Evolutionary changes in the chromatin landscape reshape a developmental gene regulatory network during rapid life history divergence in sea urchins

Phillip L. Davidson¹, Maria Byrne², Gregory A. Wray^{1*}

1. Department of Biology, Duke University, Durham, NC, USA

2. School of Life and Environmental Sciences, University of Sydney, Sydney, NSW, AU

*** Corresponding Author:** ggray@duke.edu

ABSTRACT

Chromatin configuration is highly dynamic during embryonic development in animals, exerting an important point of control in transcriptional regulation. Yet there exists remarkably little information about the contribution of evolutionary changes in chromatin configuration to the evolution of gene expression and organismal traits. Genome-wide assays of chromatin configuration, coupled with whole-genome alignments, can help address this gap in knowledge in several ways. In this study we present a comparative analysis of regulatory element sequences and accessibility throughout embryogenesis in three sea urchin species with divergent life histories: a lecithotroph *Heliocidaris erythrogramma*, a closely related planktotroph *H. tuberculata*, and a distantly related planktotroph *Lytechinus variegatus*. We identified distinct epigenetic and mutational signatures of evolutionary modifications to *cis*-regulatory element function in *H. erythrogramma* that have accumulated non-uniformly throughout the genome, suggesting selection, rather than drift, underlies many modifications associated with the derived life history. Specifically, regulatory elements composing the sea urchin developmental gene regulatory network are enriched for signatures of positive selection and accessibility changes which may function to alter binding affinity and access of developmental transcription factors to these sites. Furthermore, regulatory element changes often correlate with divergent expression patterns of genes involved in cell type specification, morphogenesis, and development of other derived traits, suggesting these evolutionary modifications have been consequential for phenotypic evolution in *H. erythrogramma*. Collectively, our results demonstrate that selective pressures imposed by changes in developmental life history rapidly reshape the *cis*-regulatory landscape of core developmental genes to generate novel traits and embryonic programs.

INTRODUCTION

Cis-regulatory elements are significant contributors to phenotypic evolution and trait diversification by modulating transcriptional regulation via gene network interactions (Davidson and Erwin 2006; Wray 2007; Wittkopp and Kalay 2012). Extensive empirical evidence demonstrates that the function of *cis*-regulatory elements depends on distinct molecular characteristics such as the element's sequence and epigenetic landscape that collectively provide a dynamic and heritable genomic substrate affecting inter- and intraspecific diversification. Enabled by advances in genomic technology and reduced sequencing costs, more recent work has demonstrated the importance of *cis*-regulatory elements in diverse biological contexts across a variety of animal systems at a genome-wide scale; for example: origination of human-specific traits (Shibata, et al. 2012; Swain-Lenz, et al. 2019), convergent evolution of flight loss in birds (Sackton, et al. 2019), population-level adaptation (Lewis and Reed 2019), genetic assimilation (van der Burg, et al. 2020) and color identity (Livraghi, et al. 2021) in butterfly wing patterning, embryonic regulation of gene expression in the crustacean *Parhyale hawaiiensis* (Sun and Patel 2021), limb bud development in chicken and mice (Jhanwar, et al. 2021), and single-cell resolution of *Drosophila* cell lineages (Cusanovich, et al. 2018). However, direct integration of large-scale *cis*-regulatory analysis and transcription in a defined evolutionary-developmental (evo-devo) context is lacking. For example, relationships governing divergence in chromatin and gene expression, the role of such changes in trait evolution, and the action of natural selection on these processes remain largely unexplored within an interrogable phylogenetic system. In this study, we address these important questions by comparing the function of *cis*-regulatory elements in embryonic gene expression of sea urchin species with divergent developmental life

history strategies to assess the extent *cis*-regulatory elements play in the evolution of gene expression, developmental systems, and by extension, trait variation and origination.

Among the most common and well-studied life history shifts in the animal kingdom is the evolution of feeding or non-feeding larval development (Flatt and Heyland 2011). In marine invertebrates, planktotrophy (feeding development) is characterized by species that produce many, relatively small eggs and have a longer planktonic larval period, whereas species with lecithotrophic (non-feeding) development invest more maternal resources into a fewer number of large eggs that develop through an abbreviated larval stage (Strathmann 1985). Importantly, the ecological circumstances associated with each strategy are predictable and well-studied across diverse marine clades (Vance 1973; Strathmann 1985; Pechenik 1999; Marshall, et al. 2012; Morgan 2020), providing natural systems to analyze how selection shapes often rapid phenotypic divergence in developmental regulation and physiology. For example, within Echinoderms, lecithotrophy has independently evolved from the ancestral planktotrophic state more than 20 times (Raff 1987; Wray 1996; McEdward and Miner 2001; Hart 2002), implying that recurring sets of environmental conditions and strong selection can enact evolutionary modifications to even highly conserved developmental processes. The consequences include changes at every developmental stage ranging from egg composition (Prowse, et al. 2009; Falkner, et al. 2015; Byrne and Sewell 2019; Davidson, et al. 2019) and zygotic genome activation (Wray and Raff 1990; Israel, et al. 2016; Wang, et al. 2020), cell type specification and embryonic patterning (Wray and Raff 1989; Love and Raff 2006), to construction of larval structures such as the skeleton and rudiment (Emlet 1995; Koop, et al. 2017), the latter being the region from which the adult emerges during metamorphosis. Still, the impact selection exerts on the evolution of *cis*-regulatory interactions underlying alternative developmental programs and contrasting life

history strategies is not known. This gap in knowledge is manifested more broadly in our current sparse understanding of the biological relationships governing regulatory element function, evolutionary constraints inherent to development and gene network architecture, and phenotypic diversification.

Perhaps the best studied example of the life history switch from planktotrophic to lecithotrophic development is the sea urchin *Heliocidaris erythrogramma*. Prior work has found dramatic modifications to morphogenesis and embryonic patterning in *H. erythrogramma* relative to planktotrophic sea urchin species (for reviews see Raff and Byrne 2006; Wray 2021). For instance, dorso-ventral and left-right axis establishment occurs considerably early in this species (Henry, et al. 1990), concordant with accelerated rudiment development and metamorphosis (Williams and Anderson 1975). At other embryonic domains, specification of cell lineages such as those responsible for mesenchymal tissue types have been delayed, lost, or spatially rearranged (Wray and Raff 1990; Klueg, et al. 1997; Wilson, Andrews, Rudolf Turner, et al. 2005; Davidson, et al. 2021). Recent transcriptomic studies have shown that expression profiles of key developmental gene regulatory network (dGRN) genes in this species have diverged from the highly conserved expression patterns in planktotrophic species (Israel, et al. 2016), and many of these regulatory changes are predicted to be based in *cis* (Wang, et al. 2020). The contribution of evolutionary changes in the sequence and configuration of *cis*-regulatory elements to the rapid evolution of these derived phenotype in *H. erythrogramma* is not known.

To address this, we compared chromatin accessibility, regulatory element sequence, and gene expression across six developmental stages in three sea urchin species spanning this life history switch: the lecithotroph *H. erythrogramma*, a closely-related (~ 5 my divergence) planktotroph *H. tuberculata*, and a distantly-related (~ 40 my divergence) planktotroph,

Lytechinus variegatus (**fig. 1a**). Planktotrophic (feeding) larvae represent the *ancestral* life history mode within echinoderms relative to the *derived* lecithotrophic state (Strathmann 1978, Raff 1987, Wray 1996, McEdward and Miner 2001), so recent and dramatic changes in developmental phenotypes in *H. erythrogramma* provide an exceptional natural system to analyze the consequences of strong shifts in selection and life history on the evolution of gene regulatory mechanisms. We emphasize the distinction between two molecular mechanisms, chromatin accessibility and sequence evolution, in generating change to *cis*-regulatory element function and developmental gene expression. We identify distinct epigenetic and mutational signatures of evolutionary modifications in the *H. erythrogramma* genome relative to species with the ancestral life history. Furthermore, these regulatory changes have accumulated non-uniformly throughout the genome and often correlate with derived expression patterns of genes involved in cell type specification, morphogenesis, and development of other derived traits. Importantly, the patterns of *cis*-regulatory modifications and their impact on the evolution of transcriptional regulation characterized here may be broadly applied to biological principles underlying the integration of developmental and gene regulatory mechanisms in lineage diversification.

RESULTS AND DISCUSSION

We performed ATAC-seq on six developmental stages of *H. erythrogramma* (*He*), *H. tuberculata* (*Ht*), and *L. variegatus* (*Lv*) in order to identify open chromatin regions (OCRs) representing putative *cis*-regulatory elements (CREs) during embryogenesis. In this study, we use the term “OCR” when reporting analysis results because ATAC-seq is an assay of open chromatin. However, we apply the more specific term “CRE” when discussing the biological interpretation of our results because *cis*-regulatory elements are the operational units identified

by open chromatin peaks used to understand mechanisms of gene regulatory evolution. The developmental stages analyzed here include 4th/5th, 6th/7th, and 8th/9th mixed cleavage stage embryos, as well as hatched blastula, mid-gastrula, and early larva (**fig. 1b**), enabling characterization of regulatory element dynamics from the onset of zygotic transcription through the specification and morphogenesis of complex tissue types. We then carried out alignments of chromosome-scale genome assemblies to find evolutionarily conserved OCRs between species to compare accessibility and sequence composition of CREs. These analyses provide a simultaneous comparison of gene regulatory mechanisms in two ways: 1) availability of CREs to transcription factor binding by modulating chromatin configuration (accessibility) and 2) affinity of transcription factor binding to CREs via changes in nucleotide sequence, as evidenced by motif analysis and signatures of positive selection. Importantly, the genomes of these sea urchin species were assembled and annotated using an identical set of sequencing technology and methodology (Davidson, et al. 2020; Davidson, et al. 2021), minimizing technical bias and permitting high resolution identification of changes in regulatory element activity.

Of 42,679 OCRs identified across the three species, we found 35,788 1:1:1 orthologous, high confidence OCRs proximal to 11,679 gene models for cross-species comparison (**fig. 2a**). Note that inclusion in this set of 1:1:1 orthologues does not require that chromatin be open in all three species, but rather that it be open in at least one species at one time point surveyed and that the orthologous sequence is present in all three species. This set of orthologous OCRs forms the basis of most analyses presented here, including comparisons of developmental chromatin configuration (**fig. 2b, 3**), analyses of OCR sequence changes associated with positive selection and transcription factor binding (**fig. 2d, 4**), discerning how these changes are distributed among OCRs located proximally or distally to protein coding sequences (**fig. 2c, 5**), and determining the

extent to which these regulatory modifications are associated with derived gene expression patterns (**figs. 5-7**). In addition to orthologous OCRs, we also identified sets of species-specific gains or losses of OCRs occurring on entire genomic segments present or lost in a single species (**fig. 2a**). Because these indel-based differences in OCRs are only physically present in at most two of the three sea urchin species compared in this study, they are treated as distinct sets of putative CREs. While the changes in regulatory element function examined in this study are not applicable to species-specific OCRs, it is important to note that many physical OCR gains occur nearby genes with differential expression levels in *H. erythrogramma*, suggesting gain or loss of entire CREs may be an important component of the evolution of this species' derived life history strategy (**supplementary fig. 1**).

Development and life history drive global variation patterns of accessibility

First, we assessed which biological factors may significantly contribute to variation in chromatin accessibility at a genomic scale. Principal component analysis (PCA) suggests species-level differences (PC 1: 20.3%) and the interaction of development and life history (PC 2: 18.5%) are the most important factors explaining variation in chromatin accessibility when all orthologous OCRs are considered (**fig. 1c**). Within each species, chromatin accessibility differences between stages roughly mirrors developmental progression as early and late stages cluster more closely to one another. Strikingly, though, there is far greater variation in accessibility through development in both species with the ancestral life history relative to *He*, suggesting the developmental chromatin landscape of the derived state is less dynamic during embryogenesis. This finding is consistent with delayed blastomere specification (Wray and Raff 1989) and simplified larval morphology in *He* (Emlet 1995), suggesting chromatin state changes

normally associated with establishing cell identity and tissue differentiation have been modified during the evolution of its derived developmental life history.

Interestingly, when considering only OCRs near dGRN genes, the primary driver of variation in chromatin accessibility shifts to developmental life history rather than species-level differences (**fig. 1d**; PC 1: 25.5%), a finding corroborated by variance partitioning of the dataset (**supplementary fig. 2**). This result suggests chromatin configuration of critical developmental network genes depends significantly on the species' developmental strategy, as access to regulatory elements of these genes controls the flow and deployment of embryonic programs. Therefore, changes in accessibility to CREs have the potential to modify the function and construction of developmental gene networks by altering regulatory interactions, and by extension, drive the evolution of novel developmental phenotypes.

Loss of regulatory element accessibility within the ancestral dGRN

We next quantified how the accessibility of 35,788 OCRs vary between life history strategies in order to better understand how chromatin configuration at specific CREs may influence the evolution and function of alternative embryonic programs. Across each cleavage stage examined (timepoints spanning 16-cell through ~500-cell stage embryos), we found an average of 5.7% of OCRs were differentially accessible between ancestral and derived sea urchin species, including a modest difference in sites more accessible in *Ht* and *Lv* (3.1%) or *He* (2.6%) (**fig. 3a**). Strikingly, starting at the blastula stage, the proportion of differentially accessible OCRs across the genome rises, with most sites increasing in accessibility ("opening") in the ancestral life history but not in *He* (**fig. 3a**). For example, by the larval stage, 9.5% of OCRs are differentially accessible between life histories, nearly three quarters of which (73.9%) are more accessible in *Ht* and *Lv*. Furthermore, when all OCRs are considered, including both orthologous

and non-orthologous sites across species, a much greater proportion of the *He* genome is closed compared to *Ht* and *Lv*, suggesting that, at a genome-wide scale, *He* is characterized by zygotic quiescence during early development relative to ancestral life history (**fig. 2e; supplementary fig. 3**). Our results support the prediction that as embryonic development proceeds, more regulatory sites open and become available to regulatory factor binding as zygotic transcription activates, cell lineages differentiate, tissue types are specified, and morphological structures are constructed. However, this genome “awakening” and reconfiguration of chromatin state across the embryo unfolds at a much grander scale in *Ht* and *Lv* relative to *He*. This finding supports the idea that embryonic fate specification in *He* is delayed relative to the ancestral state (Wray and Raff 1989, 1990) and collectively, is concordant with a morphologically simplified larva (Williams and Anderson 1975) and derived gene expression patterning in this species (Wilson, Andrews and Raff 2005; Wilson, Andrews, Rudolf Turner, et al. 2005; Smith, et al. 2008; Israel, et al. 2016; Koop, et al. 2017). Reduced accessibility of tissue-specific enhancers may alter gene networks tasked with the establishment of differentiated cell types in the ancestral life history, generating novel interactions and developmental phenotypes in the process.

This hypothesis is further supported when OCR accessibility of dGRN genes are compared. Near these critical developmental regulatory genes, we find a remarkable asymmetry in CRE accessibility between life history modes from the onset of development – on average, 0.5% of OCRs near dGRNs are more accessible in *He* during development, whereas this figure ranges from 2.8% (early cleavage) to 15.1% (larva) for species with the ancestral life history (**fig. 3b**). An example of CRE activation in the ancestral state that is not found in *He* is the chromatin landscape proximal to the gene *patched (ptc)* (**fig. 3c**), a receptor of the Hedgehog ligand whose misexpression results in skeletal defects and abnormal mesodermal patterning in the ancestral sea

urchin developmental program (Walton, et al. 2009). Therefore, modifications to CRE accessibility in *He* are not uniformly distributed across the genome, as would be expected of neutral processes such as drift, but instead are highly concentrated near dGRN genes (Fisher exact test, two-sided: $p < 1.31 \times 10^{-11}$). This finding suggests natural selection has sculpted the epigenetic landscape composing the sea urchin dGRN during life history evolution, generating adaptive changes to transcriptional regulation and expression patterning during embryogenesis. This finding further suggests that the evolutionary changes in putative CRE accessibility have functional consequences. If they had no impact on gene expression or some other trait, natural selection would not influence their status during development.

Developmental regulatory elements have accumulated mutational signatures of positive selection

Next, we analyzed sequence evolution of orthologous regulatory loci to identify instances of mutational pileup as a signal of positive selection within these sites. Substitution rates of 35,788 OCRs on the *He* and *Ht* branches were compared to 10 randomly selected neutrally evolving, non-coding reference sequences, and a likelihood ratio test was implemented to fit substitution rates across the tree to a branch-specific positive selection model, using *Lv* as the outgroup species (see **Methods and Materials** for details). Our goal was to identify cases of gene regulatory evolution via nucleotide variation within CRE sequences, potentially acting to modify transcription factor binding affinity of these elements and associated changes to gene network expression and connectivity.

In total, we identified 648 OCRs with evidence of positive selection in only the *Ht* genome, compared to 512 in the *He* genome (**fig. 4a**: green and orange points, respectively). Unlike OCR accessibility, a trait that changes during development and differs among tissues,

OCR sequence is an immutable genomic property within the lifetime of an individual (excluding somatic mutation and gametogenesis). This result thus suggests that, while the life history of these two sea urchin species dramatically differs, many potential changes in gene regulation associated with other aspects of their biology have accumulated within each species. At a genome-wide scale, these adaptive changes have occurred perhaps even more so in *Ht*, the species with ancestral developmental life history.

Partitioning OCR selection results by the function of the nearest gene reveals only limited enrichment of positive selection within the putative regulatory elements of any particular class of genes (**fig. 4b**, **supplementary fig. 4**). A prominent exception to this finding involves OCRs near genes in the sea urchin dGRN or near genes encoding other transcription factors. In these cases, approximately twice as many OCRs exhibit evidence of positive selection on the *He* branch relative to *Ht*. This result is especially striking considering *Ht* has 26.3% more OCRs with branch-specific evidence of positive selection across the entire genome (previous paragraph and **fig. 4b**). Moreover, nucleotide sequences of the 23 dGRN OCRs under selection in *He* are highly enriched for binding motifs of critical developmental transcription factors, including three SOX, three zinc finger, and one homeobox protein (**fig. 4c**). The SOX binding motifs are interesting because many transcription factors in this family act as “pioneer factors” that are capable of initiating chromatin opening (Zaret and Carroll 2011). Another interesting example involves NRF2: the binding motif for this transcription factor is overrepresented in OCRs under selection in *He* (**fig. 4c**) and the gene is expressed is much at much higher levels in *He* embryos relative to those in the ancestral life history (**fig. 4d**). This protein has gene regulatory functions in diverse biological contexts ranging from antioxidant response and cellular defense to metabolism and stem cell quiescence (Tonelli, et al. 2018), which in this case, could be related to derived

embryonic detoxification and physiological systems (Tsushima, et al. 1995; Davidson, et al. 2019) or early cell lineage specification programs (Wray and Raff 1989, 1990; Davidson, et al. 2021) of *He*. Taken together, regulatory elements tasked with wiring the ancestral sea urchin dGRN have accumulated a disproportionate number of mutations on the *He* branch. These likely play a significant role in regulation of divergent gene expression patterns and its highly derived developmental program. More broadly, our results provide strong evidence that natural selection can efficiently generate heritable changes concentrated within the *cis*-regulatory landscape of specific subsets of genes or gene networks to rapidly produce changes in transcriptional regulation.

Overlap of evolutionary changes in regulatory element accessibility and sequence

The previous two sections describe how two distinct modifications to regulatory element function, namely accessibility and sequence, have accumulated across the *He* genome, and particularly so near critical developmental network genes. We next asked whether there exists any interaction between these two evolutionary changes to better understand the extent to which both epigenetic and sequence changes affect the same regulatory element. At all three post-cleavage stages, we found that differentially accessible OCRs contain a significantly greater proportion of sites with evidence of selection on the *He* branch relative to non-differentially accessible OCRs, including as much as an ~50 % proportional increase at the blastula stage (Fisher exact test, two-sided: $p < 5 \times 10^{-3}$) (**fig 5a**: top). This same trend continues at early cleavage (16/32-cell embryos), although not statistically significant. In contrast, we detected no statistical enrichment for positive selection within differentially accessible OCRs in the *Ht* genome – at early cleavage this relationship does approach significance, but no such trend exists at the later developmental stages (**fig. 5a**: bottom) (Fisher exact test, two-sided: $p > 0.05$).

Significant overlap in differential accessibility and positive selection in the *He* genome suggests a possible biological interaction between the two evolutionary mechanisms that may operate in tandem to alter regulatory function. We present two simple models for how this could occur. 1) *Accessibility first*: an increase or decrease in accessibility of a CRE at a particular developmental stage or tissue type makes the CRE more or less visible to selection. Following an increase in accessibility, mutations conferring adaptive changes in gene regulation during development would be selected for and increase in frequency within the population, whereas the same mutations in a closed OCR would have no effect; CREs with reduced accessibility, on the other hand, may accumulate mutations that reinforce its derived state, ensuring that it remains closed and unable to interact with transcription factors. 2) *Mutation first*: mutations within the regulatory element drive increased accessibility of the site, perhaps through recruitment of pioneer transcription factors, histone remodelers, or other chromatin modifiers, thereby activating the element's regulatory function; alternatively, if the regulatory element's function is detrimental to regulating gene expression, then mutations that disable the element's function or decrease its accessibility via changes to chromatin configuration would be selected. Regardless of mechanism, the disproportionate occurrence of positive selection within differentially accessible OCRs only in the *He* genome suggests a potentially synergistic set of gene regulatory processes may operate to drive developmental change during the evolution of its derived life history. The observation that this association is restricted to the *He* branch suggests that this interaction often produces adaptive changes in transcription during development. Tandem accessibility and mutational changes may be a common gene regulatory method to efficiently modify gene expression for organismal adaptation and innovation across organismal lineages.

Relationship of evolutionary regulatory mechanisms between proximal and distal elements

Next, we asked whether evolutionary changes in OCR sequence and chromatin accessibility differ between core promoters and more distal elements such as enhancers. We quantified changes in proximal (< 500 bp to translational start site of a gene) and distal (≥ 500 bp to translational start site of a gene) OCRs to test if either change covaries with promoter or enhancer evolution between species. Although many proximal and distal OCRs show evidence of positive selection on the *Ht* and *He* branches, we found no significant enrichment between the two categories of putative CREs (**fig. 5b**). This finding suggests mutations in both enhancers and promoters are likely important contributors in driving gene expression change, via altered binding and recruitment of regulatory factors and transcriptional machinery.

In contrast, accessibility changes shows a much greater asymmetry between proximal and distal elements (**fig. 5c**). While accessibility distributions are comparable at early cleavage (16/32-cell embryos), many more accessibility differences accumulate in proximal OCRs relative to distal ones later in development, after zygotic activation and tissue specification has commenced. For example, by the gastrula and larval stages, more than 20% of proximal elements are significantly more accessible in *Lv* and *Ht*, whereas this figure is ~5% for distal elements.

There are several possible biological and technical explanations for this result. Unlike enhancers, whose activity is often restricted to a single tissue or stage (Thurman, et al. 2012), promoters need to be accessible whenever the gene is transcribed, regardless of space and time. This has important implications for interpreting differences in chromatin configuration among species. At a biological level, greater promoter accessibility in the embryos of species with the ancestral lie history (*Ht* and *Lv*) may merely reflect overall greater transcriptional activity and

more changes in expression during development (**fig. 1,3; supplementary fig. 5**) (Israel, et al. 2016; Wang, et al. 2020). For a given gene, the accessibility of a promoter region is normally greater than that of nearby enhancers (Thurman, et al. 2012), which may lend greater statical power in detecting accessibility differences between species and life histories. In other words, smaller effect sizes associated with overall reduced accessibility of enhancers relative to promoters may contribute to the apparent enrichment of promoter accessibility changes. In addition, these chromatin configuration data were collected from whole embryos, so cell or lineage-specific differences in accessibility will be averaged within the sample, a technical factor that may be especially important for detecting accessibility changes of enhancer elements with tissue-specific dynamics. Lastly, promoter accessibility changes may simply be the most efficient way to rapidly modulate large changes in transcription, at least for some genes. Adjusting enhancer accessibility can certainly induce changes gene expression, but activating or shutting down a promoter may be a much more potent mechanism for large-scale changes in transcriptional regulation, such as those involved in rapid developmental evolution.

Relationships governing magnitude of gene expression and chromatin accessibility changes

Multiple lines of evidence presented above suggest that changes in the sequence or accessibility of *cis*-regulatory elements contributed to the evolution of development in *He*. The association between these changes and divergence in gene expression, however, is unclear. To understand this relationship, we compared mRNA expression fold-change of genes with accessibility fold-change of nearby OCRs at four developmental stages to measure instances of correlated changes in the magnitude of expression and chromatin configuration (**fig. 5d-e**). When all genes and OCRs are considered (**fig. 5e**: top row), there is no clear association between the magnitude and direction of accessibility and expression differences across developmental life

histories. This result is not surprising at a global scale, given that expression levels and accessibility are conserved between species for most genes and OCRs, respectively. However, a notable set of proximal sites have greater accessibility in species with the ancestral life history at the blastula, gastrula, and larval stages, consistent with overall greater promoter accessibility in *Ht* and *Lv* (see previous section). A similar result occurs when only differentially expressed genes and nearby OCRs are considered (**fig. 5e**: middle row). One possible explanation for the lack of an obvious relationship is that, while only genes that are differentially expressed between life histories are plotted, each of these genes has an average of 3.1 OCRs assigned to it (based on proximity). Each of these OCRs may not be active at a given stage of development, or they may function antagonistically to one another (i.e. regulatory elements can enhance or repress transcription), thus collectively dampening a correlative signal between gene expression and accessibility.

Lastly, we considered only differentially expressed genes and nearby OCRs with differential accessibility. This analysis reveals interesting relationships between transcription and chromatin configuration (**fig. 5e**: bottom row). As development progress, the relatively diffuse relationship between CRE accessibility and gene expression quickly tightens (compare 16/32-cell stage with larva). This likely reflects synchronization of zygotic genome activation with the embryo's *cis*-regulatory landscape as maternal mRNA reserves are depleted and nascent transcription and cell differentiation commences. Greater accessibility at the majority of proximal sites in *Ht* and *Lv* is still found in this subset of genes and OCRs, though smaller sets of promoters with greater accessibility in *He* and corresponding expression increases are detected (faint gray spots). Among distal elements, we observed a considerably more diffuse relationship between accessibility, expression, and life history, possibly reflecting the potential activating or

repressive functions of enhancers and downstream gene network interactions. At each developmental stage, the most common relationship was increased gene expression and accessibility in *Ht* and *Lv*, consistent with the preponderances of activating (as opposed to repressive) interactions within the ancestral dGRN. Despite these relationships, our results highlight the difficulty in correlating the magnitude of cross-species changes in gene expression and regulatory element accessibility at a genomic scale considering: **i)** Expression and accessibility of most genes and CREs are conserved; **ii)** CRE activity can have antagonistic effects on transcriptional regulation as activators or repressors, dependent on both the CRE sequence and *trans* environment specific to cell types or stages of development; **iii)** The activity and quiescence of multiple CREs regulating the same gene dampens regulatory signals when analyzed at the whole-embryo level; and **iv)** CRE accessibility is permissive rather than directive of regulatory factor binding and downstream gene expression such that “openness” does not necessarily indicate the CRE is actively regulating transcription.

Developmental trajectories of chromatin configuration and gene expression have evolved most frequently along the *H. erythrogramma* branch

To address the challenges associated with understanding the relationship between expression and accessibility changes of putative regulatory elements, we applied the “jump score” method to measure the magnitude of evolutionary changes in developmental trajectories, which are represented as distance in principal component space (Israel, et al. 2016). This metric is particularly useful because the jump score summarizes patterns of conservation or divergence in gene expression across development, rather than a single stage. Here we apply this approach to evolutionary changes in the accessibility of CREs across developmental trajectories.

We compared the expression jump score of 9,423 genes and the mean jump score of nearby OCRs to test how developmental changes in transcriptional regulation and chromatin status are distributed along the *He* and *Ht* branches. For example, expression of *insm*, a transcription factor involved in neurogenesis in sea urchins (McClay, et al. 2018), is conserved in across *He*, *Ht*, and *Lv* (**fig. 6a**), reflected by a low difference in jump scores between *He-Lv* and *Ht-Lv* and tight clustering of expression profiles in principal component space (**fig. 6b**). Similar expression profiles are accompanied by conserved *cis*-regulatory landscape dynamics through development in each species, exemplified by a near-zero mean OCR jump score difference and accessibility profile of ten nearby OCRs (**fig. 6c,d**). However, *brn 1/2/4*, a regulator of midgut patterning (Yuh, et al. 2005), has a divergent developmental gene expression trajectory in *He* relative to the ancestral state and large jump score differences in *He-Lv* and *Ht-Lv* branch comparisons (**fig. 6e,f**). This expression change is accompanied by significant reduction in chromatin accessibility at the blastula through larval stages in *He*, represented by a large mean difference in OCR jump scores along the *He* branch (**fig. 6g,h**). This locus is especially interesting considering *Hery* larvae do not feed and have lost development a functional through-gut.

Given the highly derived developmental program of *He* relative to the ancestral life history, we predicted that the most frequent changes in transcriptional regulation would occur along the *He* branch. Indeed, concerted modifications to OCR accessibility and gene expression profiles during *He* development are the most frequent set of changes among genes and OCRs with divergent expression and chromatin accessibility trajectories (1,582 genes) (**supplementary fig. 6; Data S7**). Furthermore, the least common combination of changes includes both expression and accessibility changes occurring on the *Ht* branch (1,114 genes), consistent with

this species' more conserved embryonic gene expression program. Thus, we find that local patterns of CRE accessibility have evolved in concert with divergent developmental expression patterns during life history evolution in *He*. Collectively, our results suggest that modified transcription factor access to regulatory elements can generate rapid and adaptive evolutionary changes to gene expression, and by extension, dGRN functionality.

Distinct *cis*-regulatory modifications differentially effect evolution of gene expression

Lastly, we tested whether changes in nucleotide sequence and accessibility within OCRs are distributed near genes that are differentially expressed between developmental life histories in order to better understand what evolutionary mechanisms may drive evolutionary changes in transcriptional regulation. We found a modest increase in differential expression of genes with nearby differentially accessible OCRs relative to genes lacking differentially accessible OCRs at blastula, gastrula and larval stages (~5-10% increase), but no such relationship exists during early cleavage, which may be expected given the transcriptome is composed almost entirely of maternally-derived transcripts during early cleavage (**fig. 7a**). Interestingly, this relationship between differential expression and *cis*-regulatory modification is magnified when comparing the frequency of signatures of positive selection within OCRs of *He* to OCRs lacking evidence of positive selection (~6-14% increase) and especially to OCRs with evidence of positive selection in *Ht* (15-33% increase) (**fig. 7b**). This finding is concordant with our predictions of evolutionary mechanisms of gene expression change, as selective changes in CRE sequences in *He* most frequently result in concordant divergence in developmental gene expression possibly associated with their derived life history strategy. Positive selection within CRE sequences of *Ht*, on the other hand, are much less often linked to changes in gene expression associated with life history, as *Ht* has retained the ancestral sea urchin developmental program. Thus, while the absolute

number of accessibility changes between developmental life histories is typically greater than detected instances of positive selection (**figs. 3-4**), mutations within CRE sequences are more likely to be associated with significant changes in gene regulation and expression relative to a no-change state.

This result may be attributable to the distinct characteristics of the two modes of *cis*-regulatory modification. CRE accessibility changes are transient and can be regulated within an organism across developmental stages and tissue types. Thus, relatively frequent changes in chromatin accessibility may be detected within a species at any given time as altered access to CREs may be less consequential for *fixed* expression differences within a species. Though speculative, frequent changes in chromatin accessibility may be an important mechanism for accumulating mutations within regulatory elements, as the “visibility” of mutations to natural selection can fluctuate between cell types, developmental stages, or across generations. Mutations, on the other hand, represent hardwired alterations to the *cis*-regulatory landscape of an individual as heritable sets of sequence changes. Although it may take longer for mutations to accumulate within CREs (and as a result may be relatively less common during divergence between species), their impact on gene expression may be detected more frequently as selection has favored the regulation change imparted by the CRE’s altered function. It is important to highlight that our conclusions regarding patterns of gene regulatory evolution are drawn from summarizing thousands of genes, so regulatory evolution and constraints associated with any specific gene will potentially include a unique arrangement and sequence of changes.

Conclusions

In this study, we utilized a recent life history switch in sea urchins to better understand the relationships between evolutionary change in *cis*-regulatory mechanisms, evolutionary

mechanisms, and divergence in embryonic gene expression. We show that two distinct molecular mechanisms within cis-regulatory elements contribute to evolutionary divergence in gene expression; changes in their DNA sequence and changes in the accessibility of those sequences. The former influences transcription factor binding while the latter limits access of transcription factors to DNA and prevent interactions between distal elements and proximal promoters. We document an elevated rate of evolutionary changes in both mechanisms in the derived relative to the ancestral life history, and particularly so near critical developmental network genes whose regulatory connections have been conserved for 10s to 100s million years. These concentrated evolutionary modifications may operate independently or synergistically to have rapidly reshaped transcriptional dynamics during developmental evolution in *H. erythrogramma* through alterations to cell fate specification and ultimately morphogenesis.

Collectively, our results indicate that natural selection can efficiently alter the influence of regulatory elements on transcription, through changes in both chromatin accessibility and transcription factor binding. These results provide support for the idea that evolutionary changes in chromatin configuration can contribute to adaptive changes in life history through their influence on developmental gene expression. More generally, the comparative approaches developed in this study present new ways of analyzing multimodal genomic datasets to understand the evolutionary mechanisms that operate on regulatory elements and their consequences for gene expression and organismal traits.

MATERIALS AND METHODS

Embryo Rearing and Sample Preparation

H. erythrogramma and *H. tuberculata* adult animals were collected near Sydney, NSW, Australia and maintained at 21°C in aquaria with circulating natural seawater. *L. variegatus*

individuals were collected near the Duke University Marine Lab in Beaufort, NC, USA and maintained at 23°C in aquaria at Duke University, Durham, NC, USA. Each sea urchin species was induced to spawn via injection of 0.5 M KCl solution into the coelom. Eggs were washed twice in 20 µm filtered seawater (FSW). Sperm was collected dry and stored at 4°C until usage. Washed eggs were fertilized with 1 µL of sperm, and after 10 minutes, were washed three times in FSW. For each species, three unique male-female pairs were crossed to produce three biologically-independent replicates of sea urchin embryos. Embryos were reared at 23°C in large glass culture dishes supplied with FSW, and water was changed every six hours. Because these species exhibit different developmental rates, developmental milestones, rather than absolute developmental time, were selected to maximize developmental synchrony within cultures and between species for comparison. These stages include 1) 4th/5th (EC), 2) 6th/7th (MC), and 3) 8th/9th (LC) cleavage stage embryos, as well as 4) hatched blastula (B), 5) mid-gastrula (G), and 6) early larva (L) (two-armed pluteus for *H. tuberculata* and *L. variegatus*; 32 hours post-fertilization for *H. erythrogramma*). Once a culture reached the appropriate stage, a subset of the culture was collected into a 1.5 mL Eppendorf tube for ATAC-seq processing.

ATAC-seq Protocol and Sequencing

ATAC sample preparation was carried out according to the Omni-ATAC-seq protocol (Corces, et al. 2017). For each replicate, embryos were washed once in 1 µm filtered seawater, lysed, and 50,000 nuclei were isolated for the transposition reaction using the Illumina TDE1 enzyme and tagmentation (TD) buffer (Cat. No. 20034197 and 20034198) (San Diego, CA, USA). Sequencing libraries for each replicate were amplified via qPCR as described in (Buenrostro, et al. 2015), and libraries were purified and size selected using Ampure XP Beads at a 1.8:1 bead volume:library volume (Beckman Coulter, Brea, CA, USA). Library quality and

transposition efficiency was accessed using a Fragment Analyzer and PROSize 2.0 (Agilent, Santa Clara, CA) at the Duke University Center for Genomic and Computational Biology. *H. erythrogramma* and *L. variegatus* libraries were sequenced on an Illumina HiSeq 4000 instrument using 50 bp SE sequencing. *H. tuberculata* libraries were sequenced on an Illumina NovaSeq 6000 instrument using 50 bp PE sequencing (only SE were used for data analysis).

ATAC-seq Data Analysis

Raw ATAC-seq reads were trimmed for quality and sequencing adapters using “cutadapt” (Martin 2011) v2.3 with the following parameters: -a CTGTCTCTTATACACATCT -q 20 --trim-n -m 40. Trimmed reads were then aligned to each species’ respective genome via “stampy” (Lunter and Goodson 2011) v.1.0.28 using the “—sensitive” set of parameters. ATAC-seq alignments were filtered for mitochondrial sequences and required an alignment quality score of at least 5 using “samtools” (Li, et al. 2009) v.1.9. Filtered alignments are available in BAM and Bedgraph format (see “Data Availability” below).

In this study, we aimed to compare evolution of orthologous regulatory elements so to accomplish this, we performed a series of liftOvers (Hinrichs, et al. 2006) to convert ATAC-seq alignments between genomic coordinates of each sea urchin species. We took an iterative, reciprocal strategy described below to minimize possible reference bias associated with converting between genome assemblies: 1) *H. erythrogramma*: $He \rightarrow Lv \rightarrow He$; 2) *H. tuberculata*: $Ht \rightarrow Lv \rightarrow He$; 3) *L. variegatus*: $Lv \rightarrow He \rightarrow Lv \rightarrow He$. At the end of filtering and coordinate conversion, all ATAC-seq alignments were referenced to the *H. erythrogramma* genome. LiftOver files are available in (Davidson, et al. 2021).

ATAC-seq peaks were called from these filtered, converted alignments using the “MACS2” v2.1.2 (Zhang, et al. 2008) callpeak function (parameters: -nomodel, -keep-

dup=auto, -shift 100, -extsize 200) for each developmental stage separately at an FDR threshold of 5%. Peak coordinates were merged using the “bedtools” (Quinlan and Hall 2010) v2.25 *merge* function requiring a peak overlap of at least 200 bp to be merged into a single peak. Lastly, for each sample, accessibility of each peak was measured with the “bedtools” *multiBamCov* function and imported into R v4.0.2 for statistical analysis (Data S1).

Estimates of positive selection

To measure branch-specific signatures of positive selection in the non-coding genome, the “adaptiPhy” (Berrio, et al. 2020) pipeline (<https://github.com/wodanaz/adaptiPhy>) for global tests of natural selection was followed. First, orthologous sequences for non-coding sites of interest (ATAC-seq peaks) were extracted from each species’ genome into FASTA format. Sequences were trimmed to include only contiguous DNA sequence using the *prunning.py* script and filtered using the *filtering.py* script, requiring a minimum alignment length of 75 base pairs. These trimmed and filtered alignments serve as “query” sequences of tests for selection. To generate a neutral reference for comparison, ten orthologous, neutral sites were randomly selected from a curated set of ~88,000 putatively neutral sequences (Davidson, et al. 2021) (Data S2) and concatenated into a single neutral reference sequence. Then, substitution rates of both the query and randomly concatenated neutral reference were estimated using phyloFit (Siepel and Haussler 2004), and the zeta score was calculated as the ratio of the query substitution rate to the neutral reference substitution rate. In addition, p-values of likelihood ratio tests for significant levels of branch-specific positive selection were calculated with adaptiPhy (Berrio, et al. 2020) pipeline using HyPhy (Pond, et al. 2005). For each query site, tests for positive selection were repeated 10 times against a unique neutral reference to ensure reproducible estimates of zeta and significance values for each non-coding site of interest. P-values and

substitution rates for all query and neutral sites were then imported in R v4.0.2. For each query site, the median zeta score and p-value for each test of positive selection across the ten replicate calculations were used for downstream analyses.

OCR Filtering

After accessibility and estimates of positive selection were calculated for each ATAC-seq peak, herein referred to as an “open chromatin region” (OCR), a series of filtering and quality control metrics were carried out to ensure only high confidence and quality peaks were compared between species. These filtering steps are as follows: 1) each OCR is required to have at least 75bp of contiguous, single copy sequence (see previous section) for accurate estimations of selection ; 2) a local composition complexity (LCC) (Konopka 2001) value of 1.9 or more was required of the OCR for each species to remove repetitive or other low-complexity sequences that may generate inaccurate estimations of selection (module: biopython.org/docs/1.75/api/Bio.SeqUtils.lcc.html); 3) a CPM of 3 or more was required in at least 9 samples to remove OCRs with extremely low accessibility across samples; 4) the midpoint of the OCR must lie within 25 Kb (in either direction) of the translational start site of a gene model. The 25 Kb window for assigning OCRs to genes was selected on the basis that the median and mean intergenic distance within the *He* genome are 21.4 Kb and 31.9 Kb, respectively. Given nearly no prior knowledge is known of the *cis*-regulatory landscape for these sea urchin species, these stringent filtering methods were carried out in order to maximize confidence in comparisons of non-coding sequence evolution and function. This resulted in a final set 35,788 high-confidence OCRs for cross-species analysis (Data S3).

ATAC-seq Statistical Analysis

Principal component analysis and variance partitioning was performed on the regularized log-transformed (rlog) data (Love, et al. 2014) of filtered OCR accessibility counts via the *prcomp* function and the “variancePartition” v1.18.3 package (Hoffman and Schadt 2016) in R using the formula: $\sim (1 \mid \text{Stage}) + (1 \mid \text{Species}) + (1 \mid \text{Life History}) + (1 \mid \text{Genus}) + (1 \mid \text{Stage: Life History})$. Raw accessibility counts of the filtered OCRs were loaded into DESeq2 (Love, et al. 2014) v1.30 to calculate differential accessibility between sample groups. Differentially accessible sites between life history strategies were classified as having a 2-fold accessibility difference and supported by an FDR of 10% in *He* vs *Ht* and well as *He* vs *Lv* comparisons (including the sign of the accessibility difference being in the same direction) (see **supplementary fig. 7** for all significance cutoff combinations of differential accessibility results). For graphical purposes, the average fold-change in the *He* vs *Ht* and *He* vs *Lv* comparisons is displayed within figures. Significant levels of positive selection in either the *He* or *Ht* branch were classified as OCRs having a zeta value greater than 1.25 and supported by a median false-discovery rate less than 10% in one species, and failing to meet both criteria in the other. Motif enrichment analyses were carried out with “Homer” v4.11 (Heinz, et al. 2010) via the findMotifsGenome.pl script using all OCRs as the background set and regions size of 200 bp. Sea urchin GRN genes are provided in Data S4

RNA-seq Expression Analyses

Raw RNA-seq reads from eggs, 4-cell, 16-cell + 32-cell (collectively EE) , blastula (B), gastrula (G), and early larva (L) stage embryos of *He*, *Ht*, and *Lv* were retrieved from (Israel, et al. 2016), trimmed and filtered for low quality bases and reads with Trimmomatic (Bolger, et al. 2014), and aligned to each species respective genomes and gene models with STAR (Dobin, et al. 2013). From these alignments, mRNA expression was estimated with Salmon (Patro, et al.

2017) and loaded to R for statistical analysis. Read counts for summed to each gene's best match to the *S. purpuratus* v4.2 gene models to generate a common reference for expression comparisons between species as described in (Israel, et al. 2016) (Data S5). Differentially expressed genes between life histories were called as having a fold-change in expression > 2 and supported by an FDR of 10% or less in *He* vs *Ht* and well as *He* vs *Lv* comparisons (including the sign of the accessibility difference being in the same direction) (Data S6). These criteria were chosen to make RNA and ATAC-seq analyses between species comparable as possible.

Principal Component Distances of Developmental Profile

To identify OCRs and genes whose developmental regulatory profiles diverged in *H. erythrogramma* or *H. tuberculata*, we calculated principal component distances of chromatin accessibility and gene expression profile shapes between species throughout development as described in (Israel, et al. 2016). Mean accessibility and expression profiles of regularized log-transformed (rlog) OCR and genes counts were standardized using the *standardize* function of the “mfuzz” v2.48.0 package (Kumar and M 2007). From these data, scaled distance measures (“jump scores”) between orthologous non-coding sites or genes were calculated using loadings of the first two principal components (Data S7): $d_{HeLv} =$

$$\sqrt{PC1(x_{Lv} - x_{He})^2 + PC2(y_{Lv} - y_{He})^2} ;$$

$d_{HtLv} = \sqrt{PC1(x_{Lv} - x_{Ht})^2 + PC2(y_{Lv} - y_{Ht})^2}$. A significant change in the developmental trajectory of gene expression or OCR accessibility was determined to have a jump score difference of at 0.15 on one branch over the other .

DATA AVAILABILITY

Raw ATAC-seq reads are available in FASTQ format on NCBI's Short Read Archive (accession number pending publication). Filtered alignments of ATAC-seq reads before and after between-

species liftOver files are available on Dryad (accession number pending publication). Results files of ATAC-seq analyses are available as Supplementary Data files.

AUTHOR CONTRIBUTIONS

PLD and GAW conceived and designed the study. PLD and MB performed sample collection. PLD performed the data analysis. PLD and GAW wrote the manuscript. All authors revised the manuscript.

ACKNOWLEDGEMENTS

We would like to thank the Sydney Institute of Marine Sciences and their staff for assistance with animal care, laboratory set up, and carrying out experiments. We also thank Allison Edgar and Hannah Devens for assistance preparing sea urchin embryo cultures associated with this work. This work was supported by the National Science Foundation Division of Integrative Organismal Systems (award no. 1929934 to GAW).

COMPETING INTERESTS

The authors have no competing interests to declare.

REFERENCES

- Berrio A, Haygood R, Wray GA. 2020. Identifying branch-specific positive selection throughout the regulatory genome using an appropriate proxy neutral. BMC Genomics 21:359.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.
- Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol 109:21 29 21-29.
- Byrne M, Sewell MA. 2019. Evolution of maternal lipid provisioning strategies in echinoids with non-feeding larvae: selection for high-quality juveniles. Marine Ecology Progress Series 616:95-106.
- Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, et al. 2017. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods 14:959-962.
- Cusanovich DA, Reddington JP, Garfield DA, Daza RM, Aghamirzaie D, Marco-Ferreres R, Pliner HA, Christiansen L, Qiu X, Steemers FJ, et al. 2018. The cis-regulatory dynamics of embryonic development at single-cell resolution. Nature 555:538-542.
- Davidson EH, Erwin DH. 2006. Gene regulatory networks and the evolution of animal body plans. Science 311:796-800.
- Davidson PL, Guo H, Swart JS, Massri AJ, Edgar A, Wang L, Berrio A, Devens HR, Koop D, Cisternas P, et al. 2021. Recent reconfiguration of an ancient developmental gene regulatory network in *Heliocidaris* sea urchins.
- Davidson PL, Guo H, Wang L, Berrio A, Zhang H, Chang Y, Soborowski AL, McClay DR, Fan G, Wray GA. 2020. Chromosomal-level genome assembly of the sea urchin *Lytechinus variegatus* substantially improves functional genomic analyses. Genome Biology and Evolution 12:1080-1086.
- Davidson PL, Thompson JW, Foster MW, Moseley MA, Byrne M, Wray GA. 2019. A comparative analysis of egg provisioning using mass spectrometry during rapid life history evolution in sea urchins. Evolution & Development 21:188-204.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15-21.
- Emlet RB. 1995. Larval spicules, cilia, and symmetry as remnants of indirect development in the direct developing sea urchin *Heliocidaris erythrogramma*. Dev Biol 167:405-415.
- Falkner I, Sewell MA, Byrne M. 2015. Evolution of maternal provisioning in ophiuroid echinoderms: characterisation of egg composition in planktotrophic and lecithotrophic developers. Marine Ecology Progress Series 525:1-13.
- Flatt T, Heyland A. 2011. Mechanisms of life history evolution: the genetics and physiology of life history traits and trade-offs: Oxford University Press.
- Hart MW. 2002. Life history evolution and comparative developmental biology of echinoderms. Evolution & Development 4:62-71.
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38:576-589.
- Henry JJ, Wray GA, Raff RA. 1990. The dorsoventral axis is specified prior to 1st cleavage in the direct developing sea urchin *Heliocidaris erythrogramma*. Development 110:875-

884.
Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F. 2006. The UCSC genome browser database: update 2006. *Nucleic acids research* 34:D590-D598.
Hoffman GE, Schadt EE. 2016. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics* 17:483.
Israel JW, Martik ML, Byrne M, Raff EC, Raff RA, McClay DR, Wray GA. 2016. Comparative developmental transcriptomics reveals rewiring of a highly conserved gene regulatory network during a major life history switch in the sea urchin genus *Helicoidaris*. *PLoS Biol* 14:e1002391.
Jhanwar S, Malkmus J, Stolte J, Romashkina O, Zuniga A, Zeller R. 2021. Conserved and species-specific chromatin remodeling and regulatory dynamics during mouse and chicken limb bud development. *Nat Commun* 12:5685.
Klueg KM, Harkey MA, Raff RA. 1997. Mechanisms of evolutionary changes in timing, spatial expression, and mRNA processing in the *msp130* gene in a direct-developing sea urchin, *Helicoidaris erythrogramma*. *Developmental Biology* 182:121-133.
Konopka AK. 2001. Sequence complexity and composition. eLS.
Koop D, Cisternas P, Morris VB, Strbenac D, Yang JY, Wray GA, Byrne M. 2017. Nodal and BMP expression during the transition to pentamery in the sea urchin *Helicoidaris erythrogramma*: insights into patterning the enigmatic echinoderm body plan. *BMC Dev Biol* 17:4.
Kumar L, M EF. 2007. Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* 2:5-7.
Lewis JJ, Reed RD. 2019. Genome-wide regulatory adaptation shapes population-level genomic landscapes in *Heliconius*. *Mol Biol Evol* 36:159-173.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078-2079.
Livraghi L, Hanly JJ, Van Bellghem SM, Montejo-Kovacevich G, van der Heijden ES, Loh LS, Ren A, Warren IA, Lewis JJ, Concha C, et al. 2021. Cortex cis-regulatory switches establish scale colour identity and pattern diversity in *Heliconius*. *eLife* 10.
Love AC, Raff RA. 2006. Larval ectoderm, organizational homology, and the origins of evolutionary novelty. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 306:18-34.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550.
Lunter G, Goodson M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome research* 21:936-939.
Marshall DJ, Krug PJ, Kupriyanova EK, Byrne M, Emler RB. 2012. The biogeography of marine invertebrate life histories. *Annual Review of Ecology, Evolution, and Systematics*, Vol 43 43:97-+.
Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* 17:10-12.
McClay DR, Miranda E, Feinberg SL. 2018. Neurogenesis in the sea urchin embryo is initiated uniquely in three domains. *Development* 145.
McEdward LR, Miner BG. 2001. Larval and life-cycle patterns in echinoderms. *Canadian*

Journal of Zoology 79:1125-1170.

Morgan SG. 2020. Life and death in the plankton: larval mortality and adaptation. Ecology of marine invertebrate larvae: 279-321.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14:417-419.

Pechenik JA. 1999. On the advantages and disadvantages of larval stages in benthic marine invertebrate life cycles. Marine Ecology Progress Series 177:269-297.

Pond SL, Frost SD, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676-679.

Prowse TAA, Falkner I, Sewell MA, Byrne M. 2009. Long-term storage lipids and developmental evolution in echinoderms. Evolutionary Ecology Research 11:1069-1083.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-842.

Raff RA. 1987. Constraint, flexibility, and phylogenetic history in the evolution of direct development in sea urchins. Developmental Biology 119:6-19.

Raff RA, Byrne M. 2006. The active evolutionary lives of echinoderm larvae. Heredity 97:244-252.

Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE, Gardner PP, Clarke JA, Baker AJ, Clamp M, et al. 2019. Convergent regulatory evolution and loss of flight in paleognathous birds. Science 364:74-78.

Shibata Y, Sheffield NC, Fedrigo O, Babbitt CC, Wortham M, Tewari AK, London D, Song L, Lee B-K, Iyer VR, et al. 2012. Extensive evolutionary changes in regulatory element activity during human origins are associated with altered gene expression and positive selection. PLOS Genetics 8:e1002789.

Siepel A, Haussler D. 2004. Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. Mol Biol Evol 21:468-488.

Smith MS, Turner FR, Raff RA. 2008. Nodal expression and heterochrony in the evolution of dorsal-ventral and left-right axes formation in the direct-developing sea urchin *Heliocidaris erythrogramma*. Journal of Experimental Zoology Part B-Molecular and Developmental Evolution 310b:609-622.

Strathmann RR. 1978. The evolution and loss of feeding larval stages of marine invertebrates. Evolution 32:894-906.

Strathmann RR. 1985. Feeding and nonfeeding larval development and life history evolution in marine invertebrates. Annual Review of Ecology and Systematics 16:339-361.

Sun DA, Patel NH. 2021. Identification and classification of cis-regulatory elements in the amphipod crustacean *Parhyale hawaiiensis*. bioRxiv:2021.2009.2016.460328.

Swain-Lenz D, Berrio A, Safi A, Crawford GE, Wray GA. 2019. Comparative analyses of chromatin landscape in white adipose tissue suggest humans may have less beigeing potential than other primates. Genome Biology and Evolution 11:1997-2008.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al. 2012. The accessible chromatin landscape of the human genome. Nature 489:75-82.

Tonelli C, Chio IIC, Tuveson DA. 2018. Transcriptional Regulation by Nrf2. Antioxid Redox Signal 29:1727-1745.

Tsushima M, Byrne M, Amemiya S, Matsuno T. 1995. Comparative biochemical studies of carotenoids in sea urchins. 3. Relationship between developmental mode and

carotenoids in the Australian echinoids *Heliocidaris erythrogramma* and *H. tuberculata* and a comparison with Japanese species. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* 110:719-723.

van der Burg KRL, Lewis JJ, Brack BJ, Fandino RA, Mazo-Vargas A, Reed RD. 2020. Genomic architecture of a genetically assimilated seasonal color pattern. *Science* 370:721-725.

Vance RR. 1973. On the reproductive strategies of marine benthic invertebrates. *The American Naturalist* 107:339-352.

Walton KD, Warner J, Hertzler PH, McClay DR. 2009. Hedgehog signaling patterns mesoderm in the sea urchin. *Dev Biol* 331:26-37.

Wang L, Israel JW, Edgar A, Raff RA, Raff EC, Byrne M, Wray GA. 2020. Genetic basis for divergence in developmental gene expression in two closely related sea urchins. *Nat Ecol Evol* 4:831-840.

Williams DHC, Anderson DT. 1975. Reproductive system, embryonic development, larval development and metamorphosis of sea urchin *Heliocidaris erythrogramma* (Val) (Echinoidea-Echinometridae). *Australian Journal of Zoology* 23:371-403.

Wilson KA, Andrews ME, Raff RA. 2005. Dissociation of expression patterns of homeodomain transcription factors in the evolution of developmental mode in the sea urchins *Heliocidaris tuberculata* and *H. erythrogramma*. *Evolution & Development* 7:401-415.

Wilson KA, Andrews ME, Rudolf Turner F, Raff RA. 2005. Major regulatory factors in the evolution of development: the roles of gooseoid and Msx in the evolution of the direct-developing sea urchin *Heliocidaris erythrogramma*. *Evolution & Development* 7:416-428.

Wittkopp PJ, Kalay G. 2012. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics* 13:59-69.

Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* 8:206-216.

Wray GA. 2021. Extreme phenotypic divergence and the evolution of development. In. *Current Topics in Developmental Biology*.

Wray GA. 1996. Parallel evolution of nonfeeding larvae in echinoids. *Systematic Biology* 45:308-322.

Wray GA, Raff RA. 1989. Evolutionary modification of cell lineage in the direct-developing sea urchin *Heliocidaris erythrogramma*. *Developmental Biology* 132:458-470.

Wray GA, Raff RA. 1990. Novel origins of lineage founder cells in the direct-developing sea urchin *Heliocidaris erythrogramma*. *Dev Biol* 141:41-54.

Yuh CH, Dorman ER, Davidson EH. 2005. Brn1/2/4, the predicted midgut regulator of the endo16 gene of the sea urchin embryo. *Dev Biol* 281:286-298.

Zaret KS, Carroll JS. 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* 25:2227-2241.

Zhang Y, Liu T, Meyer CA, Eickhout J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome biology* 9:R137.

824 **FIGURE and LEGENDS**

825

826 **Figure 1: Global variation in chromatin accessibility between sea urchin embryos is largely**

827 **driven by developmental stage and life history strategy. A)** Three focal sea urchin species of

828 this study, including two planktotrophic species representing the ancestral life history (light

829 green: *Lytechinus variegatus*; dark green: *Heliocidaris tuberculata*) and one lecithotrophic

830 species representing the derived life history (orange: *Heliocidaris erythrogramma*). This color

831 scheme is consistent throughout the study. **B)** Developmental stages assayed in this study for

832 each species. The ancestral life history is illustrated above, with corresponding developmental
 833 stages of the derived life history below. Developmental times are shown for *Heliocidaris*
 834 *tuberculata* at 21°C. Note, however, samples were collected on the basis of developmental stage,
 835 not hours post-fertilization, for all three species. **C-D)** Principal component analysis of variation
 836 in normalized chromatin accessibility among **C)** all orthologous OCRs and **D)** only OCRs near
 837 dGRN genes through embryogenesis. OCR: open chromatin region; dGRN: developmental gene
 838 regulatory network.

Figure 2: Summary statistics of regulatory element gains, losses, accessibility, and positive selection. **A)** Union set of all OCRs, including structural differences between species such as lineage-specific gains and losses as well as 35,788 1:1:1 orthologous OCRs shared between all three species. These orthologous OCRs are the primary focus of this study. Breakdown of **B)** OCR differentially accessibility between life history strategies and **C)** OCR distance to gene

845 models for these orthologous sites. **D)** Overlap of OCR differential accessibility and evidence of
 846 positive selection on the *He*, *Ht*, or both branches. **E)** Percentage of each species genome that is
 847 “open” at each stage of development. “Open” is defined as the cumulative portion of the genome
 848 represented by all OCRs (orthologous and lineage-specific) with significant levels of
 849 accessibility for a given stage and species (see **supplementary fig. 3** for additional information).
 850 OCR: open chromatin region; *He*: *Heliocidaris erythrogramma*; *Ht*: *Heliocidaris tuberculata*.

851
 852 **Figure 3: The chromatin landscape of the derived life history is far less accessible than that of species with the ancestral life**
 853 **history mode, particularly near core dGRN genes. A-B)** Number and proportions of differentially accessible OCRs that are
 854 significantly more open in *H. erythrogramma* (orange, top) or species with the ancestral life history strategy (green, bottom) for **A)** all
 855 orthologous OCRs and **B)** only OCRs near dGRN genes. **C)** Example of reduced chromatin accessibility of a proximal and distal OCR
 856 in *H. erythrogramma* during embryogenesis near the gene *patched*. OCR: open chromatin region; dGRN: developmental gene
 857 regulatory network.

Figure 4: Sequence changes indicative of positive selection are enriched within *cis*-regulatory elements near dGRN genes and other transcription factors in the *H. erythrogramma* genome. A) Zeta values (ratio of OCR substitution rate to that of neutrally evolving reference sequences) for *H. erythrogramma* (top) and *H. tuberculata* (bottom) across the genome. OCRs with significant

862 support for lineage-specific evidence of positive selection are noted for *H. erythrogramma* (orange) and *H. tuberculata* (dark green),
863 and those nearby dGRN genes are outlined in purple. **B)** Fold-change in proportion of OCRs with more frequent evidence of positive
864 selection in either *H. erythrogramma* or *H. tuberculata*, partitioned by functional categories of nearby genes (see **supplementary fig.**
865 **5** for all categories). **C)** Transcription factor binding motif enrichment results of OCRs near dGRN genes with evidence of positive
866 selection in *H. erythrogramma*. Detection mode indicates OCR sequences were searched against a (i) “known” set of optimized motif
867 models or (ii) enriched motifs within the OCR were identified “de-novo” and matched to a known set of motifs. **D)** Embryonic mRNA
868 expression levels for *Nuclear factor erythroid 2-related factor (NRF)* for each species. OCR: open chromatin region; dGRN:
869 developmental gene regulatory network.

871 **Figure 5: Independent and synergistic changes to regulatory element function have evolved during life history evolution in *H.***
 872 ***erythrogramma*. A)** Proportion of differentially accessible or non-differentially accessible OCRs with evidence of positive selection
 873 on the *H. erythrogramma* branch (orange, top) or *H. tuberculata* branch (dark green, bottom). Two-sided, Fisher's exact test: ** $p \leq 5$
 874 $\times 10^{-3}$; *** $p \leq 5 \times 10^{-4}$. **B)** Proportion of proximal of distal OCRs with evidence of positive selection in either *Heliocidaris* species. **C)**
 875 Differential accessibility of proximal and distal OCRs at four stages of development. Orange indicates significantly greater
 876 accessibility in *H. erythrogramma*, dark green indicates significantly greater accessibility in *H. tuberculata*. **D-E)** Fold-change of
 877 mRNA expression vs nearby OCR accessibility between life history strategies for proximal (left) and distal (right) OCRs.
 878 Relationships are shown for all genes and OCRs (E: top row), significantly differentially expressed genes (E: middle row), and
 879 significantly differentially expressed genes *and* differentially accessible OCRs (E: bottom row). OCR: open chromatin region; DE:
 880 differentially expressed (genes); DA: differentially accessible (OCRs).

Figure 6: Quantifying developmental expression and accessibility changes in principal component space predicts divergent patterns of gene regulation. A-D) An example of conserved developmental gene expression and chromatin accessibility for *insm*. A) Developmental mRNA expression profile of *insm* for each sea urchin species and B) their coordinates in PCA space. C) Coordinates in PCA space of developmental accessibility of orthologous OCRs near *insm* and D) normalized chromatin accessibility coverage of these same OCRs through embryogenesis for each species. E-H) An example of divergent developmental gene expression and chromatin accessibility associated with life history for *brn 1/2/4*. E) Developmental mRNA expression profile of *brn 1/2/4* for each

888 sea urchin species and **F)** their coordinates in PCA space. **G)** Coordinates in PCA space of developmental accessibility of orthologous
889 OCRs near *brn 1/2/4* and **H)** normalized chromatin accessibility coverage of these same OCRs through embryogenesis for each
890 species. OCR: open chromatin region; PCA: principal component analysis.

Figure 7: Evolutionary changes to *cis*-regulatory element function are frequently associated with expression changes of nearby genes underlying life history evolution in sea urchins.

Percentage of genes that are differentially expressed between developmental life history strategies at four developmental stages, distinguished as having at least one nearby OCR with **A)** differential accessibility or **B)** evidence of lineage-specific positive selection. OCR: open chromatin region.