The role of heterochronic gene expression and regulatory architecture in early developmental divergence

Authors: Nathan D. Harry[1], Christina Zakas[1*]

Affiliation: [1] North Carolina State University
Department of Biological Sciences
112 Derieux Place
Raleigh, NC 27607

*Corresponding author
email: czakas@ncsu.edu

ABSTRACT

New developmental programs can evolve through adaptive changes to gene expression. The annelid *S. benedicti* has a developmental dimorphism, which provides a unique intraspecific framework for understanding the earliest genetic changes that take place during developmental divergence. Using comparative RNAseq through ontogeny, we find that only a small proportion of genes are differentially expressed at any time, despite major differences in larval development and life-history. These genes shift expression profiles across morphs by either turning off any expression in one morph or changing the timing or amount of gene expression. We directly connect the contributions of these mechanisms to differences in developmental processes. We examine F_1 offspring— using reciprocal crosses— to determine maternal mRNA inheritance and the regulatory architecture of gene expression. These results highlight the importance of both novel gene expression and heterochronic shifts in developmental evolution, as well as the *trans*-acting regulatory factors in initiating divergence.

INTRODUCTION

Small changes in development can result in vast morphological differentiation and divergence. Through the history of evolutionary developmental biology, researchers have proposed that changes in the timing of traits during development can produce most morphological changes [1–3]. Though this theory has been further refined in the field since Haeckel first coined the term “heterochrony” in development (1875), the timing of developmental changes remains a prominent mechanism of diversification [4–7]. At the molecular level, changes in gene expression timing (**heterochronic** genes) or expression amount (**heteromorphic** genes) can underlie major morphological differences [8]. However, these heterochronic shifts are usually investigated on a per-gene basis to reveal the extent that morphological change is achieved through gene expression perturbation (as reviewed in [2]). Overall, the extent that heterochronic gene expression contributes to developmental differences at the molecular level has not been quantified. Therefore, the underlying regulatory changes that result in gene expression changes also remain elusive. In this study we quantify total gene expression over sequential developmental time to determine the extent that gene expression differences occur between divergent developmental and life-history modes. Furthermore, we use genetic crosses between the developmental morphs to quantify the mode of regulatory change that is responsible for these differences.

Changes in gene expression (heterochrony and heteromorphy) are now well established drivers of both interspecific [9–12] and intraspecific [13,14] differentiation. Knowing the extent that these biological processes are at play is fundamental to understanding developmental evolution [1,7,15–19]. While somewhat subtle differences in gene expression timing and amount can drive morphological changes, *exclusive* genes – where genes are simply turned on or off, and possibly gained or lost in one lineage – are also possible [20,21]. Despite the importance of these mechanisms for developmental and evolutionary change, long-standing questions in the field remain: What are the extents to which modifications of gene expression change developmental programs and morphological outcomes? What are the molecular factors that regulate these changes and how do they act?

One difficulty in determining the genetic basis of gene expression shifts is that it typically involves parsing developmental programs across divergent taxa, and therefore requires querying numerous genetic differences that could have arisen by either selection or drift over millions of years. Creating hybrids across such divergent species presents its own difficulties. Here we determine the extent that gene expression differences contribute to developmental divergence over small evolutionary timescales using a model with two developmental modes within a single species. We examine the prevailing prediction that heterochrony is the primary driver of morphological change and developmental evolution [2,15,22]. While numerous interspecific gene expression studies have assessed the occurrence of both heterochronic genes and exclusive genes [23–26], no study has assessed the contributions of these mechanisms to producing differences in developmental processes. By using an intraspecific model of developmental divergence where genetic crosses are possible, we determine the architecture of regulatory differences that drive gene expression and ultimately life-history differences.

RESULTS

Study system and embryology

We compare the gene expression over developmental time for two morphs of a marine annelid, *Streblospio benedicti*, which has an intraspecific developmental dimorphism. There are two distinct developmental morphs which differ in their egg size, embryological development time, larval ecology and morphology. Offspring are either obligately-feeding *planktotrophic* (PP) larvae or non-feeding *lecithotrophic* (LL) larvae. Despite these developmental differences, as adults they are morphologically indistinguishable, outside of some reproductive traits, and occupy the same environmental niches. The larval traits are heritable, meaning the differences in development are genetic and not plastic [27]. The two morphs have been characterized extensively in terms of life-history and genetic differences [27–32]**.** Intriguingly, crosses between the morphs are viable with no obvious fitness effects, and these F_1 offspring can have intermediate larval traits compared to the parentals [27,32]. Embryological differences between the two types (and thus egg sizes) have been briefly described [33] but here we detail the full time course of embryogenesis from the one cell embryo through the larval phase in detail for both morphs, as well as F_1 crosses (PL or LP) between the morphs in both directions.

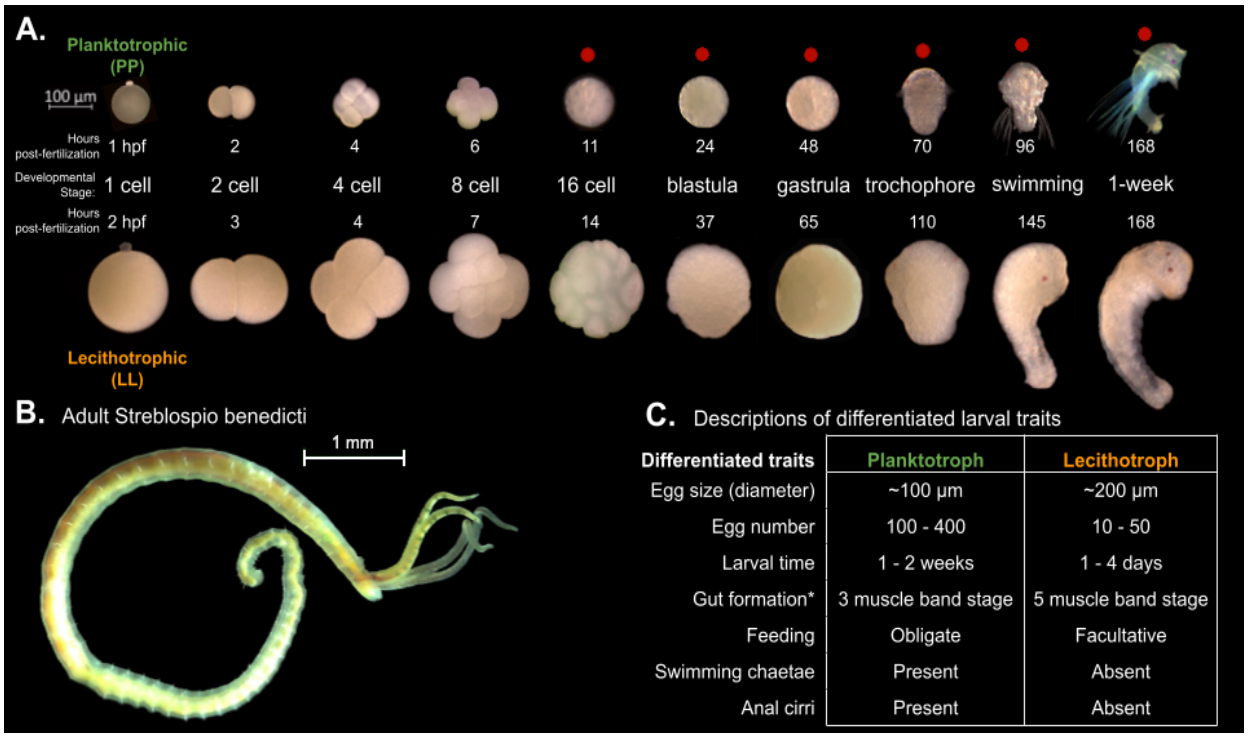
Spiralian animals have a famously conserved pattern of embryological cleavage (reviewed in [34]), so unsurprisingly we find the two morphs proceed through the same developmental stages despite starting from eggs of different sizes (4x volume difference). Despite the similarity in embryo morphology, there are some notable differences between the

two morphs: the absolute time between each stage is shifted such that the LL embryos take longer to reach an equivalent larval stage, which is expected as larger, more yolky cells can take longer to divide [33]. At the swimming larval stage there are some notable morphological and behavioral differences (reviewed in [32], Figure 1), notably the PP larvae are obligately feeding and have feeding structures, while the LL do not. While the embryological stages may not be morphologically different other than size, it is reasonable to expect that they could be expressing different genes or changing the relative timing of expression to produce the morphological and behavioral differences seen in the larvae.

Total RNA Expression Analysis

We measured total gene expression from the six developmental stages using RNAseq with at least four biological replicates per morph at each stage (Figure 2A). Using the full dataset, the first two principal components (PCA) show most of the variance in total gene expression is due to developmental stage and morph (PC1; Figure 2B) As expected the LL individuals appear to transcriptionally fall behind the PP offspring at each morphologically defined, equivalent developmental stage. We expect this pattern as LL offspring develop more slowly to reach these first developmental stages but reach the juvenile stage more quickly in absolute time than PP offspring, which develop in the water column for a longer period (Fig 1). Notably, pre-gastrulation development is distinctly separated from post-gastrulation by the second principal component.

As expected for an intraspecific comparison, most genes are conserved, having the same expression levels at all stages in both morphs. Only 36.2% of all expressed genes are significantly differentially expressed (DE) between PP and LL at any stage in this dataset. We find that early in development over a third of these DE genes are significantly different between the morphs, but these differences tend to be quite small in magnitude. At gastrulation the number of significant genes decreases to less than 5% of the total DE genes, however these remaining expression differences are much larger in magnitude (Figure 2C). It appears that the two morphs are more functionally distinct during early development, likely because of the different metabolic requirements imposed on them by the differences in maternal egg provisioning [32,37–39].



Differentiated traits	Planktotroph	Lecithotroph
Egg size (diameter)	~100 µm	~200 µm
Egg number	100 - 400	10 - 50
Larval time	1 - 2 weeks	1 - 4 days
Gut formation*	3 muscle band stage	5 muscle band stage
Feeding	Obligate	Facultative
Swimming chaetae	Present	Absent
Anal cirri	Present	Absent

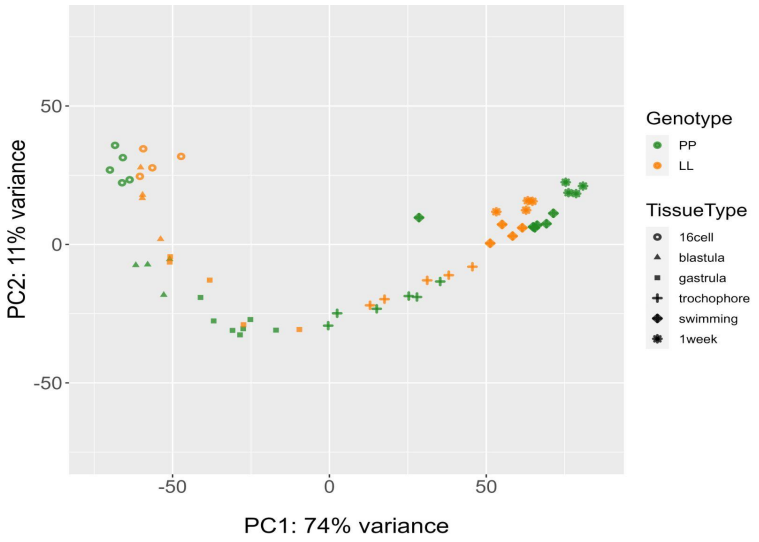
Figure 1: Embryology and development of S. benedicti ***(A)*** *Timeline of early development of the two morphs. The 'swimming' stage is the onset of swimming ability in both morphs. The '1-week' old stage (7 days post fertilization) occurs approximately 1-4 days following larval release from maternal brood pouches under natural conditions. PP larvae at this stage were not fed. Stages used in this study are marked with a red dot.* ***(B)*** *Adult male. The adults of both morphs are indistinguishable outside of reproductive traits.* ***(C)*** *Table of larval traits. LL offspring are facultatively feeding, meaning that feeding is not obligatory, but they are capable of feeding [35]. Gut formation timing [36] where the 3-band stage is our swimming stage, and the 5-band stage is shortly before our 1-week stage.*

A.

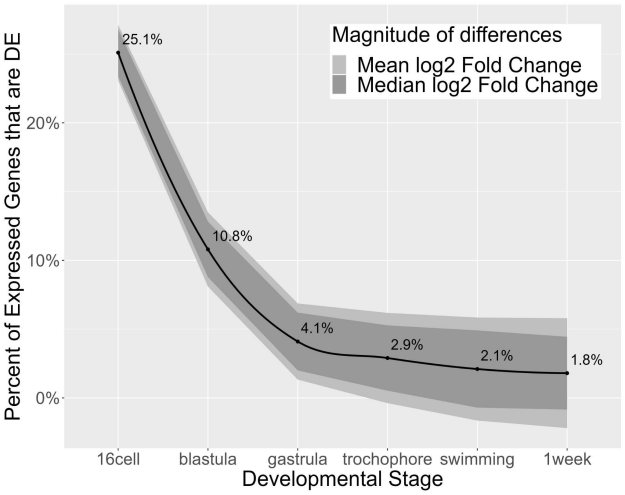
Cross ID	16cell	blastula	gastrula	trochophore	swimming	1-week	complete set
Bay2Z	23			16			
BayGG			9	8	8		
BayHH		11	18	13	15		
BayII			37	13			
BayLL					28	10	
BayOO	19						
BayPP	14	20	7	17	31	10	✓
BayRR	13		13				
BayVV	17						
BayXX	13	11	12	16	60	14	✓
BayYY	18	20	26		15	10	
Total	**7**	**4**	**7**	**6**	**6**	**4**	**2**
F1Bay1	14	23	24	15	11	18	✓
F1Bay2	16	13	51	12	18	15	✓
F1Bay3	9		10	24	16	21	
F1Bay4		19		21			
F1Bay5		21	12	12	9	12	
F1K	12	14	14	9			
F1LB2			12		21		
F1LB3					23		
F1LB4	11		10		7		✓
F1LB5			18		22		
F1LB6	14		13		8		✓
F1LB7	116		122		10		✓
F1LB8					10		
F1LB9	15		15		4		✓
Total	**8**	**5**	**11**	**6**	**12**	**4**	**6**
LBA	13		13	29			
LBnewJ	23	14	9	23	27	19	✓
LBnewL	15	13	10	6	12	14	✓
LBnewO	17	44	10	18	11	10	✓
LBnewQ			33			17	
LBnewS		62		21	13		
Total	**4**	**4**	**5**	**5**	**4**	**4**	**3**

Figure 2: DE genes across six developmental stages in the two developmental morphs. ***(A)*** *Sequenced libraries. Colors indicate morphs: Green = PP, Orange = LL, Purple = (F_1) PL, Pink = (F_1) LP. Numbers in boxes are the sample's sequencing depth in million reads. Total number of replicates for each timepoint are in bold. Samples for which all six stages were collected from the same cross are indicated as 'complete" with a check mark.* ***(B)*** *Principal component analysis of PP and LL samples. PC1 represents developmental time, whereas PC2 separates pre-gastrulation from post-gastrulation* ***(C)*** *DE between morphs over development. Black line is the percentage of total expressed genes that are DE at each stage, ribbon displays the relative mean and median log2 fold-change between morphs for those genes that are DE.*

B.



C.



Differential Expression Analysis

We use our RNAseq time course dataset to quantify the extent that modifications to gene expression timing and amount contribute to developmental differences. We define **heteromorphic** genes as homologous genes whose expression differs significantly between larval morphs, but the pattern of expression (expression profile) does not change (Figure 3). The significant difference in heteromorphic genes could be at only one or a few discrete time points. **Heterochronic** genes are those whose overall expression pattern and timing change between the developmental morphs, as discussed in the clustering algorithm below. We categorize genes that are only expressed in one morph over the full time course of development as **exclusive** genes. These genes are exclusive to only one developmental type.

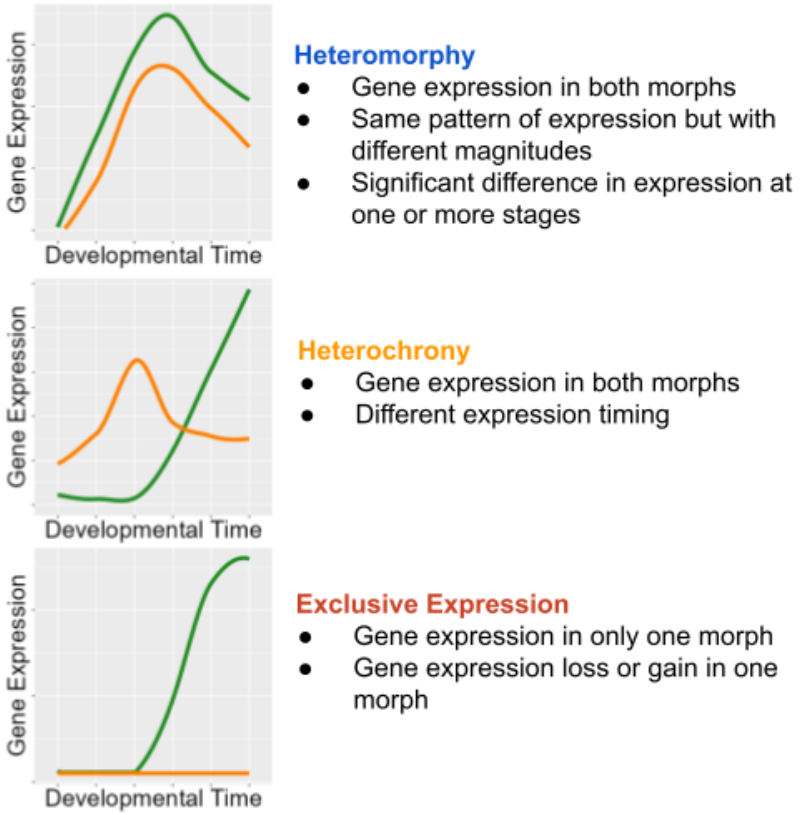


Figure 3: Classifications with examples of gene expression categories. Functional heterochrony is a change in the developmental stage at which a particular pattern of gene expression appears instead of a difference in absolute time.

To differentiate heterochronic shifts in gene expression from heteromorphic ones, we clustered all gene expression patterns from PP into a representative set of expression profiles using Mfuzz (v2.60.0). (No additional clusters were found when using LL). We selected six clusters based on the criteria that the average correlation between cluster pairs increased with each additional cluster, such that the final clusters generated sufficiently represented the diversity of gene expression patterns (Figure 4). We then assigned each gene to a cluster in the PP and LL datasets independently. A gene's expression is considered heterochronic when it appears in one cluster for PP and a different cluster for LL.

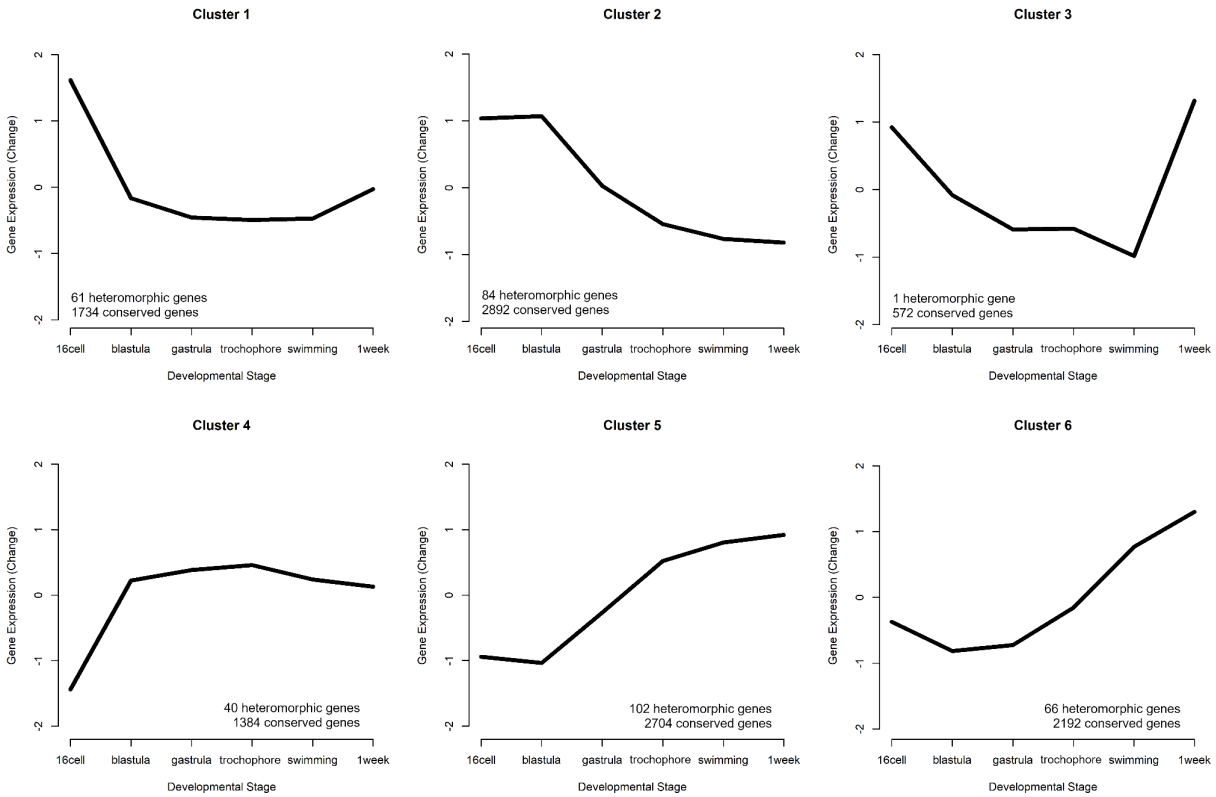


Figure 4: Six gene expression profiles representing the patterns of expression. The number of conserved genes and exclusive genes that match each cluster is listed. Clusters 2 (early expression) and 5 (late expression) represent the most genes.

We identified 354 genes from our set of DE genes (45.9%) where both the PP and LL expression pattern matched the same cluster. (These are heteromorphic genes and are significantly DE in at least one developmental stage, but do not have different profiles of expression between PP and LL). Approximately half of these are assigned to clusters 2 and 5. Cluster 2, which shows a pattern of maternal transcript degradation with no subsequent zygotic expression, is likely to contain genes associated with embryogenesis that are shared by both morphs. Cluster 5 shows a pattern of largely post-gastrulation zygotic gene expression. Based on this pattern, these genes are likely to be associated with shared larval features. Although, there are some late-appearing heteromorphic genes that are associated with discrete larval differences: “chitin catabolic processes” for example, are overexpressed in PP at the swimming stage (when they grow swimming chaetae) and are lowly expressed in LL which do not make swimming chaetae (Supplemental file).

Next, we identified heterochronies, which we define as a gene that switches clusters – and therefore expression profiles – between PP and LL. There are 224 heterochronic genes, which is 29% of DE genes. As we reduced the complexity of the entire gene expression dataset into six clusters, some genes with similar profiles in both PP and LL were ultimately assigned to different clusters, generating a false positive. To ensure against these false positives, we filtered

genes by Pearson correlation between sample means in PP and LL. Genes with a correlation r > 0.85 were not counted as heterochronic.

Heterochronies are a change in the gene expression profile depending on developmental morph, but how different are the patterns of heterochronic changes? Parsimoniously, we would expect most heterochronies to have a similar overall shape that is simply shifted by a stage or two (like a switch from cluster 1 to cluster 2; Figure 4). Examining the six clusters above, it is obvious that some expression profiles have similar trajectories (Cluster 5 and 6 for example both show increasing expression post blastula), while others are essentially inverted expression patterns (Cluster 1 and 4 for example). Clusters are ordered in a correlation heatmap (Figure 5A) from early gene expression to late gene expression. We find that most genes that switch expression profiles between PP and LL do so between the most similar clusters, and thus are near the diagonal. But where we see opposite profiles, the trend is early expression in PP genes compared to LL (top left corner). This is expected given the delay in embryogenesis for LL compared to PP. That said, many genes are expressed earlier in LL compared to PP (bottom right orange triangle in Figure 5A) but most of these are relatively small shifts.

We quantify the magnitude of change in gene expression between PP and LL independently of the cluster assignment by estimating the Pearson correlation directly. This shows that most heterochronies are minor changes in the timing of gene expression. There is a left-tailed distribution indicating that a few genes have a large difference in expression patterns (like a profile inversion) between PP and LL (Figure 5B).

Gene ontology (GO) enrichment tests show heterochronic genes that are expressed earlier in PP are functionally enriched for mesoderm specification, cell fate specification pathways and several organogenesis pathways (others include BMP signaling, mesoderm specification, intestinal epithelial cell differentiation, and gene silencing; Supplemental file). Notably, heterochronic genes which are shifted earlier in LL are enriched for many metabolic functions. The shift in expression of metabolic genes is consistent with the life-history and feeding differences between the larvae, and it may be that most of what is differentiating a PP and LL larvae is the onset of feeding and gut related developmental programs.

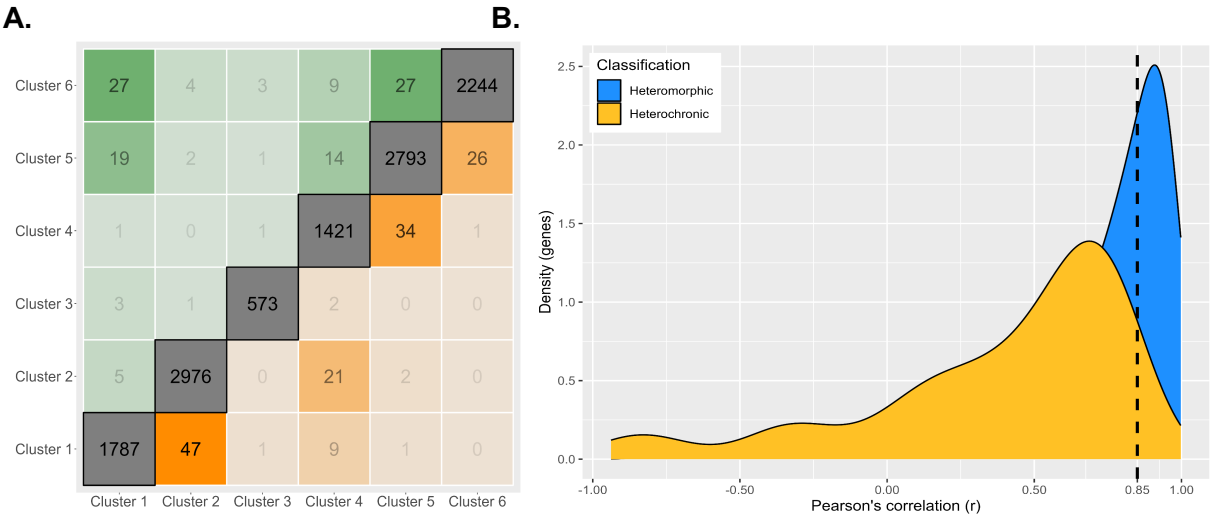


Figure 5: Magnitude of expression profile changes. ***(A)*** *Number of switches between each cluster. Clusters are arranged on the axes in order from high early developmental expression*

(left/bottom) to high late developmental gene expression (right/top). The diagonal represents genes which are assigned to the same cluster in both PP and LL samples, boxes near the diagonal are switches between similar cluster profiles, and boxes furthest from the diagonal are extreme cluster switches. Genes expressed earlier in PP are above diagonal (green) and earlier in LL are below (orange). ***(B)*** *Density plot of genes correlations between PP and LL. Heterochronic genes are yellow and heteromorphic genes in blue. Most heterochronies are minor changes in gene expression timing but the left-tailed distribution shows that few genes have large differences in timing.*

Exclusive genes, by definition, have no expression in one morph, and cannot be assigned to a cluster. To quantify exclusive genes, we use a data-adaptive method (DAFS) to calculate expression thresholds for each sample, below which genes are considered non-expressed. Using this approach, we identify 195 genes (25.3% of DE genes) which are only expressed in one morph (Figure 6). We find considerably more P-specific genes (150 genes in PP vs 45 genes in LL), which tend to be expressed in later developmental stages. These include genes which code for proteins such as Fibropellin, Forkhead-box, Crumbs-like, Notch-2, and many zinc-finger proteins. While the function of these genes during development remains to be determined, it is possible that they maintain PP-specific larval traits.

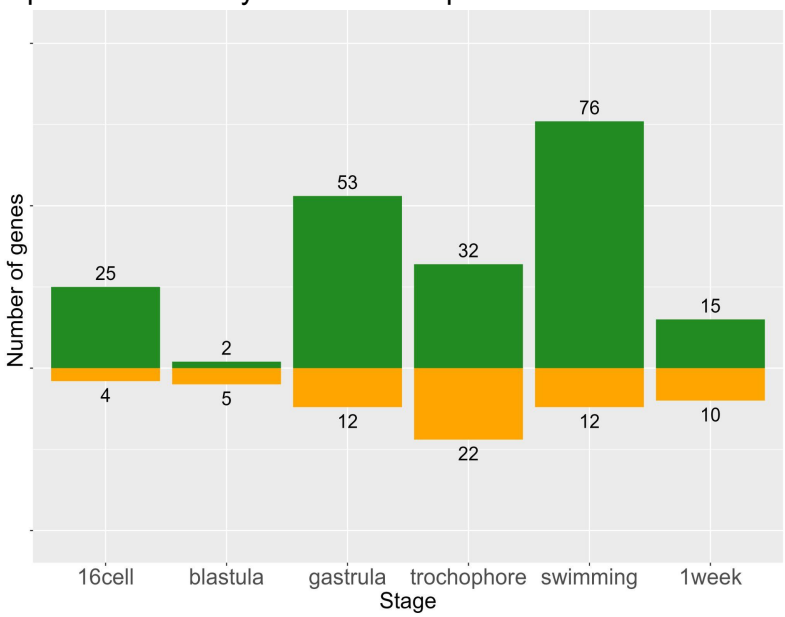


Figure 6: Exclusive gene expression by stage. These genes are never expressed in the opposite morph, but may be expressed at 1-6 stages within a morph.

Expression of functionally distinct categories of genes may evolve by different mechanisms. Exclusive genes may be involved in very different functional processes than heterochronic genes in this system. GO enrichment tests for exclusive genes expressed in PP found a few functions related to cell fate specification and signaling pathways (GO enrichment tests for biological process, $p < 0.05$; no significant functional enrichments for LL-specific genes), but a distinct lack of genes involved in metabolic processes. This is because metabolic pathways genes shifted earlier in the development of LL offspring are still required in both

offspring morphs. But the exclusive genes are not required for development and may be modified by different evolutionary mechanisms. These findings also imply that PP embryos require a few specific cell types to produce a planktotrophic larval form that are reduced or lost in LL.

Overall, we find that heteromorphic changes in expression accounts for most gene expression differences, which is expected as this is the smallest change in gene expression of the three categories. Gene expression need only be significantly different at one stage to fit this category, and such differences may not necessarily translate into biological differences in development. Exclusive and heterochronic changes make up very similar proportions of the remaining differences (Figure 7). This demonstrates that while true heterochronies are common, they are not the main driver of gene expression differences in development.

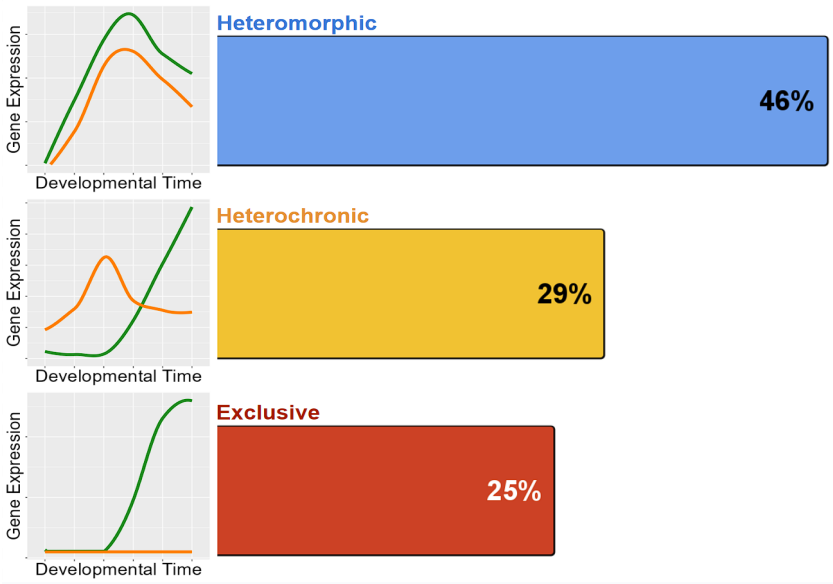


Figure 7: Distribution of the 772 genes with differentiated expression between morphs. Heteromorphy makes up nearly half of all differentiated genes, while heterochrony and exclusive genes split the remainder nearly equally.

Gene expression of genetic crosses

Because we are using an intraspecific comparison, we can extend our analysis to understand the regulatory architecture behind these expression differences. We included RNAseq time course data for offspring from reciprocal crosses between the two morphs, meaning PP and LL parents were crossed in both directions alternating the role of the mother. Cross of the two developmental morphs to produce F_1 offspring, which can have a range of intermediate traits, but typically closely resemble their mother's phenotype [30] . This is particularly useful to disentangle maternal effects, as both F_1s (PL and LP; mother's genotype is listed first) are heterozygotes, but they originate from different egg sizes and mothers with different genetic backgrounds. F_1s allow us to identify the regulatory architecture underlying DE, and to assess the impact of maternal background on gene expression.

F_1s have a general pattern of intermediate expression values but have more variability within replicates and clear cases of outliers (Figure 8). They have considerably more variability

across replicates, and some genes are *misexpressed*— where the F_1 expression value is outside the range of the parental difference (Figure 8). Misexpression is reported in hybridization studies and is thought to be the result of epistatic interactions between divergent genomes that reduce the fitness of offspring contributing to sympatric speciation [40,41]. Whether misexpression affects the fitness of the embryos in *S. benedicti* is unclear, although F_1s typically develop normally in the lab and no systematic reduction in viability has been observed. In previous studies of gene expression in eggs of the two morphs, we saw a similar pattern where F_1 eggs typically have intermediate expression values compared to parental genotypes, but high levels of misexpression [39].

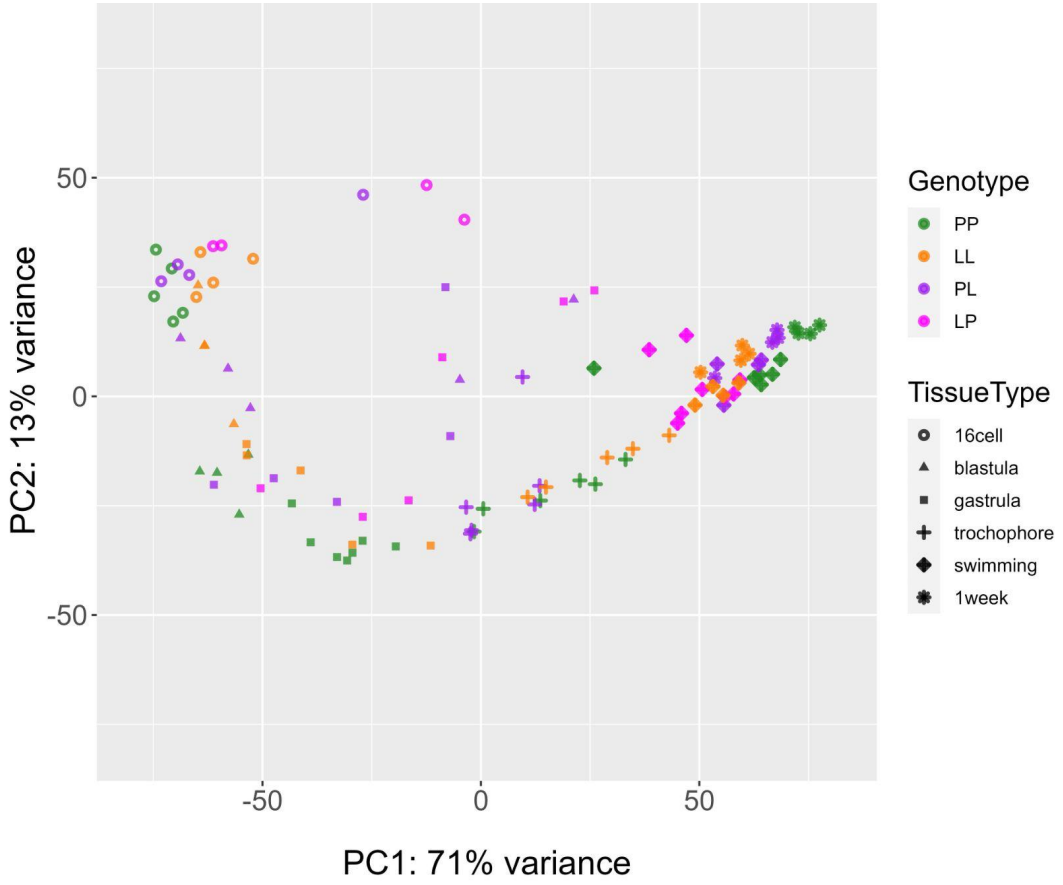


Figure 8: Principal component analysis (PCA) of gene expression including F_1 (PL or LP) offspring. F_1s have more misexpression in early stages but converge after the swimming stage and become intermediate to PP and LL samples.

We find that most genes (>8,000 genes on average, out of 14,327 genes expressed) have extremely conserved expression patterns among PP, LL, and F_1 embryos at each developmental stage. The high variability and misexpression of F_1s negatively impacts our statistical power by introducing variation and results in fewer genes being confidently assigned gene regulatory mechanisms. As a result, almost no genes can be identified as having a dominant or additive inheritance pattern, while 150 genes are identified as overdominant over the course of development. Notably, very few exclusive genes are expressed at a significant

level in either F_1s [between 0 (PP) and 3 (LL) genes], strongly suggesting negative regulation of expression that acts in *trans* for these genes.

Regulatory architecture

We leverage the F_1 offspring to dissect the regulatory architecture underlying developmental gene expression differences; we use allele-specific expression patterns in F_1 offspring– tracking the expression of the maternal or paternal allele– to assign gene regulatory differences as either *cis* or *trans*-acting modifications (or both; [42]). Typically this approach is used in hybrids [43–47], but we have adapted it to intraspecific F_1 offspring to test whether early divergence is consistent with predictions in the literature [39].

We determine the regulatory architecture underlying DE by calculating P and L-allele specific expression in reads from F_1 samples and assigned primary regulatory modes according to established criteria [44,45,47,48]. As this is an intraspecific comparison with little fixed genomic differences between the morphs, not all genes have alleles that can be assigned parentage; however, a set of ~two-thousand DE genes have morph-differentiating SNPs enabling the parental assignment of F_1 transcript reads. The regulatory modes of these DE genes are assigned as "*cis*", "*trans*", "cis + trans", "cis x trans" or "compensatory" (**STable 1** has formal criteria). Of this set, misexpression in F_1 offspring causes many genes to receive ambiguous assignments. As a result, only 143 genes have informative regulatory mode assignments. While this is a small number, we expect that the genes we can classify are proportionally representative of the remaining regulatory architecture.

The primary mode of regulatory change throughout development is *trans*-acting (Figure 9). No differences are caused by purely *cis*-acting regulatory modifications, though we did find some cis+trans and cis-x-trans interactions. We found that the number of genes with compensatory regulatory modifications increases sharply after gastrulation, indicating that gene expression may be more tightly controlled past that point. Studies across Drosophila species, for example, have indicated that maternal effects are regulated differently than zygotic effects, but in this case there are larger *trans*-acting factors in early (maternally controlled) developmental stages [49]. We do not see such a clear effect; *trans*-acting factors make up most of the regulatory architecture throughout development. Interestingly, this result is different from previous studies of egg mRNA expression of this species, where *cis* and *trans*-regulatory modifications were found at similar rates for maternal gene expression [39]. Comparison across the egg (maternal) and embryo (maternal and zygotic) regulatory landscape shows greater *cis*-acting regulation in the maternally expressed genes. These results demonstrate the possibility that maternal and zygotic gene regulatory architecture evolve through distinct mechanisms and on separate timescales.

We tested the possibility that the reciprocal F_1s generated in the study (for the 3 stages that have both LP and PL samples) might have different regulatory mode changes due to parental effects. Splitting our analyses into separate PL and LP F_1 parental arrangements and assessing the mode of regulatory change in each did not, however, yield significantly different results - only even fewer total genes able to be assigned a mode.

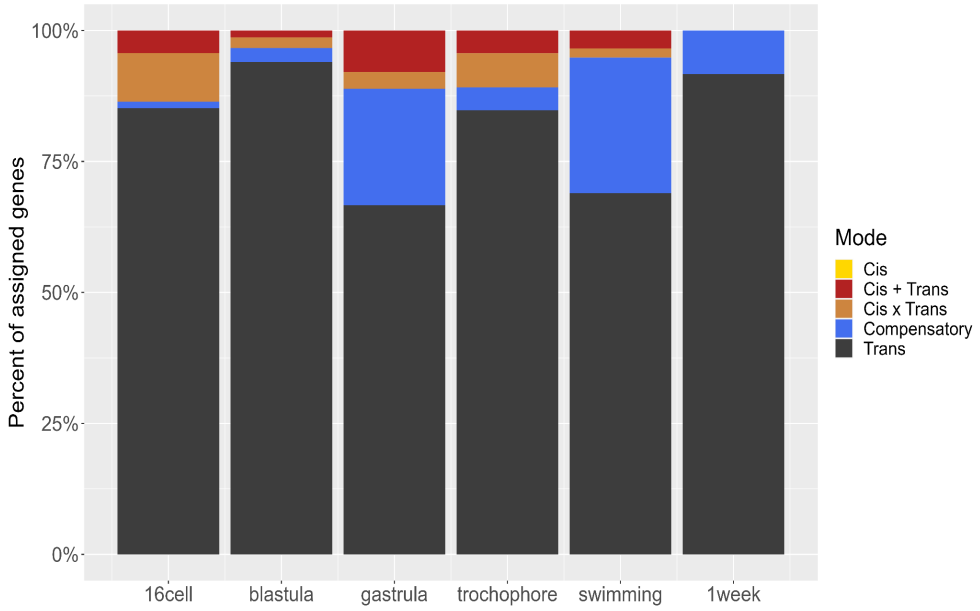


Figure 9: Distribution of regulatory mode over time. The majority of regulatory changes act in trans, and compensatory changes emerge after gastrulation. 162 genes are assigned.

We also investigated parent-of-origin effects on gene expression using F_1s. These occur when an allele's expression is at least partially dependent on which parent contributed that allele. For instance, maternal effects may involve polymorphisms that affect the development of an offspring only when contributed as part of the maternal genome (and which impose no developmental variation when contributed by the paternal genome). These effects are a form of epigenetic inheritance which can allow for complex, adaptive maternal-zygote interactions such as certain metabolic genes being specifically activated in the offspring to match the nutrient content of its mother's egg. Due to animal constraints where LL mothers produce far fewer offspring per clutch (10-40 on average), we were only able to collect samples for F_1 offspring with L mothers (LP) for three developmental stages (16-cell, gastrula, and swimming). All six time points were sampled for F_1's with PP mothers (PL). To find parental effects (maternal or paternal) we conducted a differential expression analysis contrasting the PL samples to the LP samples. Tests for differential expression show that many significant differences in gene expression between the F_1 morphs arise at early stages (ostensibly due to the presence of maternal transcripts inherited through the egg) and then drop sharply after gastrulation (Figure 10; SFigure 5). This indicates that while a few parental effects on gene expression may persist throughout development, most parental effects are limited to early embryogenesis, which is consistent with expectations for maternal transcription degradation.

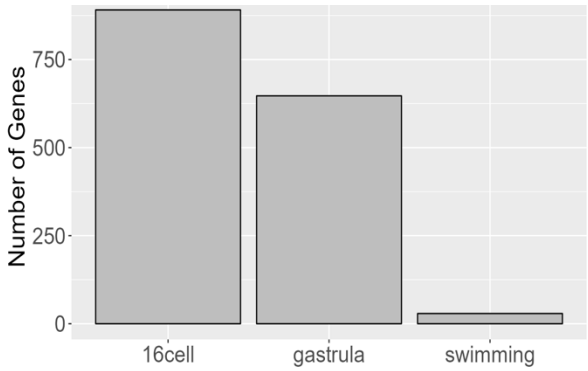


Figure 10: *Genes with parental effects. Early developmental stages have significantly more genes, likely due to maternal transcript differences. Maternal transcripts are degraded around the gastrula stage and any remaining parental effects observed are likely due to other mechanisms.*

DISCUSSION

By comparing gene expression at the relative embryological stages of the two morphs we determine the extent that expression changes contribute to life-history differences. LL offspring take longer to reach the same embryonic developmental stages as PP in absolute time, although the stages are similar, and morphological differences are not observed until later development. Numerous morphological and life-history differences occur between the morphs at the swimming larval stage (reviewed in [32]), but here we detail differences in the embryology as well. Evolutionarily, we expect that LL is the more derived developmental mode, which likely arose from a maternal increase in egg size followed by adaptive changes in the genome that increased lecithotrophic fitness (reviewed in [32]). This generates predictions about the types of genes that we expect to change in patterns and magnitude. For example, we may expect that genes associated with feeding processes, growth, and cell cycle and specification initiate faster in the smaller, PP embryo [50–54]. The non-obligatory-feeding LL larvae have a delay in mesoderm formation and through-gut development [29,55] so we would expect a delay in LL for mesoderm and gut specification, which we do see in the GO terms for the genes expressed earlier in PP.

The pattern of differential expression we observe over development— where numerous but small expression differences occur early followed by a few genes with large expression differences (Fig 2C)— is consistent with the importance of gastrulation to canalize development. Gastrulation is the 'phylotypic stage' at which different phyla are the most morphologically and molecularly similar to each other during their development [56–61]. Adults of both types are also extremely similar, so we would not expect persistent, large-scale gene expression differences at these late stages. The few genes that remain significantly DE post gastrulation are candidates for maintaining morphological and life-history differences between the morphs. ('Chitin biosynthesis' genes, for example, remain significantly different at later stages, and upregulated in PP). However, these genes' expression levels might converge much later in development as both morphs reach adulthood.

We see similar amounts of heterochronic and exclusive genes. Heterochronic genes, and their pathways, are ideal candidates for modifying larval and life-history traits. Exclusive genes are of particular interest in *S. benedicti* as previous genetic and transcriptomic work suggests there is very little differentiation of the two morphs at the genomic level [31,62]. It is possible that genomic rearrangements [63], gene duplications [64], or co-options [65] underlie the expression of exclusive genes. However, exclusive genes are more likely driven by regulatory differences in enhancing or silencing elements (as in [66,67]).

We expect that incipient species that are early in their divergence have more genetic regulatory changes due to *trans*-acting factors; this is because *trans*-acting factors may be more pleiotropic and initiate numerous developmental changes parsimoniously. Essentially a few *trans*-acting regulatory modifications, such as changes to transcription factors, could account for most of the developmental and phenotypic differences between morphs. As species divergence time increases, *cis*-acting elements can arise and refine gene expression of individual genes [44,45,49]. These regulatory architecture trends have been reported in other species, but none have examined the short evolutionary window of divergent morphs within the same species. This suggests the life-history differences we see in this early evolutionary divergence are driven by a few developmentally upstream transacting factors with pleiotropic effects.

We quantify the relative contribution of heteromorphic, heterochronic, and exclusive gene expression pattern modifications in the evolution of a developmental dimorphism. We find that heterochronic and exclusive genes contribute similarly to gene expression differentiation. Additionally, we find that the regulatory architecture of differential expression is predominantly *trans*-acting modifications, supporting the hypothesis that early gene expression evolution occurs by few, highly pleiotropic *trans*-acting regulatory modifications.

METHODS

Animal rearing and sample collection

We use lab-reared male and female *S. benedicti* originally sampled from Newark Bay Bayonne, New Jersey (PP) and Long Beach, California (LL). All experimental procedures and growth incubations are carried out at 20°C unless otherwise noted. To produce offspring for sampling we crossed virgin females with one male for each test cross (CrossID in Figure 2A). Planktotrophic offspring were fed small quantities of our lab's standard feeding algae after the swimming stage had been reached to avoid starvation. As clutch sizes can be quite small, we took advantage of the multiple broods that can be produced by a single mating event. We used 3-5 consecutive broods per female where all offspring were full sibs. This is necessary as embryo number is limited (10-40 embryos per clutch for LL and 100-400 for PP) and 10-40 pooled embryos are required to produce sufficient input mRNA. In this study brood number and developmental stage are confounded within sample groups (CrossID).

Embryo Timeline Construction

To build a timeline of development, we removed embryos from the maternal brood pouch and observed development in a petri dish in artificial sea water in an incubator at 20C (previous work indicates development is normal outside the brood pouch). We selected six distinct timepoints that we confirmed with cell counts (nuclei) using Hoechst: 16-cells, blastula (64 cells), and gastrula (124 cells). Time to each development stage was averaged over at least four clutch observations. Later embryo and larval stages were identified by morphological differences (appearance of ciliated band trochophores and eye development). Six stages capture a broad scope of developmentally critical periods while having enough timing separation to be distinct (each embryonic time point was ~12 hours apart). Images were captured using a 40x objective on a Zeiss Axio inverted microscope with an indicator for scale which was subsequently used to scale images to the same relative size in Figure 1.

RNAseq

Embryos were collected from a single female, and 10-40 were processed for total mRNA extraction at each timepoint. If the number of embryos was insufficient for all time points, we used multiple broods. We used a minimum of 20 LL embryos for stages 16-cell, blastula, gastrula, and 10 LL embryos for stages trochophore, swimming, 1-week. PP females have large clutches, so embryos were divided equally (~40 embryos/clutch) among the six stages. With this approach we could not collect all timepoints for the same females, but there are at least 2 complete sets and at least 4 replicates per stage for each morph (Figure 2A).

Embryo RNA extraction used the Arcturus PicoPure (Ref: 12204-01) including the DNAse step. We used a Qubit RNA kit to measure RNA yields for pilot data but once we could estimate RNA yield per embryo we bypassed this step to maximize RNA yields. Libraries were constructed with the NEB UltraII Stranded RNA library prep kit (cat# E7760S) for Illumina. Libraries were sequenced on two lanes of 150bp on the Illumina NovaSeq.

Sequencing read quality trimming and mapping

We used TrimGalore (cutadapt) [68] and FastP [69] for quality assessment and trimming. Reads were mapped to a reference of all transcript sequences (transcripts extracted with GFFread, [70]) that are annotated in the *S. benedicti* reference genome [62] using Salmon 1.10 [71] using the default scoring parameters. Individual sample transcript expression quantification estimates were summarized to the gene-level using the Tximport R package [72].

Library diagnostics

As the *S. benedicti* genome is from PP individuals [62], we checked the mapping rate of PP and LL samples at each developmental stage to ensure similar mapping rates. (SFigure1). We calculated mapping by dividing the sum of expression estimates for all genes by the sequencing depth for each sample individually, and then averaging samples from each developmental stage by morph to produce an average mapping rate (Supp figure 1). We find LL and PP mapping rates are comparable.

Samples which received fewer than 2 million total reads were excluded from further analyses since our expected sequencing depth was between 20 and 40 million reads (six samples total). Samples were normalized using the DESeq2 R package [73] and expression estimates were transformed using the variance stabilizing transformation function to perform principal component analysis (PCA) using the function *prcomp* [74], which is plotted for the first two principal components with ggplot2 [75] (Figure 2B).

Differential expression

We used DESeq2 to determine significance in expression differences [73] using the variable grouping method recommended by the DESeq2 manual. We used two factors for characterization: developmental stage and genotype (PP, LL, PL, LP) which generate a factor level for each stage/genotype combination (ie. '16cell_Lecithotroph') which we called 'multiFactor' for DESeq2 as "design = ~multiFactor". The Benjamini-Hochberg false discovery rate (FDR) algorithm was used with P-values between factor levels of 'multiFactor' to reduce the incidence of false positives for differential expression. Throughout this study, the threshold for

significant gene expression differences is an FDR adjusted p-value of 0.05 or less, and an expression fold change greater than 2-fold.

Expression profile clustering

To summarize expression clusters, we used the Mfuzz 2.58 R package [76,77] Clusters are based on mean expression estimates from P samples. Normalization was (DESeq2, [73]), log2-transformed, and then filtered for genes with low variability/expression and standardized using the Mfuzz functions '*filter.std'* and '*standardize'* respectively. Mfuzz requires *a priori* cluster numbers. To capture the representative sample of expression patterns, we found the highest number of clusters for which Pearson correlations of pairs of cluster centroids (the single most cluster-representative gene for each cluster) did not exceed 0.85 ($r < 0.85$) for any pair of clusters. This resulted in 6 clusters. The fuzzifier coefficient (m) required by Mfuzz was estimated to be 1.71 by the function '*mestimate'* [76,77]. Following cluster generation with the PP expression data, the LL expression data was mapped onto the clusters using the Mfuzz function '*membership'* so both morphs could be compared. Clusters were plotted using the function '*mfuzz.plot2'* (Figure 4).

Expression Classification

In order to identify genes as heterochronic we first identify genes that have different expression patterns between the morphs and then filter those genes based on their Pearson correlation coefficients. Genes for which the PP and LL expression are on different clusters are categorized as heterochronic. Some of these genes fall onto different clusters but still have relatively similar expression profiles. Therefore, we also filter by a measure of the Pearson correlation coefficient. The threshold is based on the distribution of correlation coefficients of all PP to LL genes' expression in the dataset (Figure 5b). Based on the distribution, a threshold of $r < 0.85$ was selected to filter genes which would be classified as switching clusters but are similar enough in their expression between the two morphs to be removed. This filter disqualified 38 genes.

Morph-exclusive expression

To categorize exclusive expression, we use a gene expression thresholding approach. Spurious gene expression resulting in low levels of estimated expression (few RNAseq reads) is categorically and functionally distinct from robust gene expression [78]. We establish meaningful expression thresholds based on the assumption that genes with robust expression will have values which are normally distributed. Individual sample thresholds were calculated using a data-adaptive flag method (DAFS) [79]. Thresholds were used to identify genes which are expressed at significant levels in only PP or only LL offspring. For example, if a gene is expressed above its threshold in three or more (out of five) P samples at one or more developmental stages and in no more than one LL sample at any developmental stage, then that gene was categorized as exclusive.

PP samples above expression threshold	LL samples above expression threshold	Classification
≤ 1	≥ 3	Lecithotrophic-specific expression
≥ 3	≤ 1	Planktotrophic-specific expression

Table 2: Logical table showing criterion for morph-specific expression classification of genes based on number of samples in each group that have gene expression greater than their expression level threshold as assessed by DAFS. An allowance for a single mismatched replicate is made in this criterion. This is because we do not believe that a single incidence within a morph (across all developmental stages) is a functionally significant and biologically relevant level of gene expression in this case.

F_1 expression time-series collection

We concurrently sequenced RNA libraries from F_1 offspring (Figure 2A). We generated F_1 offspring from reciprocal crosses PxL and LxP (both mother-father directions) using the same sample collection method as stated above. However, we only sequenced samples from 3 of the developmental stages for the F_1's in the LP direction. All analyses were performed as above.

Mode of inheritance and F_1 mis-expression

Using F_1 data, we classified the mode of inheritance for each gene according to established criteria and differential expression tests from DESeq2 [39,45,47]. Genes which were significantly DE between reciprocal F_1 offspring (PL vs LP) at any developmental stage were not considered for this analysis. The remaining genes were classified as either 1) conserved, 2) additive, 3) dominant for one genotype, or 4) mis-expressed (over/under-dominant).

Parent of origin effects

DE genes between PL and LP offspring are different due to parental effects. DE was detected with DESeq2 with the contrasts 'c("multiFactor", "PL_sixteencell", "LP_sixteencell")', 'c("multiFactor", "PL_gastrula", "LP_gastrula")', and 'c("multiFactor", "PL_swimming", "LP_swimming")' similar to PP and LL samples. We evaluate these at three of the six developmental stages. The direction of the parental effect was determined by matching the expression change with the direction of each identified gene's expression in PP and LL samples (where those genes were differentially expressed in PP and LL samples).

Mode of regulatory change

To measure allele-specific expression, we assigned sequencing reads from F_1 samples to either P or L parentage by identifying fixed SNPs within the transcript sequences of the parental types. We used HyLiTE [80] to identify SNPs and assign reads as a P or L allele. We then categorized genes' regulatory mode according to established empirical methods [39,44,45,47,48] for each developmental stage independently. This required three comparisons of each gene's expression which we performed using DESeq2: 1) the contrast of PP to LL samples, 2) the contrast of P alleles to L alleles within F_1 samples, and 3) a ratio of the differential expression of

PP:LL to the differential expression of P:L alleles. For 3) we use a special contrast in DESeq2, applying the design (~ Geno * Ori) where Geno identifies reads as either a P or L allele and Ori identifies the reads as originating from the parentals or F_1 samples. DE genes were categorized as either in "cis", "trans", "cis + trans", or "cis * trans." [39,44,45,47,48].

DATA ACCESSIBILITY

All raw reads will be submitted to NCBI's sequence read archive (SRA). All analyses and datasets are available in the supplemental files, which contain estimated transcript expression data generated by Salmon 1.10 and an Rmarkdown file containing annotated code used to conduct these analyses.

ACKNOWLEDGEMENTS

Sarah Cole for help with collecting embryos, animal rearing, constructing the sampling timeline. Thanks to Matthew Rockman and Greg Wray for comments on the manuscript.

REFERENCES

1. Gould, S. J. *Ontogeny and Phylogeny*. (Harvard University Press, 1985).
2. Dobreva, M. P., Camacho, J. & Abzhanov, A. Time to synchronize our clocks: Connecting developmental mechanisms and evolutionary consequences of heterochrony. *J Exp Zool Pt B* **338**, 87–106 (2022).
3. deBeer, G. R. Embryology and Evolution. *Humana Mente* **5**, 482–484 (1930).
4. Wray, G. A. & McClay, D. R. Molecular heterochronies and heterotopies in early echinoid development. *Evolution* **43**, 803–813 (1989).
5. Wray, G. A. & Raff, R. A. The evolution of developmental strategy in marine invertebrates. *Trends in Ecology & Evolution* **6**, 45–50 (1991).
6. Smith, K. K. Sequence Heterochrony and the Evolution of Development. *J. Morphol.* **252**, 82–97 (2002).
7. Smith, K. K. Time's arrow: heterochrony and the evolution of development. *Int J Dev Biol* **47**, 613–621 (2003).
8. Erwin, D. H. & Davidson, E. H. The last common bilaterian ancestor. *Development* **129**, 3021–3032 (2002).
9. King, M.-C. & Wilson, A. C. Evolution at Two Levels in Humans and Chimpanzees: Their macromolecules are so alike that regulatory mutations may account for their biological differences. *Science* **188**, 107–116 (1975).
10. Carroll, S. B. Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**, 479–485 (1995).
11. Wray, G. A. The Evolution of Transcriptional Regulation in Eukaryotes. *Molecular Biology and Evolution* **20**, 1377–1419 (2003).
12. Fay, J. C. & Wittkopp, P. J. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity* **100**, 191–199 (2008).
13. López-Maury, L., Marguerat, S. & Bähler, J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat Rev Genet* **9**, 583–593 (2008).
14. Hamann, E. *et al.* Rapid evolutionary changes in gene expression in response to climate fluctuations. *Mol Ecol* **30**, 193–206 (2021).
15. Raff, R. A. & Wray, G. A. Heterochrony: Developmental mechanisms and evolutionary results. *J Evolution Biol* **2**, 409–434 (1989).
16. Gunter, H. M., Koppermann, C. & Meyer, A. Revisiting de Beer's textbook example of

heterochrony and jaw elongation in fish: calmodulin expression reflects heterochronic growth and underlies morphological innovation in the jaws of belonoid fishes. *EvoDevo* **5**, 8 (2014).
17. Vaglia, J. L. & Smith, K. K. Early differentiation and migration of cranial neural crest in the opossum, Monodelphis domestica. *Evol Dev* **5**, 121–135 (2003).
18. Carleton, K. L. *et al.* Visual sensitivities tuned by heterochronic shifts in opsin gene expression. *BMC Biol* **6**, 22 (2008).
19. Willink, B., Duryea, M. C., Wheat, C. & Svensson, E. I. Changes in gene expression during female reproductive development in a color polymorphic insect. *Evolution* **74**, 1063–1081 (2020).
20. Hilgers, L., Hartmann, S., Hofreiter, M. & Von Rintelen, T. Novel Genes, Ancient Genes, and Gene Co-Option Contributed to the Genetic Basis of the Radula, a Molluscan Innovation. *Molecular Biology and Evolution* **35**, 1638–1652 (2018).
21. Luna, S. K. & Chain, F. J. J. Lineage-Specific Genes and Family Expansions in Dictyostelid Genomes Display Expression Bias and Evolutionary Diversification during Development. *Genes* **12**, 1628 (2021).
22. McNamara, K. J. Heterochrony: The Evolution of Development. *Evo Edu Outreach* **5**, 203–218 (2012).
23. Ruvkun, G. & Giusto, J. The Caenorhabditis elegans heterochronic gene /in-14 encodes a nuclear protein that forms a temporal developmental switch. **338**, (1989).
24. Simola, D. F., Francis, C., Sniegowski, P. D. & Kim, J. Heterochronic evolution reveals modular timing changes in budding yeast transcriptomes. *Genome Biology* **11**, R105 (2010).
25. Capra, J. A., Pollard, K. S. & Singh, M. Novel genes exhibit distinct patterns of function acquisition and network integration. *Genome Biol* **11**, R127 (2010).
26. Schmitz, J. F., Chain, F. J. J. & Bornberg-Bauer, E. Evolution of novel genes in three-spined stickleback populations. *Heredity* **125**, 50–59 (2020).
27. Levin, L. A. & Creed, E. L. Effect of temperature and food availability on reproductive responses of *Streblospio benedicti* (Polychaeta: Spionidae) with planktotrophic or lecithotrophic development. *Marine Biology: International Journal on Life in Oceans and Coastal Waters* **92**, 103–113 (1986).
28. Levin, L. A. Multiple Patterns of Development in *Streblospio benedicti* Webster (Spionidae) from Three Coasts of North America. *Biological Bulletin* **166**, 494–508 (1984).
29. Gibson, G., MacDonald, K. & Dufton, M. Morphogenesis and phenotypic divergence in two developmental morphs of *Streblospio benedicti* (Annelida, Spionidae). *Invertebrate Biology* **129**, 328–343 (2010).
30. Zakas, C. & Rockman, M. V. Dimorphic development in *Streblospio benedicti:* genetic analysis of morphological differences between larval types. *Int. J. Dev. Biol.* **58**, 593–599 (2014).
31. Zakas, C., Deutscher, J. M., Kay, A. D. & Rockman, M. V. Decoupled maternal and zygotic genetic effects shape the evolution of development. *eLife* **7**, e37143 (2018).
32. Zakas, C. Chapter Seventeen - *Streblospio benedicti*: A genetic model for understanding the evolution of development and life-history. in *Current Topics in Developmental Biology* (eds. Goldstein, B. & Srivastava, M.) vol. 147 497–521 (Academic Press, 2022).
33. McCain, E. R. Poecilogony as a tool for understanding speciation: Early development of *Streblospio benedicti* and *Streblospio gynobranchiata* (polychaeta:spionidae). *Invertebrate Reproduction and Development* **51**, 91–101 (2008).
34. Lyons, D. C., Perry, K. J., Lesoway, M. P. & Henry, J. Q. Cleavage pattern and fate map of the mesentoblast, 4d, in the gastropod Crepidula: a hallmark of spiralian development. *EvoDevo* **3**, 21 (2012).
35. Pernet, B. & McArthur, L. Feeding by larvae of two different developmental modes in *Streblospio benedicti* (Polychaeta: Spionidae). *Marine Biology* **149**, 803–811 (2006).

36. Pernet, B. & McHugh, D. Evolutionary changes in the timing of gut morphogenesis in larvae of the marine annelid *Streblospio benedicti*. *Evolution and Development* **12**, 618–627 (2010).
37. McAlister, J. S. & Moran, A. L. Relationships among Egg Size, Composition, and Energy: A Comparative Study of Geminate Sea Urchins. *PLoS One* **7**, e41599 (2012).
38. Pettersen, A. K., Schuster, L. & Metcalfe, N. B. The Evolution of Offspring Size: A Metabolic Scaling Perspective. *Integrative and Comparative Biology* **62**, 1492–1502 (2022).
39. Harry, N. D. & Zakas, C. Maternal patterns of inheritance alter transcript expression in eggs. *BMC Genomics* **24**, 191 (2023).
40. Norrström, N., Getz, W. M. & Holmgren, N. M. A. Selection against Accumulating Mutations in Niche-Preference Genes Can Drive Speciation. *PLoS ONE* **6**, e29487 (2011).
41. Moran, B. M. *et al.* The genomic consequences of hybridization. *eLife* **10**, e69016 (2021).
42. Davidson, E. H. & Peter, I. S. Chapter 4 - Genomic Control Processes in Adult Body Part Formation. in *Genomic Control Process* (eds. Davidson, E. H. & Peter, I. S.) 133–200 (Academic Press, 2015). doi:10.1016/B978-0-12-404729-7.00004-6.
43. Tirosh, I., Reikhav, S., Levy, A. A. & Barkai, N. A Yeast Hybrid Provides Insight into the Evolution of Gene Expression Regulation. *Science* **324**, 659–662 (2009).
44. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85–88 (2004).
45. Coolon, J. D., McManus, C. J., Stevenson, K. R., Graveley, B. R. & Wittkopp, P. J. Tempo and mode of regulatory evolution in Drosophila. *Genome Res* **24**, 797–808 (2014).
46. McManus, C. J. *et al.* Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Res* **20**, 816–825 (2010).
47. Wang, L. *et al.* Genetic basis for divergence in developmental gene expression in two closely related sea urchins. *Nat Ecol Evol* **4**, 831–840 (2020).
48. Graze, R. M., McIntyre, L. M., Main, B. J., Wayne, M. L. & Nuzhdin, S. V. Regulatory Divergence in Drosophila melanogaster and D. simulans, a Genomewide Analysis of Allele-Specific Expression. *Genetics* **183**, 547–561 (2009).
49. Cartwright, E. L. & Lott, S. E. Evolved Differences in cis and trans Regulation Between the Maternal and Zygotic mRNA Complements in the Drosophila Embryo. *Genetics* **216**, 805–821 (2020).
50. Wray, G. A. & Lowe, C. J. Developmental Regulatory Genes and Echinoderm Evolution. *Systematic Biology* **49**, 28–51 (2000).
51. Strathmann, R. R. Functional design in the evolution of embryos and larvae. *Seminars in Cell & Developmental Biology* **11**, 395–402 (2000).
52. Lowe, C. J., Issel-Tarver, L. & Wray, G. A. Gene expression and larval evolution: changing roles of distal-less and orthodenticle in echinoderm larvae. *Evol Dev* **4**, 111–123 (2002).
53. Segers, F. H. I. D., Berishvili, G. & Taborsky, B. Egg size-dependent expression of growth hormone receptor accompanies compensatory growth in fish. *Proc. R. Soc. B.* **279**, 592–600 (2012).
54. Figueroa, Á., Brante, A. & Cárdenas, L. RNA-Seq reveals divergent gene expression between larvae with contrasting trophic modes in the poecilogonous polychaete *Boccardia wellingtonensis. Scientific Reports* **11**, 1–13 (2021).
55. Pernet, B. Persistent Ancestral Feeding Structures in Nonfeeding Annelid Larvae. *Biological Bulletin* **205**, 295–307 (2003).
56. Slack, J. M. W., Holland, P. W. H. & Graham, C. F. The zootype and the phylotypic stage. *Nature* **361**, 490–492 (1993).
57. Duboule, D. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development* **1994**, 135–142 (1994).
58. Richardson, M. K. Heterochrony and the Phylotypic Period. *Developmental Biology* **172**,

412–421 (1995).
59. Irie, N. & Kuratani, S. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat Commun* **2**, 248 (2011).
60. Irie, N. & Kuratani, S. The developmental hourglass model: a predictor of the basic body plan? *Development* **141**, 4649–4655 (2014).
61. 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).
62. Zakas, C., Harry, N. D., Scholl, E. H. & Rockman, M. V. The Genome of the Poecilogonous Annelid *Streblospio benedicti*. *Genome Biol Evol* **14**, evac008 (2022).
63. Janssen, P. J. Strain-specific genes of Helicobacter pylori: distribution, function and dynamics. *Nucleic Acids Research* **29**, 4395–4404 (2001).
64. Long, M., Betrán, E., Thornton, K. & Wang, W. The origin of new genes: glimpses from the young and old. *Nat Rev Genet* **4**, 865–875 (2003).
65. Linksvayer, T. A. & Wade, M. J. The Evolutionary Origin and Elaboration of Sociality in the Aculeate Hymenoptera: Maternal Effects, Sib-Social Effects, and Heterochrony. *The Quarterly Review of Biology* **80**, 317–336 (2005).
66. Zhao, J., Zhou, Y., Li, X., Cai, W. & Hua, H. Silencing of juvenile hormone epoxide hydrolase gene (Nljheh) enhances short wing formation in a macropterous strain of the brown planthopper, Nilaparvata lugens. *Journal of Insect Physiology* **102**, 18–26 (2017).
67. Matlosz, S., Sigurgeirsson, B., Franzdóttir, S. R., Pálsson, A. & Jónsson, Z. O. methylation differences during development distinguish sympatric morphs of Arctic charr (Salvelinus alpinus). *Molecular Ecology* **31**, 4739–4761 (2022).
68. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet* **17**, 3 (2011).
69. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
70. Pertea, G. & Pertea, M. GFF Utilities: GffRead and GffCompare. *F1000Research* **9**, (2020).
71. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417–419 (2017).
72. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).
73. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
74. R Core Team. R: A language and environment for statistical computing. (2020).
75. Wickham, H. *ggplot2: Elegant graphics for data analysis*. (Springer, 2016).
76. Futschik, M. E. & Carlisle, B. Noise-robust soft clustering of gene expression time-course data. *J. Bioinform. Comput. Biol.* **03**, 965–988 (2005).
77. Kumar, L. & Futschik, M. E. Mfuzz: A software package for soft clustering of microarray data. *Bioinformation* **2**, 5–7 (2007).
78. Hebenstreit, D. *et al.* RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Molecular Systems Biology* **7**, 497 (2011).
79. George, N. I. & Chang, C.-W. DAFS: a data-adaptive flag method for RNA-sequencing data to differentiate genes with low and high expression. *BMC Bioinformatics* **15**, 92 (2014).
80. Duchemin, W., Dupont, P.-Y., Campbell, M. A., Ganley, A. R. & Cox, M. P. HyLiTE: accurate and flexible analysis of gene expression in hybrid and allopolyploid species. *BMC Bioinformatics* **16**, 8 (2015).