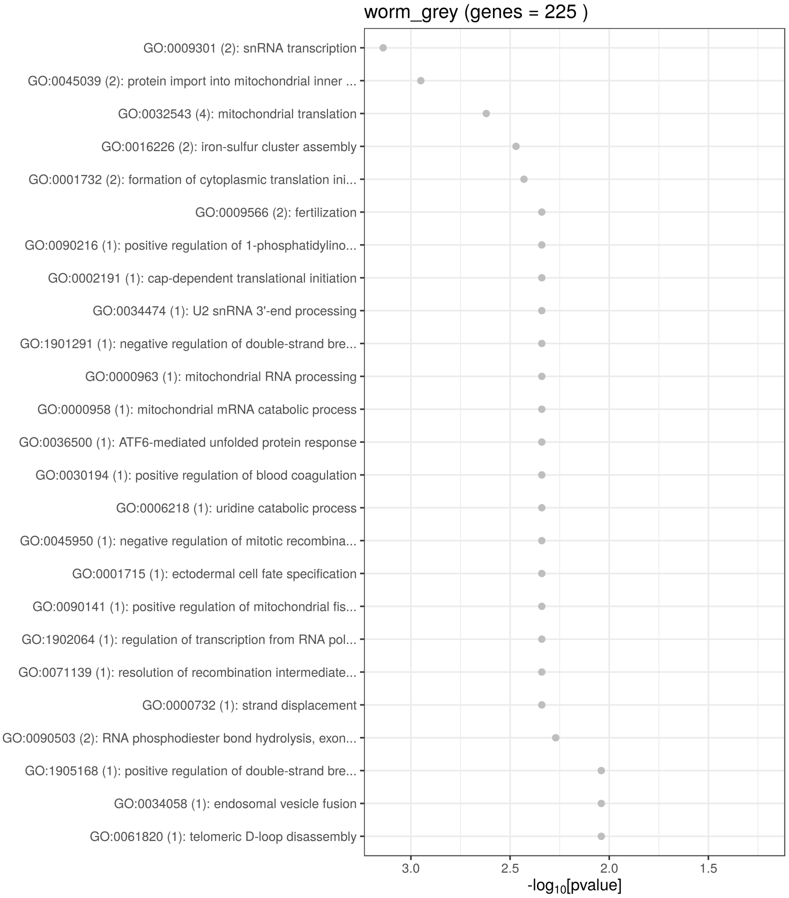



worm_brown (genes = 1,335)
GO Term (number of genes)
GO:0051301 (45): cell division
GO:0016925 (8): protein sumoylation
GO:0006334 (15): nucleosome assembly
GO:0000070 (33): mitotic sister chromatid segregation
GO:0006383 (13): transcription from RNA polymerase III pr...
GO:0006270 (9): DNA replication initiation
GO:0051382 (10): kinetochore assembly
GO:0000398 (34): mRNA splicing, via spliceosome
GO:0048557 (9): embryonic digestive tract morphogenesis
GO:0051321 (42): meiotic cell cycle
12.5
10.0
7.5
5.0
-log$_{10}$[pvalue]

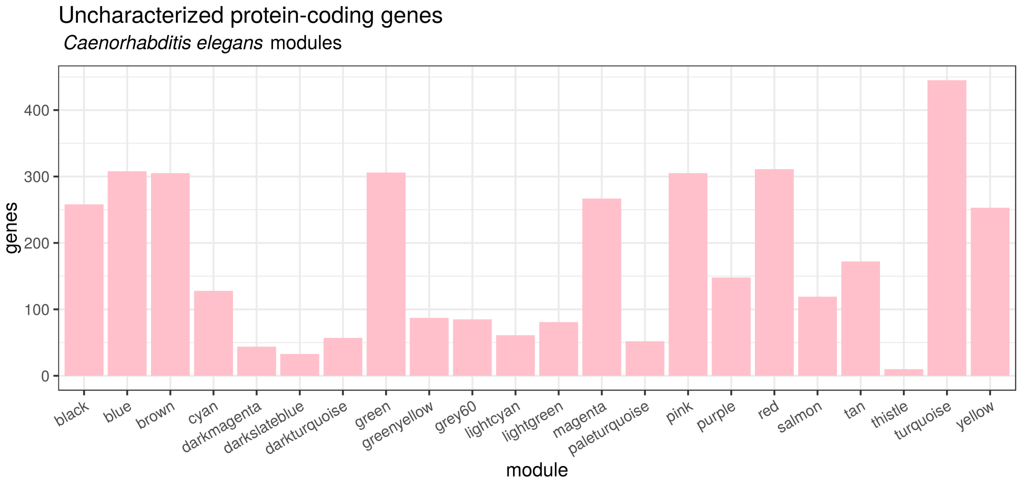

Module worm_brown (genes = 1,335) changepoints at = 230, 430, 640, 740 mpf
Maternal-to-zygotic transition: 70 to 90 mpf / Gastrulation: 150 to 330 mpf
VST average expression
6
5
4
3
0
200
400
600
800
minutes post-fertilization [mpf]

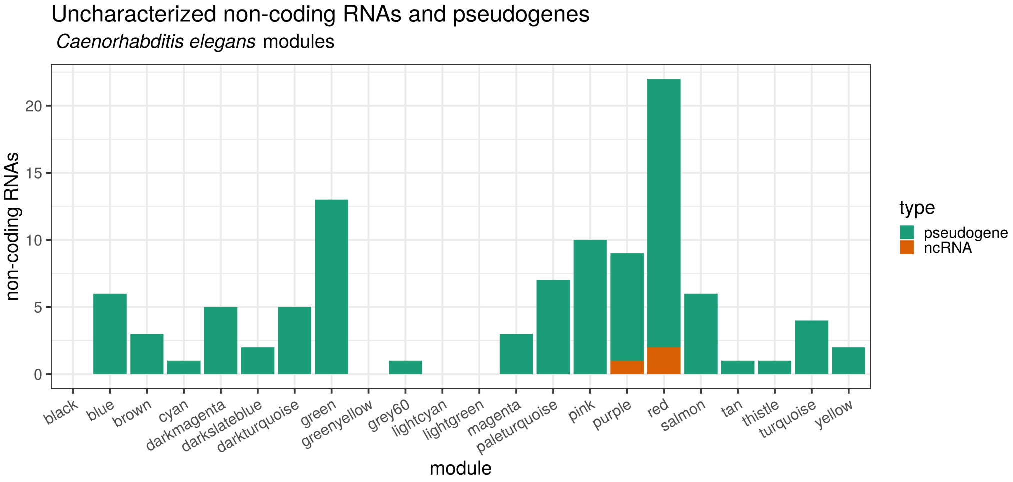
10.2.7 Grey module



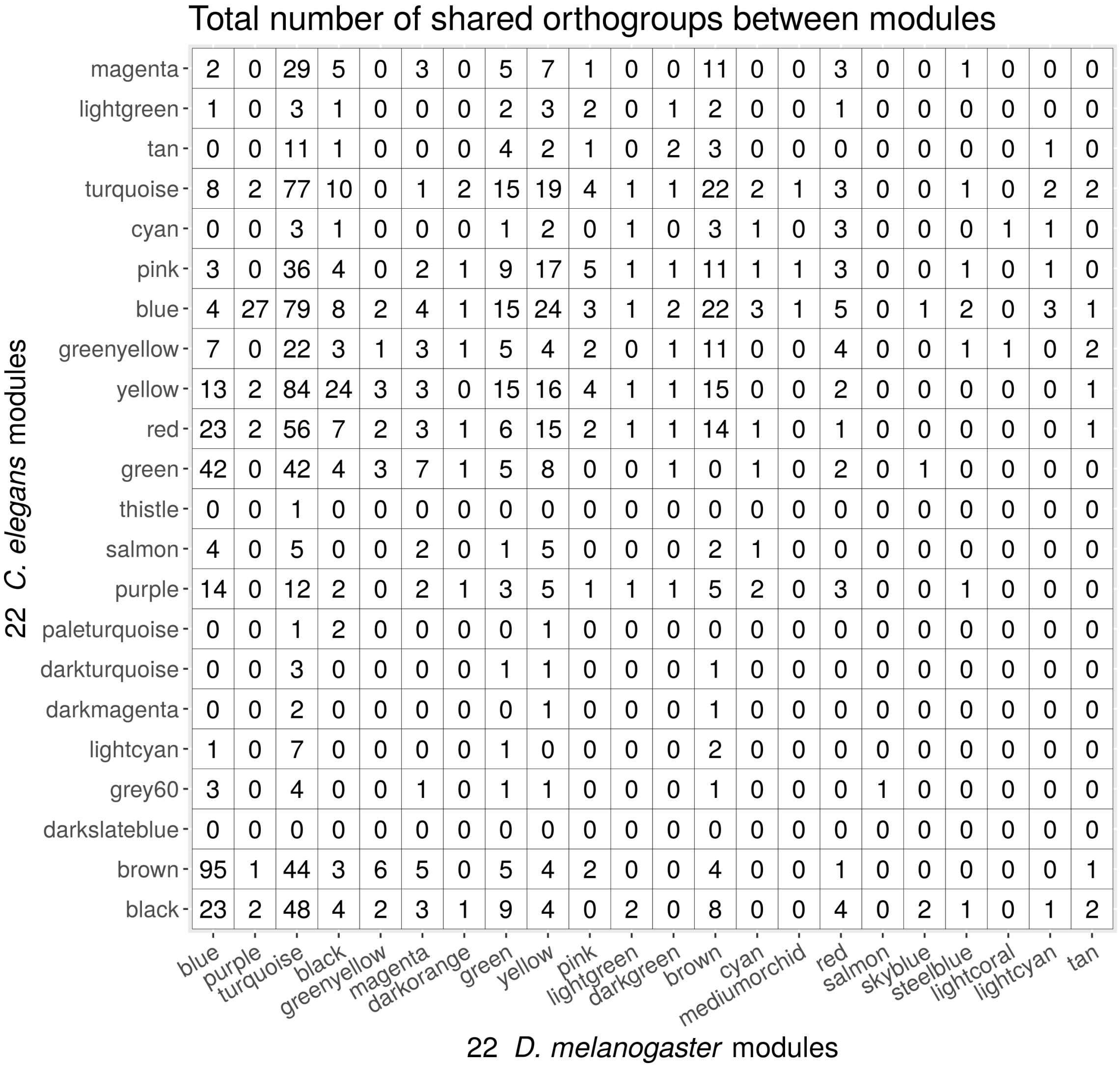


10.2.8 Uncharacterized genes in modules





10.3 Orthogroup content between modules

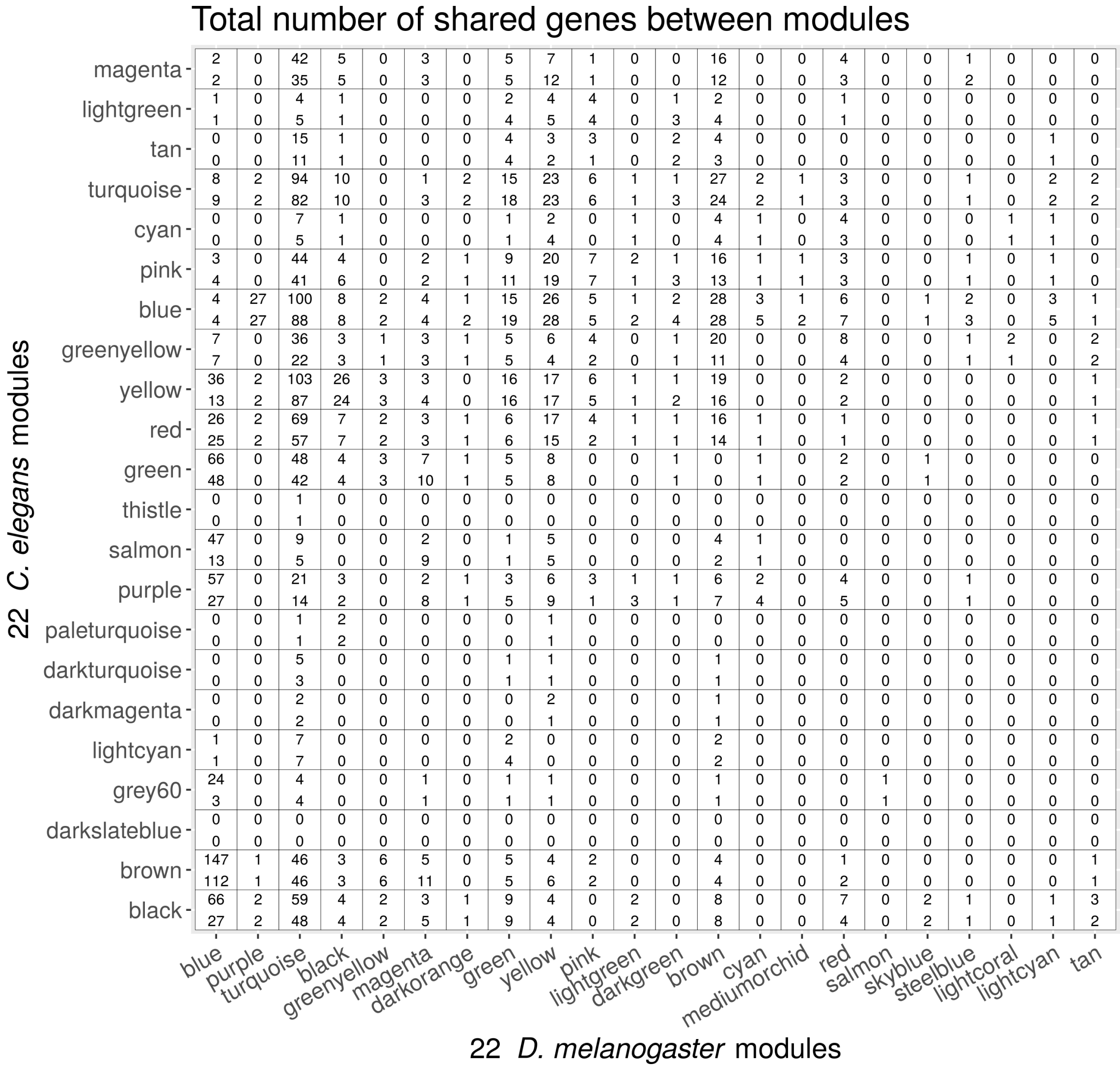


Total number of shared orthogroups between modules

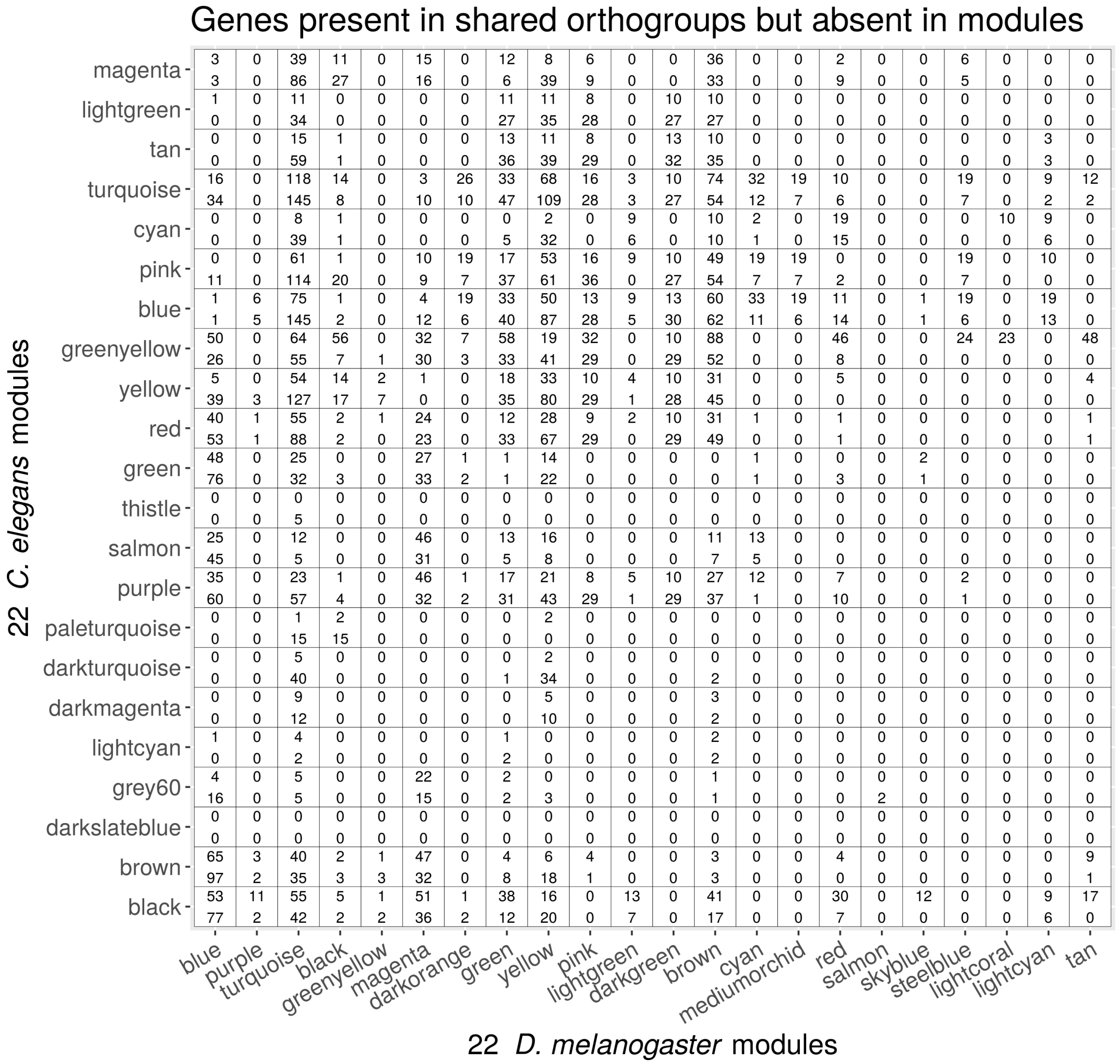
	blue	purple	turquoise	black	greenyellow	magenta	darkorange	green	yellow	pink	lightgreen	darkgreen	brown	cyan	mediumorchid	red	salmon	skyblue	steelblue	lightcoral	lightcyan	tan
magenta	2	0	29	5	0	3	0	5	7	1	0	0	11	0	0	3	0	0	1	0	0	0
lightgreen	1	0	3	1	0	0	0	2	3	2	0	1	2	0	0	1	0	0	0	0	0	0
tan	0	0	11	1	0	0	0	4	2	1	0	2	3	0	0	0	0	0	0	0	1	0
turquoise	8	2	77	10	0	1	2	15	19	4	1	1	22	2	1	3	0	0	1	0	2	2
cyan	0	0	3	1	0	0	0	1	2	0	1	0	3	1	0	3	0	0	0	1	1	0
pink	3	0	36	4	0	2	1	9	17	5	1	1	11	1	1	3	0	0	1	0	1	0
blue	4	27	79	8	2	4	1	15	24	3	1	2	22	3	1	5	0	1	2	0	3	1
greenyellow	7	0	22	3	1	3	1	5	4	2	0	1	11	0	0	4	0	0	1	1	0	2
yellow	13	2	84	24	3	3	0	15	16	4	1	1	15	0	0	2	0	0	0	0	0	1
red	23	2	56	7	2	3	1	6	15	2	1	1	14	1	0	1	0	0	0	0	0	1
green	42	0	42	4	3	7	1	5	8	0	0	1	0	1	0	2	0	1	0	0	0	0
thistle	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
salmon	4	0	5	0	0	2	0	1	5	0	0	0	2	1	0	0	0	0	0	0	0	0
purple	14	0	12	2	0	2	1	3	5	1	1	1	5	2	0	3	0	0	1	0	0	0
paleturquoise	0	0	1	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
darkturquoise	0	0	3	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
darkmagenta	0	0	2	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
lightcyan	1	0	7	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0
grey60	3	0	4	0	0	1	0	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0
darkslateblue	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
brown	95	1	44	3	6	5	0	5	4	2	0	0	4	0	0	1	0	0	0	0	0	1
black	23	2	48	4	2	3	1	9	4	0	2	0	8	0	0	4	0	2	1	0	1	2

22 *C. elegans* modules

22 *D. melanogaster* modules

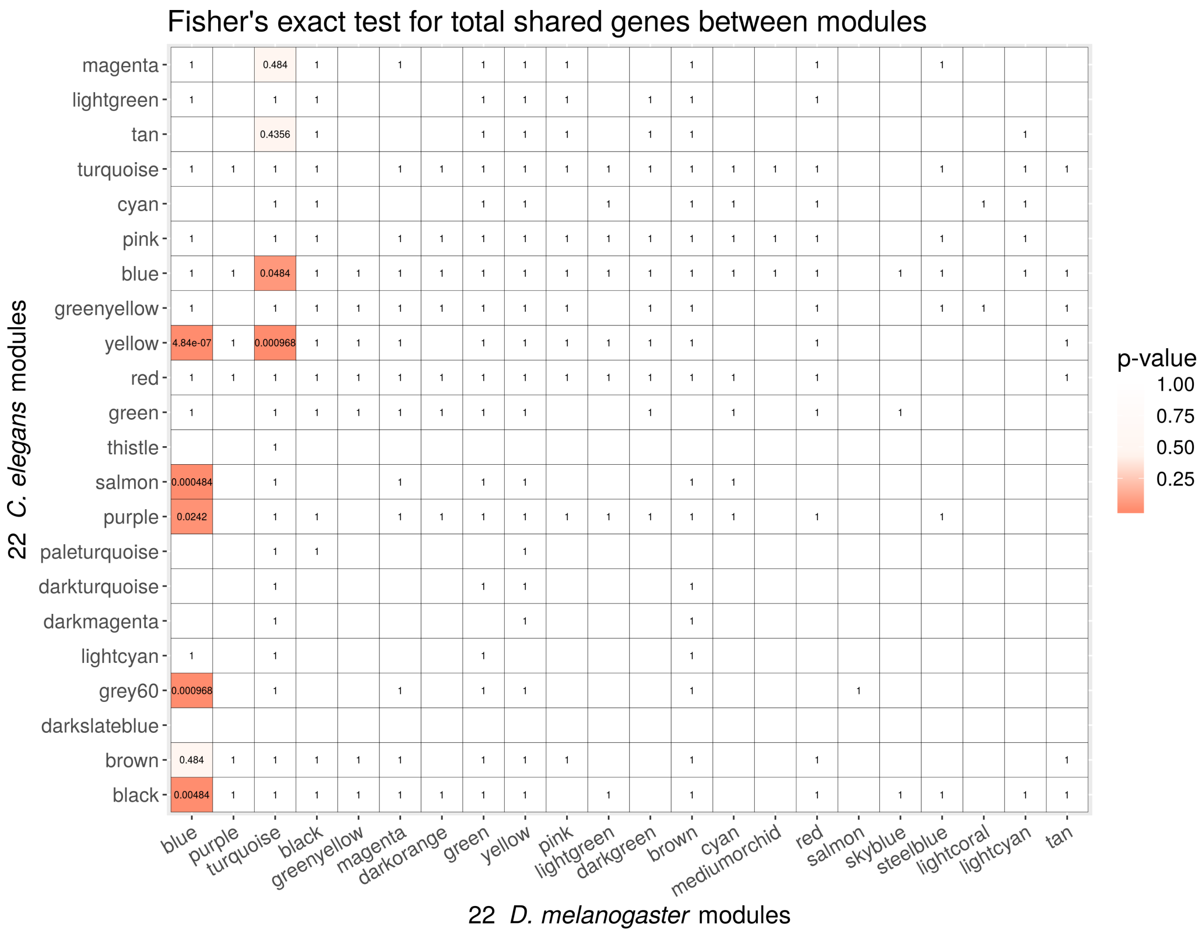


	blue	purple	turquoise	black	greenyellow	magenta	darkorange	green	yellow	pink	lightgreen	darkgreen	brown	cyan	mediumorchid	red	salmon	skyblue	steelblue	lightcoral	lightcyan	tan
magenta	2 2	0 0	42 35	5 5	0 0	3 3	0 0	5 5	7 12	1 1	0 0	0 0	16 12	0 0	0 0	4 3	0 0	0 0	1 2	0 0	0 0	0 0
lightgreen	1 1	0 0	4 5	1 1	0 0	0 0	0 0	2 4	4 5	4 4	0 0	1 3	2 4	0 0	0 0	1 1	0 0	0 0	0 0	0 0	0 0	0 0
tan	0 0	0 0	15 11	1 1	0 0	0 0	0 0	4 4	3 2	3 1	0 0	2 2	4 3	0 0	0 0	0 0	0 0	0 0	0 0	0 0	1 1	0 0
turquoise	8 9	2 2	94 82	10 10	0 0	1 3	2 2	15 18	23 23	6 6	1 1	1 3	27 24	2 2	1 1	3 3	0 0	0 0	1 1	0 0	2 2	2 2
cyan	0 0	0 0	7 5	1 1	0 0	0 0	0 0	1 1	2 4	0 0	1 1	0 0	4 4	1 1	0 0	4 3	0 0	0 0	0 0	1 1	1 1	0 0
pink	3 4	0 0	44 41	4 6	0 0	2 2	1 1	9 11	20 19	7 7	2 1	1 3	16 13	1 1	1 1	3 3	0 0	0 0	1 1	0 0	1 1	0 0
blue	4 4	27 27	100 88	8 8	2 2	4 4	1 2	15 19	26 28	5 5	1 2	2 4	28 28	3 5	1 2	6 7	0 0	1 1	2 3	0 0	3 5	1 1
greenyellow	7 7	0 0	36 22	3 3	1 1	3 3	1 1	5 5	6 4	4 2	0 0	1 1	20 11	0 0	0 0	8 4	0 0	0 0	1 1	2 1	0 0	2 2
yellow	36 13	2 2	103 87	26 24	3 3	3 4	0 0	16 16	17 17	6 5	1 1	1 2	19 16	0 0	0 0	2 2	0 0	0 0	0 0	0 0	0 0	1 1
red	26 25	2 2	69 57	7 7	2 2	3 3	1 1	6 6	17 15	4 2	1 1	1 1	16 14	1 1	0 0	1 1	0 0	0 0	0 0	0 0	0 0	1 1
green	66 48	0 0	48 42	4 4	3 3	7 10	1 1	5 5	8 8	0 0	0 0	1 1	0 0	1 1	0 0	2 2	0 0	1 1	0 0	0 0	0 0	0 0
thistle	0 0	0 0	1 1	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
salmon	47 13	0 0	9 5	0 0	0 0	2 9	0 0	1 1	5 5	0 0	0 0	0 0	4 2	1 1	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
purple	57 27	0 0	21 14	3 2	0 0	2 8	1 1	3 5	6 9	3 1	1 3	1 1	6 7	2 4	0 0	4 5	0 0	0 0	1 1	0 0	0 0	0 0
paleturquoise	0 0	0 0	1 1	2 2	0 0	0 0	0 0	0 0	1 1	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
darkturquoise	0 0	0 0	5 3	0 0	0 0	0 0	0 0	1 1	1 1	0 0	0 0	0 0	1 1	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
darkmagenta	0 0	0 0	2 2	0 0	0 0	0 0	0 0	0 0	2 1	0 0	0 0	0 0	1 1	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
lightcyan	1 1	0 0	7 7	0 0	0 0	0 0	0 0	2 4	0 0	0 0	0 0	0 0	2 2	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
grey60	24 3	0 0	4 4	0 0	0 0	1 1	0 0	1 1	1 1	0 0	0 0	0 0	1 1	0 0	0 0	0 0	1 1	0 0	0 0	0 0	0 0	0 0
darkslateblue	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
brown	147 112	1 1	46 46	3 3	6 6	5 11	0 0	5 5	4 6	2 2	0 0	0 0	4 4	0 0	0 0	1 2	0 0	0 0	0 0	0 0	0 0	1 1
black	66 27	2 2	59 48	4 4	2 2	3 5	1 1	9 9	4 4	0 0	2 2	0 0	8 8	0 0	0 0	7 4	0 0	2 2	1 1	0 0	1 1	3 2

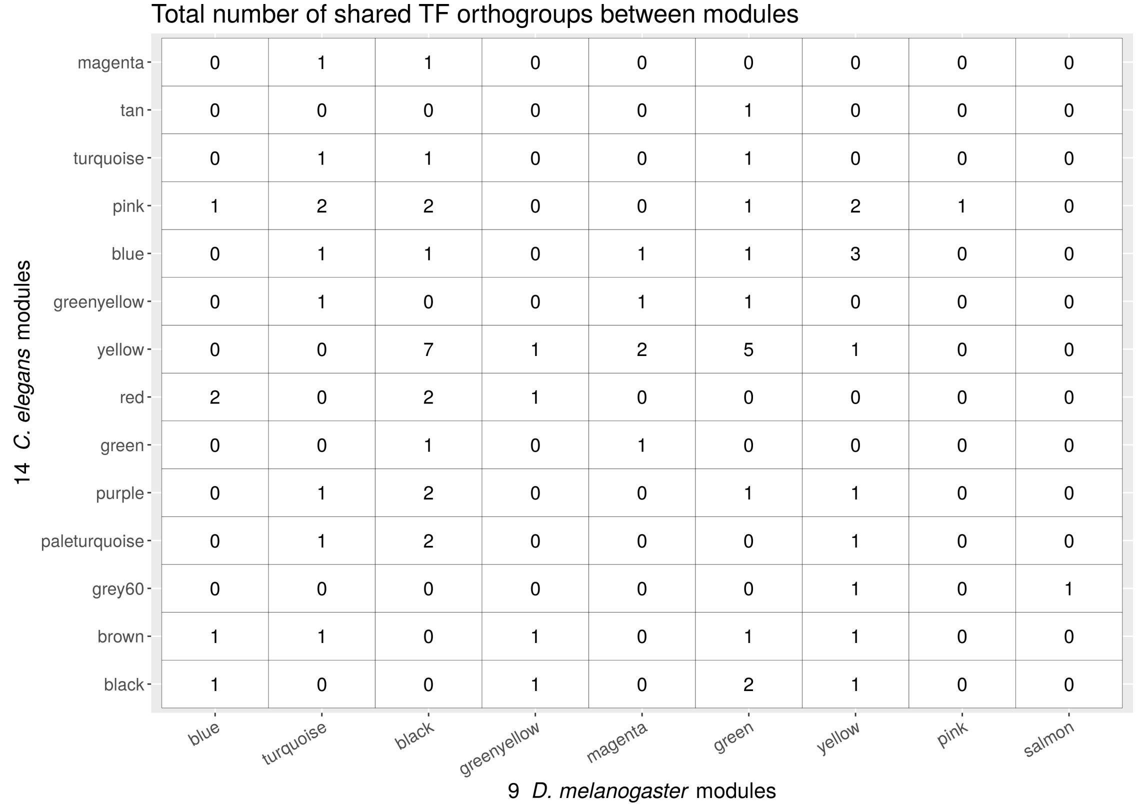


Genes present in shared orthogroups but absent in modules

	blue	purple	turquoise	black	greenyellow	magenta	darkorange	green	yellow	pink	lightgreen	darkgreen	brown	cyan	mediumorchid	red	salmon	skyblue	steelblue	lightcoral	lightcyan	tan
magenta	3 3	0 0	39 86	11 27	0 0	15 16	0 0	12 6	8 39	6 9	0 0	0 0	36 33	0 0	0 0	2 9	0 0	0 0	6 5	0 0	0 0	0 0
lightgreen	1 0	0 0	11 34	0 0	0 0	0 0	0 0	11 27	11 35	8 28	0 0	10 27	10 27	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
tan	0 0	0 0	15 59	1 1	0 0	0 0	0 0	13 36	11 39	8 29	0 0	13 32	10 35	0 0	0 0	0 0	0 0	0 0	0 0	0 0	3 3	0 0
turquoise	16 34	0 0	118 145	14 8	0 0	3 10	26 10	33 47	68 109	16 28	3 3	10 27	74 54	32 12	19 7	10 6	0 0	0 0	19 7	0 0	9 2	12 2
cyan	0 0	0 0	8 39	1 1	0 0	0 0	0 0	0 5	2 32	0 0	9 6	0 0	10 10	2 1	0 0	19 15	0 0	0 0	0 0	10 0	9 6	0 0
pink	0 11	0 0	61 114	1 20	0 0	10 9	19 7	17 37	53 61	16 36	9 0	10 27	49 54	19 7	19 7	0 2	0 0	0 0	19 7	0 0	10 0	0 0
blue	1 1	6 5	75 145	1 2	0 0	4 12	19 6	33 40	50 87	13 28	9 5	13 30	60 62	33 11	19 6	11 14	0 0	1 1	19 6	0 0	19 13	0 0
greenyellow	50 26	0 0	64 55	56 7	0 1	32 30	7 3	58 33	19 41	32 29	0 0	10 29	88 52	0 0	0 0	46 8	0 0	0 0	24 0	23 0	0 0	48 0
yellow	5 39	0 3	54 127	14 17	2 7	1 0	0 0	18 35	33 80	10 29	4 1	10 28	31 45	0 0	0 0	5 0	0 0	0 0	0 0	0 0	0 0	4 0
red	40 53	1 1	55 88	2 2	1 0	24 23	0 0	12 33	28 67	9 29	2 0	10 29	31 49	1 0	0 0	1 1	0 0	0 0	0 0	0 0	0 0	1 1
green	48 76	0 0	25 32	0 3	0 0	27 33	1 2	1 1	14 22	0 0	0 0	0 0	0 0	1 1	0 0	0 3	0 0	2 1	0 0	0 0	0 0	0 0
thistle	0 0	0 0	0 5	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
salmon	25 45	0 0	12 5	0 0	0 0	46 31	0 0	13 5	16 8	0 0	0 0	0 0	11 7	13 5	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
purple	35 60	0 0	23 57	1 4	0 0	46 32	1 2	17 31	21 43	8 29	5 1	10 29	27 37	12 1	0 0	7 10	0 0	0 0	2 1	0 0	0 0	0 0
paleturquoise	0 0	0 0	1 15	2 15	0 0	0 0	0 0	0 0	2 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
darkturquoise	0 0	0 0	5 40	0 0	0 0	0 0	0 0	0 1	2 34	0 0	0 0	0 0	0 2	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
darkmagenta	0 0	0 0	9 12	0 0	0 0	0 0	0 0	0 0	5 10	0 0	0 0	0 0	3 2	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
lightcyan	1 0	0 0	4 2	0 0	0 0	0 0	0 0	1 2	0 0	0 0	0 0	0 0	2 2	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
grey60	4 16	0 0	5 5	0 0	0 0	22 15	0 0	2 2	0 3	0 0	0 0	0 0	1 1	0 0	0 0	0 0	0 2	0 0	0 0	0 0	0 0	0 0
darkslateblue	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
brown	65 97	3 2	40 35	2 3	1 3	47 32	0 0	4 8	6 18	4 1	0 0	0 0	3 3	0 0	0 0	4 0	0 0	0 0	0 0	0 0	0 0	9 1
black	53 77	11 2	55 42	5 2	1 2	51 36	1 2	38 12	16 20	0 0	13 7	0 0	41 17	0 0	0 0	30 7	0 0	12 0	0 0	0 0	9 6	17 0


Fisher's exact test for total shared genes between modules
22 C. elegans modules
magenta
lightgreen
tan
turquoise
cyan
pink
blue
greenyellow
yellow
red
green
thistle
salmon
purple
paleturquoise
darkturquoise
darkmagenta
lightcyan
grey60
darkslateblue
brown
black
22 D. melanogaster modules
blue
purple
turquoise
black
greenyellow
magenta
darkorange
green
yellow
pink
lightgreen
darkgreen
brown
cyan
mediumorchid
red
salmon
skyblue
steelblue
lightcoral
lightcyan
tan
0.484
0.4356
0.0484
4.84e-07
0.000968
0.000484
0.0242
0.000968
0.484
0.00484
p-value
1.00
0.75
0.50
0.25

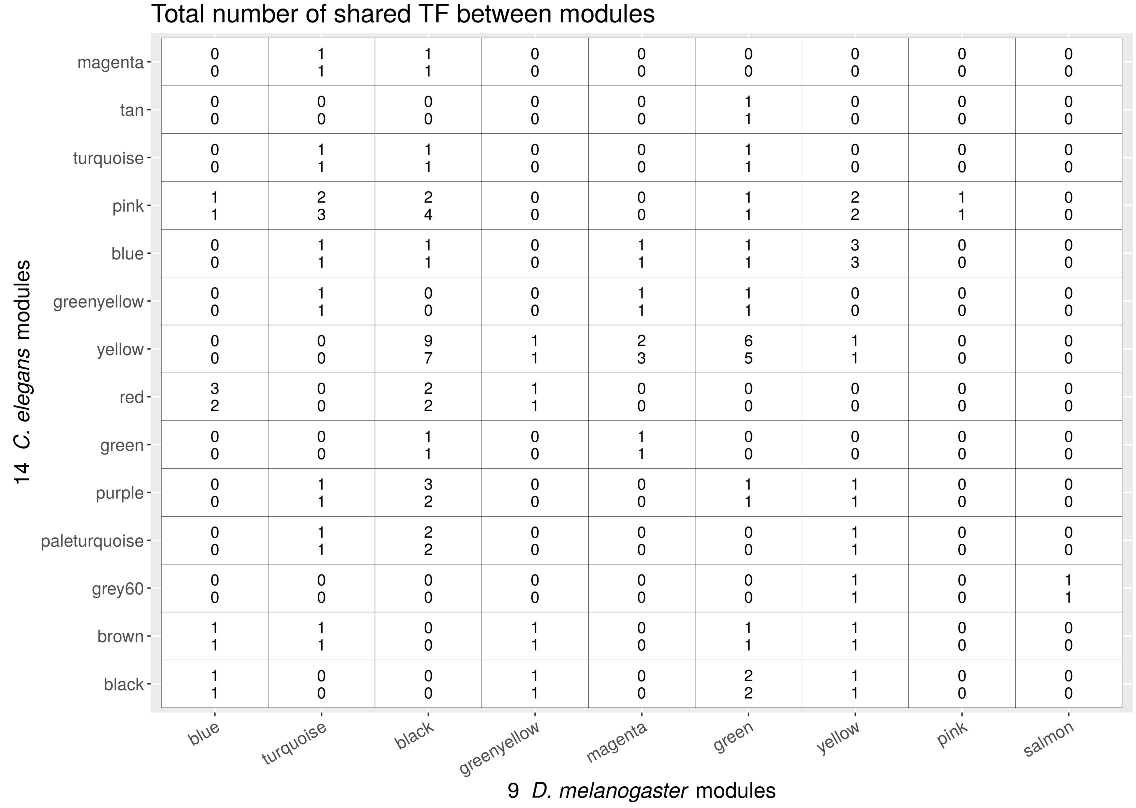
10.4 Transcription factor content between modules



Total number of shared TF orthogroups between modules

14 *C. elegans* modules
9 *D. melanogaster* modules

	blue	turquoise	black	greenyellow	magenta	green	yellow	pink	salmon
magenta	0	1	1	0	0	0	0	0	0
tan	0	0	0	0	0	1	0	0	0
turquoise	0	1	1	0	0	1	0	0	0
pink	1	2	2	0	0	1	2	1	0
blue	0	1	1	0	1	1	3	0	0
greenyellow	0	1	0	0	1	1	0	0	0
yellow	0	0	7	1	2	5	1	0	0
red	2	0	2	1	0	0	0	0	0
green	0	0	1	0	1	0	0	0	0
purple	0	1	2	0	0	1	1	0	0
paleturquoise	0	1	2	0	0	0	1	0	0
grey60	0	0	0	0	0	0	1	0	1
brown	1	1	0	1	0	1	1	0	0
black	1	0	0	1	0	2	1	0	0



Total number of shared TF between modules

14 *C. elegans* modules
9 *D. melanogaster* modules

	blue	turquoise	black	greenyellow	magenta	green	yellow	pink	salmon
magenta	0 0	1 1	1 1	0 0	0 0	0 0	0 0	0 0	0 0
tan	0 0	0 0	0 0	0 0	0 0	1 1	0 0	0 0	0 0
turquoise	0 0	1 1	1 1	0 0	0 0	1 1	0 0	0 0	0 0
pink	1 1	2 3	2 4	0 0	0 0	1 1	2 2	1 1	0 0
blue	0 0	1 1	1 1	0 0	1 1	1 1	3 3	0 0	0 0
greenyellow	0 0	1 1	0 0	0 0	1 1	1 1	0 0	0 0	0 0
yellow	0 0	0 0	9 7	1 1	2 3	6 5	1 1	0 0	0 0
red	3 2	0 0	2 2	1 1	0 0	0 0	0 0	0 0	0 0
green	0 0	0 0	1 1	0 0	1 1	0 0	0 0	0 0	0 0
purple	0 0	1 1	3 2	0 0	0 0	1 1	1 1	0 0	0 0
paleturquoise	0 0	1 1	2 2	0 0	0 0	0 0	1 1	0 0	0 0
grey60	0 0	0 0	0 0	0 0	0 0	0 0	1 1	0 0	1 1
brown	1 1	1 1	0 0	1 1	0 0	1 1	1 1	0 0	0 0
black	1 1	0 0	0 0	1 1	0 0	2 2	1 1	0 0	0 0

D. melanogaster module	Orthologous transcription factors
black	*mirr, tap, kn, Oli, scrt, Poxn, Optix, tup, disco, Hr39, Scr, zfh1, drm, klu, caup*
greenyellow	*twi, MTA1-like, l(1)sc*
magenta	*pros, oc, nerfin-1, Six4, salm*
green	*D19A, Vsx1, unc-4, B-H2, CG12605, Lim1, Pdp1, onecut, elB, B-H1, ara, Eip93F*
yellow	*CG18599, H15, salr, scro, sr, CG2889, HLH4C*
pink	*vri*
salmon	*CG12769*
turquoise*	*Jra, tj, ftz-f1, Fer2, bun, bsh*
blue*	*sc, ato, pita, retn, CG12391, amos*

C. elegans module	Orthologous transcription factors
grey60	*aptf-4, sdz-12*
paleturquoise	*nhr-89, ZK686.5, bnc-1*
purple	*ceh-41, egrh-1, ceh-51, scrt-1*
green	*ceh-36, hinf-1*
red	*hlh-14, lin-32, ngn-1, odd-1*
yellow	*ceh-5, ceh-10, ceh-31, ceh-32, ceh-34, ces-1, egl-46, hlh-8, lim-7, mab-5, pax-1, tlp-1, unc-39, zag-1, irx-1*
greenyellow	*pros-1, sbp-1, ceh-48*
blue	*ceh-24, hlh-15, lin-11, sem-4, unc-3, F19B2.6*
pink	*atf-2, cfi-1, hlh-13, hlh-17, mab-9, nhr-113, nhr-259, hlh-32, atf-8, aptf-1*
turquoise	*mbr-1, klu-2, maf-1*
tan	*unc-4*
magenta	*nhr-277*
black*	*aptf-3, zfp-2, ceh-49, F56D1.1*
brown*	*ceh-21, Y5F2A.4, Y48C3A.12, F10E7.11, ceh-99*