RECENT RECONFIGURATION OF AN ANCIENT DEVELOPMENTAL GENE REGULATORY NETWORK IN *HELIOCIDARIS* SEA URCHINS

Phillip L Davidson[1], Haobing Guo[2,3], Jane S Swart[1], Abdull J Massri[1], Allison Edgar[1], Lingyu Wang[1], Alejandro Berrio[1], Hannah R Devens[1], Demian Koop[5], Paula Cisternas[5], He Zhang[3], Yaolei Zhang[2,4], Maria Byrne[5], Guangyi Fan[2,3], Gregory A Wray[1,6]*

1. Department of Biology, Duke University, Durham, NC, USA
2. BGI-Qingdao, BGI-Shenzhen, Qingdao, China
3. State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen, China
4. Qingdao-Europe Advanced Institute for Life Sciences, BGI-Qingdao, BGI-Shenzhen, Qingdao, China
5. School of Life and Environmental Science, The University of Sydney, Sydney, NSW, Australia
6. Center for Genomic and Computational Biology, Duke University, Durham, NC, USA

* Corresponding author: gwray@duke.edu

ABSTRACT

Changes in developmental gene regulatory networks (dGRNs) underlie much of the diversity of life[1], but the evolutionary mechanisms that operate on interactions with these networks remain poorly understood. Closely related species with extreme phenotypic divergence provide a valuable window into the genetic and molecular basis for changes in dGRNs and their relationship to adaptive changes in organismal traits. Here we analyze genomes, epigenomes, and transcriptomes during early development in two sea urchin species in the genus *Heliocidaris* that exhibit highly divergent life histories and in an outgroup species. Signatures of positive selection and changes in chromatin status within putative gene regulatory elements are both enriched on the branch leading to the derived life history, and particularly so near core dGRN genes; in contrast, positive selection within protein-coding regions have at most a modest enrichment in branch and function. Single-cell transcriptomes reveal a dramatic delay in cell fate specification in the derived state, which also has far fewer open chromatin regions, especially near dGRN genes with conserved roles in cell fate specification. Experimentally perturbing the function of three key transcription factors reveals profound evolutionary changes in the earliest events that pattern the embryo, disrupting regulatory interactions previously conserved for ~225 million years. Together, these results demonstrate that natural selection can rapidly reshape developmental gene expression on a broad scale when selective regimes abruptly change and that even highly conserved dGRNs and patterning mechanisms in the early embryo remain evolvable under appropriate ecological circumstances.

39 **MAIN**

40 The well-defined dGRN of early development in sea urchins[2] provides a powerful
41 framework for investigating the evolution of embryonic patterning mechanisms. Interactions
42 within this dGRN are almost completely conserved among species that diverged ~30-40 million
43 years (my) ago[3,4], with some interactions conserved for ~225 my[5], ~275 my[6], or even ~480
44 my[7,8]. Although developmental constraints may be responsible, a confound remains untested: the
45 species with deeply conserved developmental mechanisms all share the same life history mode,
46 involving low maternal provisioning and an extended feeding larval phase. Species with derived
47 life histories involving massive maternal provisioning and highly-abbreviated, nonfeeding pre-
48 metamorphic development have evolved on multiple occasions within sea urchins[9-11], possibly in
49 response to lower or more unpredictable food availability[12]. These species reveal how conserved
50 regulatory interactions and patterning mechanisms respond to major shifts in selective regimes.
51 The Australian sea urchin genus *Heliocidaris* includes two recently diverged species: *H.*
52 *tuberculata* representing the ancestral life history and *H. erythrogramma* the derived state[13]
53 (Fig.1a). The shift to nonfeeding development radically alters natural selection on development:
54 with feeding no longer necessary, high mortality rates in the plankton[14] impose strong selection
55 to decrease time to metamorphosis[10]. Numerous anatomical features and gene expression profiles
56 of early development that are broadly conserved among sea urchins differ markedly between
57 these closely related species[15-18](Fig.1b). We sought to learn whether these recently evolved
58 differences are merely superficial and mask deeply conserved developmental mechanisms, or
59 whether they are the product of substantive evolutionary changes in early cell fate specification
60 and dGRN organization.
61

62 **Natural selection has sculpted the regulatory landscape of *H. erythrogramma***

63 We took advantage of the recent (~5 my) divergence between the two *Heliocidaris*
64 species to carry out detailed analyses of orthologous coding and noncoding regions of the
65 genome, focusing on the transcription factors and regulatory elements that constitute the
66 backbone of the dGRN and underlie cell fate specification mechanisms. Genomes of *H.*
67 *erythrogramma* and *H. tuberculata* were each sequenced, assembled into twenty-one full length
68 chromosomes, and annotated (Fig.1c). Genome sequences were then aligned to one another and
69 to that of *Lytechinus variegatus*[19], an outgroup representing the ancestral life history condition
70 (Fig.1a).

71 To understand how natural selection altered the genomes of *Heliocidaris* during the
72 evolution of nonfeeding development, we began by testing for evidence of branch-specific
73 positive selection within single-copy protein coding regions. At a genome-wide scale, we found
74 statistical support for modest enrichment of positive selection along the *H. erythrogramma*
75 branch, but not the *H. tuberculata* branch, when considering the full set of genes (Fisher exact
76 test, two-sided: $p<1.33x10^{-3}$). Of note, coding sequences of dGRN genes showed no enrichment
77 of positive selection on either branch (Fig.1d-e). This result provides little support for the idea
78 that changes in transcription factor structure and function are primarily responsible for the
79 extensive modifications in development and life history in *H. erythrogramma*. Scant evidence of
80 positive selection in the coding sequences of dGRN genes likely reflects pleotropic constraint
81 imposed by the multiple functions that their encoded transcription factors execute during cell
82 type specification and differentiation.

83 Therefore, we hypothesized that functional changes in regulatory elements are instead
84 largely responsible for these trait changes. We used ATAC-seq to identify 27,322 one-to-one

orthologous open chromatin regions (OCRs) representing putative regulatory elements in blastula-stage embryos of the two *Heliocidaris* species and of *L. variegatus*, a time point when initial cell fates have been specified in the ancestral dGRN. We then tested for branch-specific positive selection within these OCRs using an approach[20] analogous to that described above for protein-coding regions. At a global scale, these putative enhancer and promoter regions are enriched for evidence of positive selection on the branch leading to *H. erythrogramma* (545) relative to *H. tuberculata* (347) (Fisher exact test, two-sided: p<1.33e-11) (Fig.1f-g). This higher incidence in signatures of positive selection specifically within OCRs on the *H. erythrogramma* branch is indicative of weak positive selection on regulatory element function that is remarkably widespread within its genome and is consistent with our earlier finding that many expression differences between the two *Heliocidaris* species are genetically based in *cis*[21].

Strikingly, signals of *H. erythrogramma*-specific positive selection are even more enriched when only considering OCRs near dGRN genes (Fig.1f-g; difference in median zeta: 0.182; Fisher exact test, two-sided: p<5.33e-5). In all, 26 putative regulatory elements located near 23 distinct dGRN genes exhibit evidence of positive selection on the *H. erythrogramma* branch, as opposed to just four on the *H. tuberculata* branch (Fig.1c,g). These 23 genes represent 17.0% of the total within the defined dGRN with a nearby OCR, a marked enrichment compared with the remainder of the genome, where positive selection is detected in OCRs near just 5.7% of genes (Fisher exact test, two-sided: p<4.92e-4: Fig.2c).

Two distinct regulatory mechanisms underlie divergence in transcriptomes

We also observed a striking decrease in chromatin accessibility of putative regulatory elements throughout the *H. erythrogramma* genome relative to both species representing the ancestral life history (Fig.2a). Of 2,625 orthologous, differentially accessible OCRs between

developmental modes, 1,795 sites (68.4%) are significantly less accessible in *H. erythrogramma*. Because decreased chromatin accessibility can limit transcription factor access to regulatory elements and because most regulatory interactions in the early sea urchin embryo involve activation of transcription[22], widespread evolutionary reduction in chromatin accessibility throughout the genome in *H. erythrogramma* embryos suggests an important role for evolutionary changes in chromatin configuration for divergence in gene expression, in this cases associated with generally decreased or delayed zygotic transcription for many genes. This interpretation is consistent with indications of a broad delay in embryonic cell fate specification in this species[17,23-27].

In a previous study[17] we analyzed changes in temporal gene expression profiles during early development within *Heliocidaris* and found that the largest changes are concentrated on the branch leading to *H. erythrogramma* and are enriched for developmental regulatory genes generally and dGRN genes specifically. Results reported here suggest that these derived expression profiles are the product of two distinct molecular mechanisms that alter transcription factor binding: changes in nucleotide sequence and changes in chromatin configuration (Fig.2c-d). The former may alter protein:DNA binding while the latter may alter protein access to regulatory elements. Both modes of regulatory evolution are concentrated near dGRN genes relative to the rest of the transcriptome in *H. erythrogramma* (Fig.2c). Notably, accelerated sequence evolution or altered chromatin state (or both) is present in an OCR near of a differentially expressed dGRN gene ~3 times more frequently than the rest of the transcriptome in *H. erythrogramma*, while no such relationship is evident in *H. tuberculata*.

The distribution of both mechanisms of regulatory evolution is highly nonrandom within the genome (enriched near differentially expressed genes and near developmental regulatory

131 genes) and phylogenetically (enriched on the *H. erythrogramma* branch). These departures from
132 the null expectation of random distribution (i.e., resulting from genetic drift) suggest that many
133 of the specific changes are adaptive. Adults of the two *Heliocidaris* species occupy overlapping
134 habitats and ranges[28], making the suite of derived life history traits that evolved on the *H.*
135 *erythrogramma* branch the most plausible driver for many of the extensive gene regulatory
136 changes.

137 **The timing of cell fate specification has changed in *H. erythrogramma***

138 To understand how changes in the regulation of gene expression in the *H. erythrogramma*
139 embryo influenced developmental mechanisms and life history traits, we leveraged information
140 about the ancestral dGRN to examine embryonic cell fate specification. The earliest zygotic
141 patterning event in the ancestral state involves specification of skeletogenic and germ cell fates
142 following two successive unequal cleavages of vegetal blastomeres[29]. We focus here on the well-
143 characterized skeletogenic cell lineage, which rapidly establishes a distinct transcriptional state[30]
144 and, within 24 hours after fertilization, undergoes an epithelial-to-mesenchymal transition, fully
145 differentiates, and begins to synthesize a complex larval endoskeleton (Fig.1b). Specification and
146 maintenance of the skeletogenic cell fate is regulated by interactions between ~11 transcription
147 factors[31] (Fig.4a). These developmental events and most of the underlying dGRN interactions are
148 conserved across >225 my of sea urchin evolution[6].

149 Morphological development of the skeletogenic cell lineage in *H. erythrogramma* differs
150 in several regards from this ancestral state: cleavage divisions are all equal, no cells ingress
151 before gastrulation, and the larval skeleton is delayed and reduced[18] (Fig.1b). To understand
152 whether underlying developmental mechanisms are conserved despite these overt morphological
153 differences, we carried out single-cell RNA-sequencing (scRNA-seq) of blastula stage *H.*

erythrogramma embryos and compared the results with our published scRNA-seq data from *L. variegatus*[32]. In the ancestral condition, several cell clusters express distinct suites of transcription factors predicted by the dGRN, including the skeletogenic cells and germline (Figure 3a), reflecting early cell fate specification and rapid divergence in transcriptional states in *L variegatus*. These regulatory states are very similar in *Strongylocentrotus purpuratus*[33], another sea urchin representing the ancestral condition.

In *H. erythrogramma*, fewer distinct cell clusters are apparent at the same resolution and same developmental stage (Fig.3a), indicating a slower diversification of distinct transcriptional states in the early embryo. Further, the clusters do not contain similar suites of transcription factors to those in the ancestral state, and none corresponds to the distinctive skeletogenic cell lineage of *L. variegatus* and *S. purpuratus* that is established earlier in their development. These ensemble differences reflect multiple changes in the temporal and spatial expression of key transcription factors that specify the skeletogenic cell fate in the ancestral dGRN. For instance, *alx1* and *delta*, which encode two early mesodermal dGRN components (Fig.4a), in the ancestral state are expressed exclusively in the skeletogenic cell population at the blastula stage[34,35]. In *H. erythrogramma*, transcription of *alx1* is delayed, and the role of *delta* in mesodermal fate specification appears to be delayed, concordant with a lack of a distinct expression pattern in blastula single-cell transcriptomes (Fig.3b). Furthermore, a composite of skeletogenic cell marker gene expression is localized in a discrete cell population in the ancestral state but is not detected in the *H. erythrogramma* embryo (Fig.3b). These expression differences suggest evolutionary changes in the roles of key regulators of the skeletogenic cell fate.

177 **The ancient double negative gate within the dGRN is lost in *H. erythrogramma***

178 To investigate how these roles might differ in the derived developmental mode, we first
179 experimentally perturbed the function of Alx1, which is both necessary and sufficient for
180 skeletogenic cell fate specification in the ancestral state[34] (Fig.4a). Knocking down Alx1 protein
181 with a translation-blocking morpholino antisense oligonucleotide (MASO) in *H. erythrogramma*
182 eliminates both larval and adult skeleton (Fig.4b), phenocopying the results of prior experiments
183 in sea urchins representing the ancestral condition[34]. This concordance suggests that the function
184 of Alx1 in skeletogenic cell fate specification is conserved. Zygotic transcription of *alx1*[17] and
185 skeletogenesis[36] are both markedly delayed in *H. erythrogramma* relative to the ancestral state,
186 but this shift in timing does not by itself indicate a substantive change to the organization of the
187 dGRN.

188 We next examined HesC, a transcription factor that acts even earlier in the dGRN,
189 repressing transcription of *alx1* outside of the skeletogenic cell lineage (Fig.4a). Experimentally
190 eliminating HesC protein in the ancestral dGRN produces a dramatic phenotype, with most cells
191 differentiating as skeletogenic because in the absence of *hesC* repression, *alx1* is broadly
192 transcribed[37]. In *H. erythrogramma*, however, we found that embryos develop normally
193 following HesC knock-down (Fig.4b), suggesting that it no longer acts as a repressor of *alx1*
194 transcription. This interpretation is consistent with restricted spatial expression of *hesC* in *H.*
195 *erythrogramma* (Figure S1) that would seem to preclude a broad repressive function for HesC
196 outside of the skeletogenic lineage. Altered expression of *hesC* and lack of an overt knock-down
197 phenotype hint at a more profound evolution change within the dGRN.

198 We therefore turned to Pmar1, another transcriptional repressor that interacts with *hesC*
199 to form a double-negative logic gate within the dGRN[37]: throughout most of the embryo HesC

directly represses transcription of *alx1* and other genes encoding positive regulators of the skeletogenic cell fate, permitting differentiation of other cell types; in the vegetal-most cells of the embryo, however, *pmar1* is transiently expressed beginning the 16-cell stage where it represses *hesC*, allowing *alx1* transcription and thus specification of the skeletogenic cell fate[38] (Fig.4a).

Pmar1 is encoded by a cluster of tandem genes in sea urchins[39]. We identified 10 and 20 closely linked *pmar1* orthologs in *L. variegatus* and *H. tuberculata* respectively (Table S1). The homeodomain, nuclear localization signal, and two EH1 protein:protein interaction domains are typically well conserved, although a few likely pseudogenes are present in each species (Fig.4c, Figure S2). In *H. erythrogramma* we identified 11 *pmar1* orthologs (Table S1). Surprisingly, all of these copies contain numerous substitutions, deletions, and/or frameshifts, in many cases altering over half of the residues within the homeodomain (Fig.4c). In contrast, likely functional orthologs in the other two species differ by 0-3 amino acids out of 60 within the homeodomain. Furthermore, pairwise similarity between *pmar1* orthologs within a species averages greater than 88% for the entire peptide and 93% for the homeodomain in the ancestral state, while *H. erythrogramma* averages just 71.0 % and 45.3%, respectively (Fig.4d). These sequence comparisons indicate that the integrity of the *pmar1* gene family has dramatically decayed in *H. erythrogramma,* raising the question whether these genes with a crucial role in early embryonic patterning have maintained their function in the derived developmental mode.

Previous studies demonstrate that microinjecting *pmar1* mRNA into eggs produces a dramatic phenotype, with the resulting widespread overexpression of Pmar1 protein converting most of the embryo to skeletogenic cells[37,40]. Here, we utilized this assay to test the repressive function of specific *pmar1* orthologs. We separately microinjected into *L. variegatus* embryos

mRNA encoding one *pmar1* ortholog from *L. variegatus* and two from *H. tuberculata*. As expected, these treatments replicated the published phenotype, inducing extensive conversion to the skeletogenic cell fate, as confirmed by widespread expression of the larval spicule matrix protein MSP130 (Fig.4e). We then separately tested the three most intact orthologs of *pmar1* from *H. erythrogramma*. At the same and higher concentrations, none was able to produce the specific or any other discernable phenotype. (Fig.4e, Figure S3). These results indicate that the repressive role of Pmar1 is retained in *H. tuberculata* but has been lost in *H. erythrogramma*.

Together, these perturbation experiments and sequence comparisons indicate that both components of the double negative gate near the very top of the dGRN that specifies the skeletogenic cell fate do not function in *H. erythrogramma* as they do in species with the ancestral life history. Remarkably, this excision does not abort either the specification or subsequent function of skeletogenic cells: the role of Alx1, the component of the skeletogenic subcircuit immediately following the double-negative gate, remains intact (Fig.4b), structural genes characteristic of differentiated skeletogenic cells are transcribed[17], and a simplified larval skeleton is synthesized[36].

Discussion

Prior work showed that the evolution of nonfeeding development in *Heliocidaris* was accompanied by overt changes in oogenesis, cleavage geometry, morphogenesis, and larval morphology, with extensive underlying changes in gene expression[15-18,21]. Whole genome sequence analysis presented here demonstrates that these changes are not merely superficial consequences of amplified maternal provisioning. Although we find evidence for adaptive changes within some coding regions, these are dwarfed by the sheer number and widespread distribution of apparently adaptive changes in the sequences of putative regulatory elements and

246 in the regulation of their chromatin during early development (Fig.1c-g; 2). Both types of
247 molecular change are strikingly enriched on the branch where nonfeeding development evolved
248 and are over-represented among differentially expressed genes and especially among dGRN
249 genes (Fig.1f-g; 2c-d). While the potential for weak selection and changes in chromatin
250 accessibility to influence trait evolution through changes in gene regulation is widely
251 appreciated, we are aware of few cases that illustrate the influence of both so extensively at a
252 genomic scale.

253 Focusing on transcriptional regulation that patterns the early embryo provides a test of
254 the idea that evolutionary conservation of early development is the product of intrinsic
255 constraints. We examined the earliest zygotic patterning event in the sea urchin embryo, where
256 three transcription factors interact to specify two distinct cell fates and simultaneously establish
257 the primary signaling center of the embryo. There is arguably no set of interactions within the
258 dGRN that is more fundamental to patterning the early sea urchin embryo, and they are
259 conserved among sea urchins that diverged ~225 my ago[5]. Remarkably, however, Pmar1 and
260 HesC, which interact to form a crucial double-negative logic gate[37], have lost their early
261 patterning roles in *H. erythrogramma* (Fig.4b-e). The case of *pmar1* is particularly striking, as it
262 is present as a tandem array of genes; uniquely in the genome of *H. erythrogramma*, numerous
263 structural and point mutations alter about half of the homeodomain in each of 10 the orthologs,
264 rendering their proteins nonfunctional (Figure 4c-e).

265 The magnitude and extent of modifications to the earliest regulatory interactions within
266 the sea urchin dGRN in *H. erythrogramma* demonstrate that some deeply conserved embryonic
267 patterning mechanisms remain evolvable during substantial shifts in selective regimes. More
268 broadly, conservation of gene network architecture does not necessarily imply developmental

269 constraint, but may instead reflect long-term stabilizing selection for performance relative to a
270 particular environment or life history. Abrupt shifts in natural selection provide valuable natural
271 "perturbation experiments" that can reveal in detail how evolutionary mechanisms shape
272 conservation and change in gene regulation and dGRNs organization across the tree of life.
273

274 **AUTHOR CONTRIBUTIONS**

275 LW, MB, GF, and GAW conceived and designed the study. LW, DK, and PC collected tissues
276 for genomic sequencing. YZ performed genomic DNA extraction and sequencing library
277 preparations. PLD, HG, and HZ assembled the genomes. PLD performed genome annotation,
278 alignments, and data analysis. PLD collected, prepared, and analyzed ATAC-seq libraries. AJM
279 collected, prepared, and analyzed single cell RNA-seq libraries. JSS and AE performed
280 embryonic expression and injection assays. PLD and GAW wrote the manuscript and all authors
281 contributed to manuscript revisions.

282 **COMPETING INTERESTS**

283 There are no competing interests to declare.

284 **ACKNOWLEDGEMENTS**

285 We thank the Sydney Institute of Marine Science (SIMS) for facilities, as well as the SIMS staff
286 for their assistance. This work was supported by the National Science Foundation Division of
287 Integrative Organismal Systems (award no. 1929934 to GAW). and National Science Foundation
288 Graduate Research Fellowships to HRD and AJM.

289 **MATERIALS & CORRESPONDENCE**.

290 Correspondence and material requests regarding this work should be addressed to Dr. Gregory
291 Wray at gwray@duke.edu .

REFERENCES (MAIN TEXT)

1 Erwin, D. H. & Davidson, E. H. The evolution of hierarchical gene regulatory networks. *Nat Rev Genet* **10**, 141-148, doi:10.1038/nrg2499 (2009).

2 Oliveri, P. & Davidson, E. H. Gene regulatory network controlling embryonic specification in the sea urchin. *Curr Opin Genet Dev* **14**, 351-360, doi:10.1016/j.gde.2004.06.004 (2004).

3 McClay, D. R. Evolutionary crossroads in developmental biology: sea urchins. *Development* **138**, 2639-2648, doi:10.1242/dev.048967 (2011).

4 Gildor, T. & Ben-Tabou de-Leon, S. Comparative study of regulatory circuits in two sea urchin species reveals tight control of timing and high conservation of expression dynamics. *PLoS Genet* **11**, e1005435, doi:10.1371/journal.pgen.1005435 (2015).

5 Yamazaki, A. & Minokawa, T. Expression patterns of mesenchyme specification genes in two distantly related echinoids, Glyptocidaris crenularis and Echinocardium cordatum. *Gene Expr Patterns* **17**, 87-97, doi:10.1016/j.gep.2015.03.003 (2015).

6 Erkenbrack, E. M., Davidson, E. H. & Peter, I. S. Conserved regulatory state expression controlled by divergent developmental gene regulatory networks in echinoids. *Development* **145**, doi:10.1242/dev.167288 (2018).

7 Hinman, V. F., Nguyen, A. T., Cameron, R. A. & Davidson, E. H. Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. *Proc Natl Acad Sci U S A* **100**, 13356-13361, doi:10.1073/pnas.2235868100 (2003).

8 Koch, N. M. *et al.* Phylogenomic analyses of echinoid diversification prompt a re-evaluation of their fossil record. *bioRxiv*, 2021.2007.2019.453013, doi:10.1101/2021.07.19.453013 (2021).

9 Raff, R. A. Constraint, flexibility, and phylogenetic history in the evolution of direct development in sea urchins. *Developmental Biology* **119**, 6-19, doi:Doi 10.1016/0012-1606(87)90201-6 (1987).

10 Wray, G. A. Parallel evolution of nonfeeding larvae in echinoids. *Syst Biol* **45**, 308-322, doi:Doi 10.2307/2413566 (1996).

11 McEdward, L. R. & Miner, B. G. Larval and life-cycle patterns in echinoderms. *Can J Zool* **79**, 1125-1170, doi:DOI 10.1139/cjz-79-7-1125 (2001).

12 Marshall, D. J., Krug, P. J., Kupriyanova, E. K., Byrne, M. & Emlet, R. B. The biogeography of marine invertebrate life histories. *Annu Rev Ecol Evol S* **43**, 97-+, doi:10.1146/annurev-ecolsys-102710-145004 (2012).

13 Laegdsgaard, P., Byrne, M. & Anderson, D. T. Reproduction of sympatric populations of *Heliocidaris erythrogramma* and *H. tuberculata* (Echinoidea) in New South Wales. *Mar Biol* **110**, 359-374, doi:10.1007/Bf01344355 (1991).

14 Morgan, S. G. Life and death in the plankton: larval mortality and adaptation. *Ecology of marine invertebrate larvae*, 279-321 (2020).

15 Raff, R. A. Direct-developing sea urchins and the evolutionary reorganization of early development. *Bioessays* **14**, 211-218, doi:10.1002/bies.950140403 (1992).

16 Raff, R. A. & Byrne, M. The active evolutionary lives of echinoderm larvae. *Heredity* **97**, 244-252, doi:10.1038/sj.hdy.6800866 (2006).

17 Israel, J. W. *et al.* Comparative developmental transcriptomics reveals rewiring of a highly conserved gene regulatory network during a major life history switch in the sea

urchin genus *Heliocidaris*. *PLoS Biol* **14**, e1002391, doi:10.1371/journal.pbio.1002391 (2016).

18 Wray, G. A. Extreme phenotypic divergence and the evolution of development. *Seminars in Developmental Biology* (2022). *In press*.

19 Davidson, P. L. *et al.* Chromosomal-level genome assembly of the sea urchin *Lytechinus variegatus* substantially improves functional genomic analyses. *Genome Biol Evol* **12**, 1080-1086, doi:10.1093/gbe/evaa101 (2020).

20 Murrell, B. *et al.* Gene-wide identification of episodic selection. *Mol Biol Evol* **32**, 1365-1371, doi:10.1093/molbev/msv035 (2015).

21 Wang, L. *et al.* Genetic basis for divergence in developmental gene expression in two closely related sea urchins. *Nat Ecol Evol* **4**, 831-840, doi:10.1038/s41559-020-1165-y (2020).

22 Davidson, P. L., Byrne, M. & Wray, G. A. Cis-regulatory basis for developmental evolution in sea urchins. (2022). *In prep*.

23 Wray, G. A. & Raff, R. A. Evolutionary modification of cell lineage in the direct-developing sea urchin *Heliocidaris erythrogramma*. *Developmental Biology* **132**, 458-470, doi:10.1016/0012-1606(89)90242-X (1989).

24 Wray, G. A. & Raff, R. A. Novel origins of lineage founder cells in the direct-developing sea urchin *Heliocidaris erythrogramma*. *Developmental Biology* **141**, 41-54, doi:10.1016/0012-1606(90)90100-W (1990).

25 Smith, M. S., Turner, F. R. & Raff, R. A. Nodal expression and heterochrony in the evolution of dorsal-ventral and left-right axes formation in the direct developing sea urchin *Heliocidaris erythrogramma*. *J Exp Zool Part B* **310b**, 609-622, doi:10.1002/jez.b.21233 (2008).

26 Wilson, K. A., Andrews, M. E. & Raff, R. A. Dissociation of expression patterns of homeodomain transcription factors in the evolution of developmental mode in the sea urchins Heliocidaris tuberculata and H. erythrogramma. *Evol Dev* **7**, 401-415, doi:10.1111/j.1525-142X.2005.05045.x (2005).

27 Love, A. C. & Raff, R. A. Larval ectoderm, organizational homology, and the origins of evolutionary novelty. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **306**, 18-34 (2006).

28 O'Hara, T. & Byrne, M. *Australian echinoderms: biology, ecology and evolution*. (CSIRO publishing, 2017).

29 Cameron, R. A., Hough-Evans, B. R., Britten, R. J. & Davidson, E. H. Lineage and fate of each blastomere of the eight-cell sea urchin embryo. *Genes Dev* **1**, 75-85 (1987).

30 Davidson, E. H. Lineage-specific gene expression and the regulative capacities of the sea urchin embryo: a proposed mechanism. *Development* **105**, 421-445 (1989).

31 Oliveri, P., Tu, Q. & Davidson, E. H. Global regulatory logic for specification of an embryonic cell lineage. *Proc Natl Acad Sci U S A* **105**, 5955-5962, doi:10.1073/pnas.0711220105 (2008).

32 Massri, A. J. *et al.* Developmental single-cell transcriptomics in the Lytechinus variegatus sea urchin embryo. *Development* **148**, doi:10.1242/dev.198614 (2021).

33 Davidson, E. H., Cameron, R. A. & Ransick, A. Specification of cell fate in the sea urchin embryo: summary and some proposed mechanisms. *Development* **125**, 3269-3290 (1998).

34 Ettensohn, C. A., Illies, M. R., Oliveri, P. & De Jong, D. L. Alx1, a member of the Cart1/Alx3/Alx4 subfamily of Paired-class homeodomain proteins, is an essential component of the gene network controlling skeletogenic fate specification in the sea urchin embryo. *Development* **130**, 2917-2928 (2003).

35 Croce, J. C. & McClay, D. R. Dynamics of Delta/Notch signaling on endomesoderm segregation in the sea urchin embryo. *Development* **137**, 83-91, doi:10.1242/dev.044149 (2010).

36 Emlet, R. B. Larval spicules, cilia, and symmetry as remnants of indirect development in the direct developing sea urchin *Heliocidaris erythrogramma*. *Developmental Biology* **167**, 405-415, doi: 10.1006/dbio.1995.1037 (1995).

37 Revilla-i-Domingo, R., Oliveri, P. & Davidson, E. H. A missing link in the sea urchin embryo gene regulatory network: hesC and the double-negative specification of micromeres. *Proceedings of the National Academy of Sciences* **104**, 12383-12388 (2007).

38 Oliveri, P., Davidson, E. H. & McClay, D. R. Activation of pmar1 controls specification of micromeres in the sea urchin embryo. *Dev Biol* **258**, 32-43 (2003).

39 Yamazaki, A. *et al.* pmar1/phb homeobox genes and the evolution of the double-negative gate for endomesoderm specification in echinoderms. *Development* **147**, doi:10.1242/dev.182139 (2020).

40 Wu, S. Y. & McClay, D. R. The Snail repressor is required for PMC ingression in the sea urchin embryo. *Development* **134**, 1061-1070, doi:10.1242/dev.02805 (2007).

METHODS

1. Genome Sequencing and Assembly

1.1 Tissue Collection

Heliocidaris erythrogramma (*He*) and *H. tuberculata* (*Ht*) specimens were collected near Sydney Harbor in Sydney, NSW, Australia and housed in natural sea water at the Sydney Institute of Marine Science in Mosman, NSW, AU. The interpyramidal muscle of Aristotle's lantern (the sea urchin's feeding apparatus), tube feet, and the ovarian tissue were dissected from a single female individual, flash-frozen in liquid nitrogen, and stored at -80C until DNA extraction and sequencing.

1.2 Genomic DNA Sequencing

For each species, a 3rd-generation DNA library was sequenced on a PacBio sequel II CLR platform, generating 90.01 (*He*) and 89.47 (*Ht*) Gb of data with an N50 read length of

414 17.24 (*He*) and 23.70 (*Ht*) Kb. DNA from the same individual for each species was also used to
415 construct 10x Genomics linked-reads and Hi-C libraries, which were sequenced on a BGI-SEQ
416 500 platform, generating 194.11 (*He*) and 199.11 (*Ht*) Gb and 130.85 (*He*) and 229.03 Gb (*Ht*)
417 of data, respectively. Jellyfish v2.2.6[1] and GenomeScope v1.0.0[2] were deployed to conduct a k-
418 mer based survey of genome composition using linked-read sequencing data based on 17-mer
419 frequency distribution to estimate the genome size and heterozygosity of both *He* and *Ht* (Figure
420 S4 a-b).

421 *1.3 Genome Assembly*

422 PacBio sequencing data was employed to assemble a *de novo* contig-level genome
423 assembly using Canu v1.8 (minReadLength=1200; minOverlapLength=1000)[3]. Subsequently,
424 HaploMerger2 v3.6[4] was used to create breakpoints in the contigs where potential misjoins have
425 occurred by aligning allelic contigs via Lastz v1.02.00[5]. From these fragmented contigs, the
426 longest of each allelic pair was identified and selected using Redundans v0.14a[6] resulting in a
427 near-haploid level genome assembly. The output of this pipeline was polished using Pilon v1.23[7]
428 with 10X sequencing data to improve assembly quality and accuracy at single base resolution.
429 Lastly, contigs were assembled into scaffolds by mapping Hi-C read pairs to the polished
430 assembly with HiC-Pro[8], resulting in approximately 21.95% (*He*) and 32.10% (*Ht*) valid Hi-C
431 reads pairs. Juicer v1.5[9] and 3D-DNA v180419[10] were used to correct and finalize the
432 construction of chromosome-length scaffolds for each species. See Figure S4 c-d for HiC contact
433 maps.

434 *1.4 Repeat Identification and Classification*

435 Genomic repetitive elements were identified with RepeatModeler v2.01[11] to generate
436 species-specific repeat element libraries for each . Repeat families were filtered via BLASTn

v2.3.0[12] for significant hits to gene models of the well-studied sea urchin *Strongylocentrotus purpuratus* (www.echinobase.org) to prevent unintentional masking of genic regions. Repeats were masked from the genome of each species with RepeatMasker v4.1.1 (Smit et al. 2015) using the most sensitive setting (-s) to identify the location of repetitive elements. Long-terminal repeats were also secondarily identified with LTR_Finder v1.0.7[11]. Outputs of both RepeatMasker and LTR_Finder were then input into RepeatCraft v1.0[13] under default parameters to improve repeat element annotation and identification, resulting in a final genome annotation of repetitive elements. Lastly, repeats were broadly classified into functional categories described by their mode of transposition using TEclass[14].

1.5 Gene Annotation and Prediction Strategy

Previously published paired-end RNA-seq reads from six developmental stages for each *Heliocidaris* species[15] were trimmed using Trimmomatic v0.39[16] (TruSeq3-PE.fa: 2:30:10; leading: 3; trailing: 3; slidingwindow: 4:15; minlen: 36) and properly paired reads were mapped to their respective genomes using STAR v2.7.2. For each species, these RNA-seq alignments as well as protein models of the *S. purpuratus* v5.0 genome were input into BRAKER2[17] (--etpmode). This program utilizes a number of additional software as a part of its pipeline including Augustus[18], Genmark-EP+[19], Genemark-ET[20], DIAMOND[21], and Samtools[22]. Gene models from the BRAKER output for filtered for transposable elements by aligning to a combined database of transposable element sequences from the MAKER gene annotations pipeline[23] and the Dfam v3.3 transposable element database[24] using BLAST-P[12]. Lastly, gene models were improved using the PASA pipeline[25] by supplementing pre-existing gene models with a de-novo transcriptome retrieved from ref. 15. These gene models were annotated by aligning peptide sequences to three separate databases using BLAST-P v2.3.0[12] : 1) *S.*

purpuratus v4.2 gene models; 2) UniProt KnowledgeBase SwissProt protein models[26]; 3) RefSeq invertebrate protein models with *S. purpuratus* excluded (e-value cutoff: 1e-5)[27]. The list of sea urchin gene regulatory network (GRN) genes are provided in Data S1.

2. Whole Genome Alignment

Prior to whole genome alignment, each genome was soft-masked for repetitive elements using each species repeat element library. An optimal scoring matrix for whole genome alignment between each set of species was inferred using the *lastz_D_Wrapper.pl* script of HaploMerger2 v3.6[4]. Next whole genome alignment between each species pair was performed in both directions following UCSC guidelines outlined in the *runLastzChain.sh* and *doBlastzChainNet.pl* (https://github.com/ucscGenomeBrowser/kent) to produce .psl, .lav, .chain, and finally liftOver files for each whole genome alignment. In addition, .maf files were generated for *H. erythrogramma*, *H. tuberculata*, and *L. variegatus* for each chromosome using *H. erythrogramma* as the reference genome using Multiz and TBA[28].

3. ATAC-seq

3.1 Sample Preparation

For each sea urchin species (*H. erythrogramma, H. tuberculata*, and *Lytechinus variegatus*), adult animals were induced to spawn via injection of 0.5 M KCl solution into the coelom. For each species, three unique male-female pairs were crossed to produce three biologically-independent replicates of sea urchin embryos. Each culture was reared in large glass dishes supplied with 20 mm filtered sea water (FSW) that was changed every six hours. Because these species exhibit different developmental rates, a conspicuous developmental milestone, shedding of the fertilization envelope at hatched blastula-stage, was selected to maximize developmental synchrony within cultures and across species for comparison. Once a culture

483 reached the blastula stage, live embryos were collected and processed immediately for nuclei
484 preparation and transposase treatment as a part of the ATAC-seq protocol.

485 *3.2 ATAC-seq Protocol and Sequencing*

486 ATAC sample preparation was carried out according to the Omni-ATAC-seq protocol[29].
487 For each replicate, embryos were washed once in 1 mm FSW, lysed, and 50,000 nuclei were
488 isolated for the transposition reaction as described in the Omni-ATAC-seq protocol using the
489 Illumina TDE1 enzyme and tagmentation (TD) buffer (Cat. No. 20034197 and 20034198) (San
490 Diego, CA, USA). Sequencing libraries for each replicate were generated via qPCR and
491 sequencing libraries were purified and size selected using Ampure XP Beads at a 1.8:1 bead
492 volume: library volume (Beckman Coulter, Brea, CA, USA). Library quality and transposition
493 efficiency was accessed using a Fragment Analyzer and PROSize 2.0 (Agilent). *H.*
494 *erythrogramma* and *L. variegatus* libraries were sequenced on an Illumina HiSeq 4000
495 instrument using 50 bp SE sequencing at an average of 41.9 million and 37.3 million reads per
496 sample, respectively. *H. tuberculata* libraries were sequenced on an Illumina NovaSeq 6000
497 instrument using 50 bp PE sequencing (only SE were used for data analysis) at an average of
498 31.4 million reads per sample.

499 *3.3 ATAC-seq Data Analysis*

500 Raw ATAC-seq reads were trimmed for quality and sequencing adapters using cutadapt[30]
501 v2.3 with the following parameters: -a CTGTCTCTTATACACATCT -q 20 --trim-n -m 40.
502 Trimmed reads were then aligned to each species' respective genome using stampy[31] v.1.0.28
503 using the "--sensitive" set of parameters. ATAC-seq alignments were filtered for mitochondrial
504 sequences and required an alignment quality score of at least 5 using samtools v1.9[22].

505 In this study, we aimed to compare evolution of orthologous non-coding sites. To
506 accomplish this, we performed a series of liftOvers[32] to convert ATAC-seq alignments between
507 genomic coordinates of each sea urchin species (see previous section for esription of genome
508 alignments). We took an iterative, reciprocal liftOver strategy described below to minimize
509 possible reference bias associated with converting between genome assemblies: 1) *H.*
510 *erythrogramma*: *He* → *Lv* → *He*; 2) *H. tuberculata*: *Ht* → *Lv* → *He*; 3) *L. variegatus*: *Lv* → *He*
511 → *Lv* → *He*. After filtering and coordinate conversion, all ATAC-seq alignments were
512 referenced to the *H. erythrogramma* genome with average of 5.9 million alignments per sample
513 to orthologous genomic loci.
514 Following filtering and coordinate conversion, peaks were called from these alignments
515 using the MACS2 v2.1.2[33] callpeak function (parameters: –nomodel, –keep-dup=auto, –shift
516 100, –extsize 200) for each species separately. Peak coordinates were merged using the
517 bedtools[34] v2.25 *merge* function requiring a peak overlap of at least 200 bp to be merged into a
518 single peak. Lastly, for each sample, accessibility of each peak was measured with the bedtools[34]
519 v2.25 *multiBamCov* function.
520 *3.4 Tests for positive selection within OCRs*
521 In order to test for evidence of positive selection, a neutral genomic reference across all
522 species was assembled. To do this, the genome was first masked repetitive elements, coding
523 sequence, untranslated genic sequence, non-coding RNAs (including microRNAs, ribosomal
524 RNAs, small nuclear RNAs, and transfer RNAs) and ATAC-seq open chromatin regions (OCRs)
525 (see below) in the genome. The remaining, putatively neutrally evolving genome was then
526 divided into 300 bp windows, orthologous regions retrieved from each species' genome, and
527 filtered using the *filtering.py* and *pruning.py* scripts of the "adaptiphy"[35] program

(https://github.com/wodanaz/adaptiPhy). Next, branch lengths of each of these neutral sites was estimated using phyloFit[36] (--subst-mod HKY85), highly-conserved sites removed (Figure S5), relative branch lengths calculated, and sites falling within the middle 50% of relative branch lengths in the *H. erythrogramma* genome were selected as the neutral reference (88,004 sites; Data S2).

To measure branch-specific signatures of positive selection in the non-coding genome, the "adaptiPhy"[35] pipeline (https://github.com/wodanaz/adaptiPhy) for global tests of natural selection was followed. First, orthologous sequences for non-coding sites of interest were selected 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 bases. These trimmed and filtered alignments serve as "query" sequences of tests for selection. To generate a neutral reference for comparison, ten neutral sites were randomly selected (see above) and concatenated into a single neutral reference sequence. In addition, for each OCR, tests for positive selection were repeated 10 times against a unique putatively neutral reference. For each query site replicate, substitution rates of both the query and randomly concatenated neutral reference were estimated using phyloFit[36], 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[35] pipeline using HyPhy[37]. P-values and substitution rates for all query and neutral sites were then imported in R v4.0.2 for analysis (Data S3).

3.5 ATAC-seq Peak Filtering

After accessibility and rates of 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 *Section 3.4*) for accurate estimations of selection; 2) for each species, a local composition complexity (LCC)[38] value of 1.9 or more was required for the OCR 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 2 (of the 9) samples to remove OCRs with extremely low accessibility; 4) the midpoint of the OCR must lie within 25kb (in either direction) of the translational start site of a gene model; 5) the gene nearest to an OCR must be the same gene in each of the species' genomes —in other words, for each OCR and its nearest gene in the *H. erythrogramma*, the orthologous region in the *H. tuberculata* and *L. variegatus* genome must also be closest to a gene model that is orthologous (determined by annotation) to the same gene in the *H. erythrogramma* genome. 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 method resulted in a final set 27,322 high-confidence OCRs for cross-species analysis (Data S4).

3.6 ATAC-seq Statistical Analysis

Raw counts of the filtered OCRs were loaded into DESeq2[39] v1.30 to calculate differential accessibility between sample groups. For life history strategy comparisons, *H. tuberculata* and *L. variegatus* were treated as single group. Differentially accessible sites were

573 classified as having a 2-fold accessibility difference between sample groups and supported by an
574 FDR of 10%. Significant levels of positive selection were classified as having a median zeta
575 value greater than 1.5 and supported by a median false-discovery rate less than 10% across 10
576 replicates for each query site. Branch-specific evidence of positive selection met these criteria for
577 one species, but failed to meet these criteria in the other, as evidenced by a zeta score < 1.5.

578 **4. Coding Selection Analyses**

579 Single-copy orthologs between *H. erythrogramma*, *H. tuberculata*, *L. variegatus*, and an
580 unpublished of gene models of *Echinometra lucunter* were identified using OrthoFinder[40].
581 Evidence of episodic positive selection was queried on both the *H. erythrogramma* branch and
582 *H. tuberculata* branch under default parameters using BUSTED[41], by specifying either branch as
583 the "foreground" branch. P-values from these analyses are available in Data S5. Genes with
584 significant evidence of episodic positive selection were supported by a p-value <= 10% by the
585 likelihood ratio test.

586 **5. Bulk RNA-sequencing Analysis**

587 Raw RNA-seq reads from blastula stage embryos of *He*, *Ht*, and *Lv* were retrieved from
588 ref [15], trimmed and filtered for low quality bases and reads with Trimmomatic[16], and aligned to
589 each species respective genomes and gene models with STAR[42]. From these alignments, mRNA
590 expression was estimated with Salmon[43] and loaded to R for statistical analysis. Read counts for
591 summed to each gene's best match to the *S. purpuratus* v4.2 gene models to generate a common
592 reference for expression comparisons between species as described in ref [15]. Differentially
593 expressed genes between life histories were called as having a fold-change in expression > 2 and
594 supported by an FDR of 10% or less between *He* and both planktotrophic species (in the same
595 direction), and not DE between *Ht* and *Lv* (Data S6).

6. Single-cell RNA-Sequencing

6.1 H. erythrogramma embryo culturing

Female H. erythrogramma individuals were spawned via intracoelomic injection of 0.5ml of 0.5M KCl. Unfertilized eggs were washed three times in 100um filtered natural sea water (FSW). Eggs were fertilized by 2ul of concentrated sperm in .02g Para-Amino Benzoic Acid (PABA)/100 ml FSW. Following fertilization, eggs were washed three additional times in FSW to remove residual sperm and PABA. Fertilized embryos were then cultured at 22-23 °C. At 6 hours post fertilization (hpf) embryos were sampled for microscopy and dissociation, then fixed for scRNA-seq.

6.2 Embryo Dissociation and Fixation

Once embryos developed to the appropriate stage, a portion of the co-culture was taken, and washed one time in Calcium-Free Artificial SeaWater (CFASW). After washing embryos with CFASW, 3ml of embryos were added to 7ml of dissociation buffer made (1.0M Glycine and 0.25mM EDTA, pH 8.0) at 4 °C, and gently rocked on a rocker for 10 minutes. Following incubation, embryos were gently triturated 15-20 times to increase disassociation, then 10ml ice cold 100% methanol was added, and cells were incubated for 10 minutes and on the rocker. Following incubation, cells were triturated again 15-20x times, and then add another 30ml ice cold 100% methanol. Following dissociation, cells were resuspended and fixed at a final concentration of 80% Methanol for 1 hour at 4 °C. Lastly, cells were stored at -20 °C until library preparation. °C

6.3 Rehydration of Single Cells, Library Preparation, and Sequencing

Cells were centrifuged at 50xg, supernatant was discarded, and fixed cells were washed twice and rehydrated in a Sigma 3x Saline Sodium Citrate (SSC) buffer (SKU SRE0068) before

cell count and library preparation. Single cell libraries were prepared using the 10x Genomics 3' v3 gene expression kit and the 10x Chromium platform to encapsulate single cells within droplets. Library quality was verified using the Agilent 2100 Bioanalyzer. Libraries were pooled and Duke Genomics and Computational Biology Core facility sequenced samples on two NovaSeq6000 S1 flow cells with 28 x 8 x 91 bp sequencing performed.

6.4 FastQ Generation, Genome Indexing, and Quantification of scRNA-seq

Following sequencing, we used Cellranger v3.1.0 to convert Illumina-generated BCL files to fastq files using the Cellranger "mkfastq" command. scRNA-seq data for blastula stage embryos of Lytechinus variegatus were retrieved from a published scRNA-seq developmental time course of the species[44]. We then applied the "mkref" command to index the most recent Lv3.0 genome[45] (for the *Lytechinus* data) and the *H. erythrogramma* genome assembled in this study. The "count" command was used to demultiplex and count reads mapping to the respective reference *Lv* or *He* genome. The "mat2csv" command was used to obtain CSV RNA count matrix files for each sample for further downstream analysis.

6.5 Filtering and normalization

CSV RNA count matrix files were uploaded to R and a seurat object was generated for each sample. All seurat objects were then merged to undergo uniform quality control, normalization and data exploration with all samples. The merged object was then filtered to remove low quality cells with nFeature_RNA > 200, nFeature_RNA < 5500. "SCTransform", a r egularized negative binomial regression method that stabilizes variance across samples, was then applied to the merged filtered object to perform normalization and removal of technical variation[46], while preserving biological variation (Data S7).

6.6 Dimensionality reduction, visualization, and clustering

We next performed Principal Component Analysis on the SCTransformed seurat object file of raw gene expression counts, and found the nearest neighbors using 10 PC dimensions of variable gene space. UMAP (Uniform Manifold Approximation and Projection)[47] was applied to multi-dimensional scRNA-seq data to visualize the cells in a two-dimensional space. Finally, clustering was performed using graph-based Louvain Clustering with resolution, res=0.5, resulting in 11 clusters. The 11 clusters were annotated using dGRN genes and published in situ hybridization patterns as markers.

7. *H. erythrogramma* microinjection and in-situ hybridization

7.1 MASO design and microinjection

Morpholino antisense oligos (MASOs) were constructed to target the translation start site of *alx1* (ATCAATTCGGAGTTAAGTCTCGGCA) and *hesC* (ATCCAGATGTGTTAAGCATGGTTGC) and synthesized by Gene Tools (Philomath, OR, USA). Control morpholinos included a standard negative control morpholino recommended by the manufacturer (CCTCTTACCTCAGTTACAATTTATA) and a scrambled morpholino for HesC (ATCGACATCTGTTAACCATCGTTGC). Fertilized eggs of *H. erythrogramma* were injected as described in ref [48] at a concentration of 100 µM for *alx1* and 500 µM for *hesC*, then reared at 22 °C in pasteurized, 0.22-micron filtered seawater + penicillin (100 unit/mL) and streptomycin sulfate (0.1 mg/mL) (Sigma P4333A).

7.2 Fixation and ISH

H. erythrogramma embryos were fixed for in-situ hybridization (ISH) for ~16 hrs overnight at 4 °C in 4% paraformaldehyde (Sigma 158127) + 20 mM EPPS (Sigma E1894) in FSW, washed 3x in FSW, and dehydrated step-wise into 100% MeOH and stored at -20 °C. The full-length *HesC* coding sequence was synthesized in vitro by GenScript (Piscataway, NJ, USA)

and subcloned (NCBI insert number: MK749159), and used as template to make antisense RNA probes for ISH. ISH of *H. erythrogramma* was performed according to previously published methods[49]. Hybridizations were carried out at 65 °C and stringency washed at 0.1% SSC.

7.3 Imaging

Fixed *H. erythrogramma* embryos were washed with 100% EtOH, cleared and mounted in 2:1 (v/v) benzyl benzoate: benzyl alcohol (BB:BA). DIC or polarized light (PL) micrographs were taken on Olympus BX60 upright microscope with an Olympus DP73 camera. ISH images were taken on a Zeiss Upright AxioImager with a Zeiss MRm camera using ZEN Pro 2012 software.

8. Pmar1 mRNA overexpression assays

8.1 mRNA Synthesis

Sequences of *pmar1* orthologs were retrieved from each species respective genome annotations (*Lv*[44], *He* and *Ht*, this study) (Table S2). One *Lv* ortholog (LVA_25833.t1) was selected for overexpression assays as it represents the ortholog tested in previous overexpression assays[47], while two *Ht* (HTU_11636.t1 and HTU_11625.t1) and three *He* (HER_770.t1, HER_761.t1, and HER_775.t1) were selected for overexpression assays as they represent orthologs with the highest identity to the species' consensus sequence, and therefore predicted as genes most likely to be functional. Construct templates for each ortholog were ordered from Twist Biosciences (San Francisco, CA, USA) and mRNA was synthesized from these constructs with a ThermoFisher MEGAshortscript T7 Transcript Kit (AM1354).

8.2 Overexpression experiments

Female *L. variegatus* individuals were spawned via intracelomic injection of 0.5M KCl, washed in FSW, and fertilized with 1 uL of concentrated sperm in 100 ml

FSW. *Lv* constructs were injected at a concentration of 250ng/µL. *Ht* constructs were injected at a concentration of 1200ng/µL, and *He* constructs were injected at 1500ng/µL. Higher concentrations of *Heliocidaris* constructs were used to reproducibly obtain the PMC conversion phenotype, and reduced sensitivity of these assays may be attributable to less optimal cross-species interactions of Pmar1 in regulating the *L. variegatus* genome. Following injection, embryos were incubated at 23 °C and imaged at 24 hpf.

8.3 Imaging and Immunostaining

At 24 hpf, live embryos from each experiment were mounted on slides and imaged using DIC microscopy. Embryos were imaged on a Zeiss Axioplan II upright microscope controlled by Zen software.

Also at 24 hpf selected embryos were fixed in 100% ice cold methanol. Immunostaining was carried out as described in ref [50] to mark expression of Msp130 protein. Blocking and incubation of the secondary antibody was increased to 1 hr. Incubation of the primary antibody was set to 48 hours. A Zeiss 880 inverted confocal Airyscan microscope controlled by Zen software was used to take Z-stack images of stained embryos.

9. Data Availability

Sequencing reads used to assemble the *Heliocidaris* genomes, genome assemblies, and annotation results will be made available via the National Center for Biotechnology Information (NCBI) and Chinese National GeneBank (CNGB) upon manuscript acceptance for publication (accession numbers pending). Gene models are available as Data S8. Sequencing reads and alignments from the functional genomics experiments including ATAC-seq and scRNA-seq will be made available on NCBI, and results from whole-genome alignments experiments will be

made available on Dryad (accession numbers pending). Result files from these functional genomic analyses are available as Supplementary Data in this submission.

REFERENCES (METHODS)

1 Marcais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764-770, doi:10.1093/bioinformatics/btr011 (2011).

2 Vurture, G. W. *et al.* GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202-2204, doi:10.1093/bioinformatics/btx153 (2017).

3 Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**, 722-736, doi:10.1101/gr.215087.116 (2017).

4 Huang, S., Kang, M. & Xu, A. HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics* **33**, 2577-2579, doi:10.1093/bioinformatics/btx220 (2017).

5 Harris, R. S. Improved pairwise Alignment of genomic DNA. The Pennsyvania State University (2007).

6 Pryszcz, L. P. & Gabaldon, T. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res* **44**, e113, doi:10.1093/nar/gkw294 (2016).

7 Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963, doi:10.1371/journal.pone.0112963 (2014).

8 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 259, doi:10.1186/s13059-015-0831-x (2015).

9 Durand, N. C. *et al.* Juicer provides a one-click system for analyzing loop-resolution hi-c experiments. *Cell Syst* **3**, 95-98, doi:10.1016/j.cels.2016.07.002 (2016).

10 Dudchenko, O. *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92-95, doi:10.1126/science.aal3327 (2017).

11 Smit, A. F. & Hubley, R. RepeatModeler Open-1.0. (2010).

12 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421, doi:10.1186/1471-2105-10-421 (2009).

13 Wong, W. Y. & Simakov, O. RepeatCraft: a meta-pipeline for repetitive element de-fragmentation and annotation. *Bioinformatics* **35**, 1051-1052, doi:10.1093/bioinformatics/bty745 (2019).

14 Abrusan, G., Grundmann, N., DeMester, L. & Makalowski, W. TEclass--a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* **25**, 1329-1330, doi:10.1093/bioinformatics/btp084 (2009).

15 Israel, J. W. *et al.* Comparative developmental transcriptomics reveals rewiring of a highly conserved gene regulatory network during a major life history switch in the sea urchin genus *Heliocidaris*. *PLoS Biol* **14**, e1002391, doi:10.1371/journal.pbio.1002391 (2016).

16 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).

17 Bruna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform* **3**, lqaa108, doi:10.1093/nargab/lqaa108 (2021).

18 Stanke, M., Schoffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62, doi:10.1186/1471-2105-7-62 (2006).

19 Bruna, T., Lomsadze, A. & Borodovsky, M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom Bioinform* **2**, lqaa026, doi:10.1093/nargab/lqaa026 (2020).

20 Lomsadze, A., Burns, P. D. & Borodovsky, M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res* **42**, e119, doi:10.1093/nar/gku557 (2014).

21 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**, 59-60, doi:10.1038/nmeth.3176 (2015).

22 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

23 Campbell, M. S., Holt, C., Moore, B. & Yandell, M. Genome Annotation and Curation Using MAKER and MAKER-P. *Curr Protoc Bioinformatics* **48**, 4 11 11-39, doi:10.1002/0471250953.bi0411s48 (2014).

24 Storer, J., Hubley, R., Rosen, J., Wheeler, T. J. & Smit, A. F. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob DNA* **12**, 2, doi:10.1186/s13100-020-00230-y (2021).

25 Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* **9**, R7, doi:10.1186/gb-2008-9-1-r7 (2008).

26 Bateman, A. Uniprot: a universal hub of protein knowledge. *Protein Sci* **28**, 32-32 (2019).

27 O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733-745, doi:10.1093/nar/gkv1189 (2016).

28 Blanchette, M. *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res* **14**, 708-715, doi:10.1101/gr.1933104 (2004).

29 Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959-962, doi:10.1038/nmeth.4396 (2017).

30 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, 10-12 (2011).

31 Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome research* **21**, 936-939, doi:10.1101/gr.111120.110 (2011).

32 Hinrichs, A. S. *et al.* The UCSC genome browser database: update 2006. *Nucleic acids research* **34**, D590-D598 (2006).

33 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137 (2008).

34 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

35 Berrio, A., Haygood, R. & Wray, G. A. Identifying branch-specific positive selection throughout the regulatory genome using an appropriate proxy neutral. *BMC Genomics* **21**, 359, doi:10.1186/s12864-020-6752-4 (2020).

36 Siepel, A. & Haussler, D. Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol Biol Evol* **21**, 468-488, doi:10.1093/molbev/msh039 (2004).

37 Pond, S. L., Frost, S. D. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676-679, doi:10.1093/bioinformatics/bti079 (2005).

38 Konopka, A. K. Sequence complexity and composition. *eLS* (2005). http://www.els.net/WileyCDA/ElsArticle/refId-a0005260.html

39 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

40 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238, doi:10.1186/s13059-019-1832-y (2019).

41 Murrell, B. *et al.* Gene-wide identification of episodic selection. *Mol Biol Evol* **32**, 1365-1371, doi:10.1093/molbev/msv035 (2015).

42 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

43 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, doi:10.1038/nmeth.4197 (2017).

44 Massri, A. J. *et al.* Developmental single-cell transcriptomics in the *Lytechinus variegatus* sea urchin embryo. *Development* **148**, doi:10.1242/dev.198614 (2021).

45 Davidson, P. L. *et al.* Chromosomal-Level Genome Assembly of the Sea Urchin Lytechinus variegatus Substantially Improves Functional Genomic Analyses. *Genome Biol Evol* **12**, 1080-1086, doi:10.1093/gbe/evaa101 (2020).

46 Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**, 296, doi:10.1186/s13059-019-1874-1 (2019).

47 Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*, doi:10.1038/nbt.4314 (2018).

48 Edgar, A., Byrne, M. & Wray, G. A. Embryo microinjection of the lecithotrophic sea urchin *Heliocidaris erythrogramma*. *J Biol Methods* **6**, e119, doi:10.14440/jbm.2019.292 (2019).

49 Koop, D. *et al.* Nodal and BMP expression during the transition to pentamery in the sea urchin Heliocidaris erythrogramma: insights into patterning the enigmatic echinoderm body plan. *BMC Dev Biol* **17**, 4, doi:10.1186/s12861-017-0145-1 (2017).

50 Rho, H. K. & McClay, D. R. The control of foxN2/3 expression in sea urchin embryos and its function in the skeletogenic gene regulatory network. *Development* **138**, 937-945, doi:10.1242/dev.058396 (2011).

FIGURES and LEGENDS

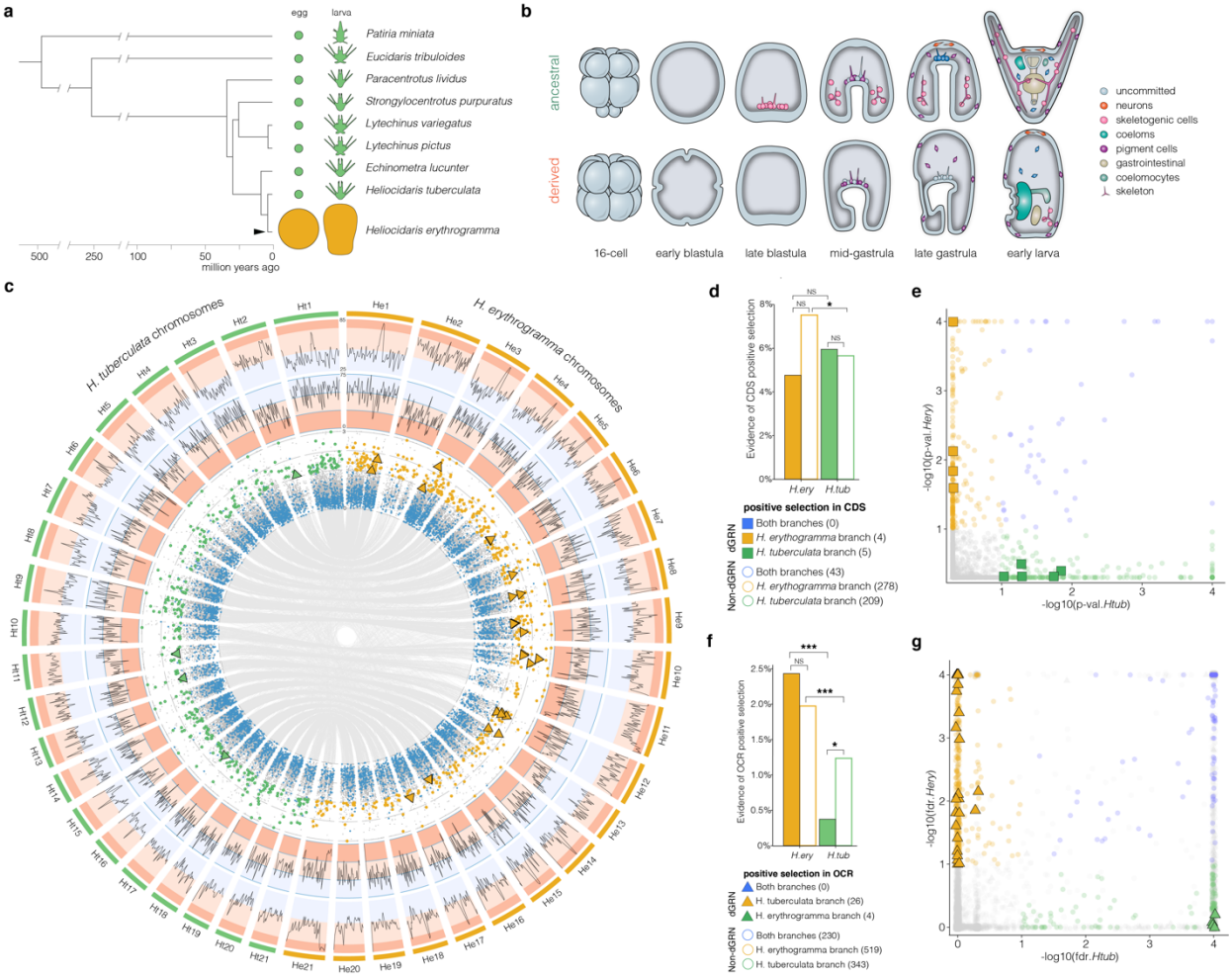
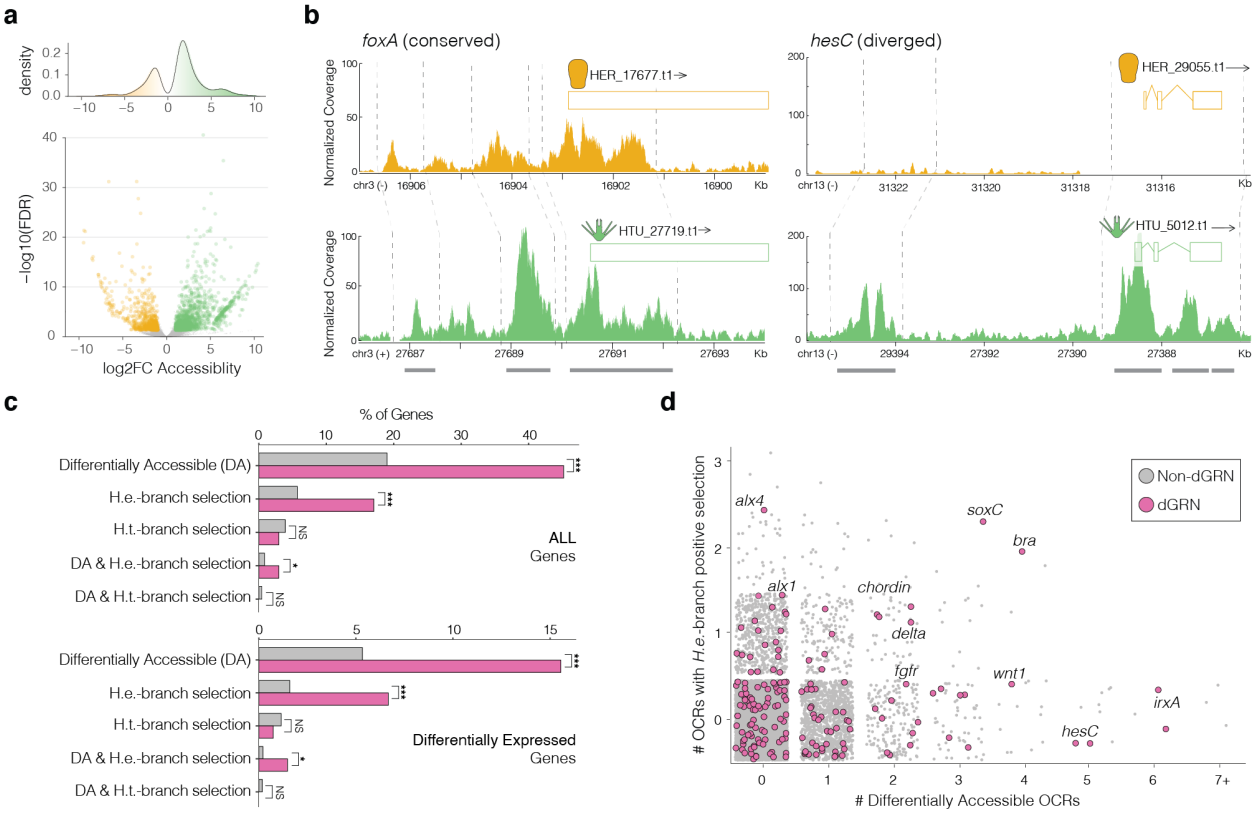


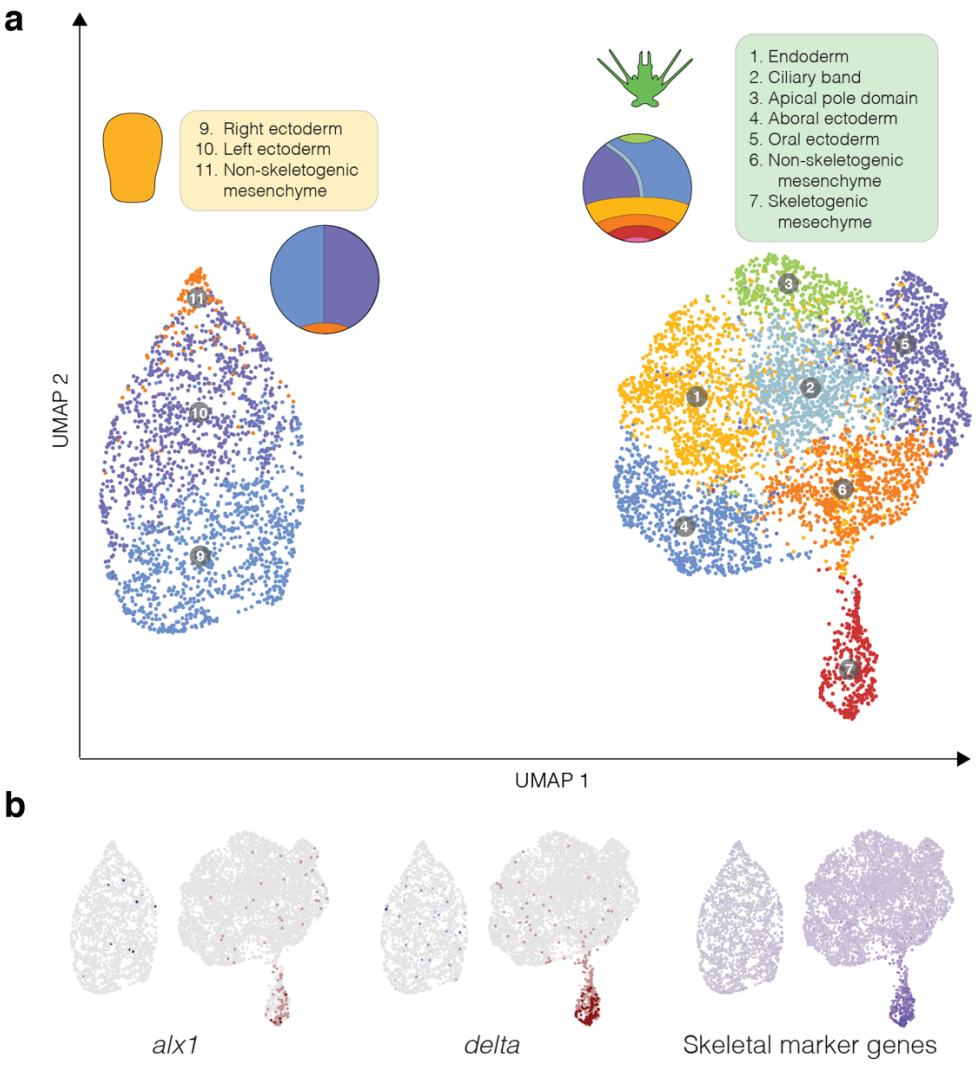
Figure 1. Evolution of life history and genomes. a, Feeding larval development (planktotrophy: green) represents the ancient and ancestral life history in sea urchins (seastar *P. miniata* represents the outgroup)[11]. Non-feeding larval development (lecithotrophy: orange) has evolved on multiple occasions, including recently within the genus *Heliocidaris* (arrowhead)[9]. **b,** Evolution of non-feeding development in *H. erythrogramma* (bottom) included dramatic modifications to otherwise broadly conserved developmental mechanisms, including changes in cleavage geometry, cell fate specification, and morphogenesis[16,18]. **c,** Chromosome-scale genome assemblies of *H. tuberculata* (green) and *H. erythrogramma* (orange). Outer ring: repetitive

element content; middle ring: gene content, inner ring: zeta values in OCRs. Colored points indicate statistically elevated zeta values (indicative of positive selection) within in a single OCR on the branch leading to *H. tuberculata* (green) or *H. erythrogramma* (orange). Blue points indicate highly conserved OCRs (top 10% of phastCon scores). Triangles denote OCRs with signature of branch-specific positive selection located near dGRN genes. **d,** Signatures of positive selection in protein-coding sequences (CDS) of 84 dGRN and 3,6999 non-dGRN single-copy orthologs. Evidence of selection is slightly enriched on the *H. erythrogramma* branch, but dGRN genes show no difference between branches. **e,** P-values of likelihood ratio test for positive selection in CDS on the branch leading to each species (color indicates significant p-values; squares indicate dGRN genes). **f,** Signatures of positive selection within single-copy OCRs near dGRN (n = 1069) and non-dGRN (n = 26,253) genes are overall much higher on the *H. erythrogramma* branch. For OCRs near dGRN genes, this difference is notably amplified: signatures of positive selection are depleted relative to non-dGRN genes on the *H. tuberculata* branch but substantially elevated on the *H. erythrogramma* branch. **g,** P-values of likelihood ratio test for positive selection in OCRs on the branch leading to each species. Fisher exact test, two-sided- *: p-value < 5e-2; ***: p-value < 5e-4. OCR: open chromatin region; dGRN: developmental gene regulatory network; CDS: coding sequence.



872

873 **Figure 2. Evolution of open chromatin landscape. a,** Density and volcano plots of
874 significantly differentially accessible orthologous OCRs between developmental life histories
875 (green = more open in *H. tuberculata* and *L. variegatus*, orange = more open in *H.*
876 *erythrogramma*). **b,** Examples of conservation (*foxA*) and change (*hesC*) in chromatin
877 accessibility landscape near dGRN genes. **c,** Relationship between chromatin, positive selection,
878 and gene expression. Percentage of all genes (top) and differentially expressed genes (bottom)
879 with at least one OCR nearby that is differentially accessible (DA), has evidence of positive
880 selection on the *H. erythrogramma* (*H.e.*) or *H. tuberculata* (*H.t.*)-branch, or is both DA and has
881 evidence of positive selection on the *H.e.* or *H.t.* branch. **d,** Number of differentially accessible
882 OCRs and OCRs with evidence of positive selection in *H. erythrogramma* for a given gene.
883 Fisher exact test, two-sided-*: p-value < 5e-2; *** p-value < 5e-4. OCR: open chromatin region;
884 dGRN: developmental gene regulatory network.



885

886 **Figure 3. Evolution of transcriptomes. a,** Single cell RNA sequencing of blastula stage

887 embryos from *H. erythrogramma* (left) and *L. variegatus* (right) with cells plotted into a single

888 UMAP space. Colors indicate inferred cell lineages based on expression of marker genes; inset

889 embryo cartoons indicate approximate spatial distribution. *H. erythrogramma* shows fewer

890 distinct transcriptional states, reflecting a broad delay in cell fate specification. **b,** Expression of

891 genes specific to skeletogenic cells: *alx1*, *delta*, and composite of 11 marker genes. The earliest

892 specification event in the ancestral dGRN, that of the skeletogenic cell lineage, takes place

893 during early cleavage[29]. Note expression of lineage-specific markers has still not occurred by

894 blastula stage in *H. erythrogramma*.

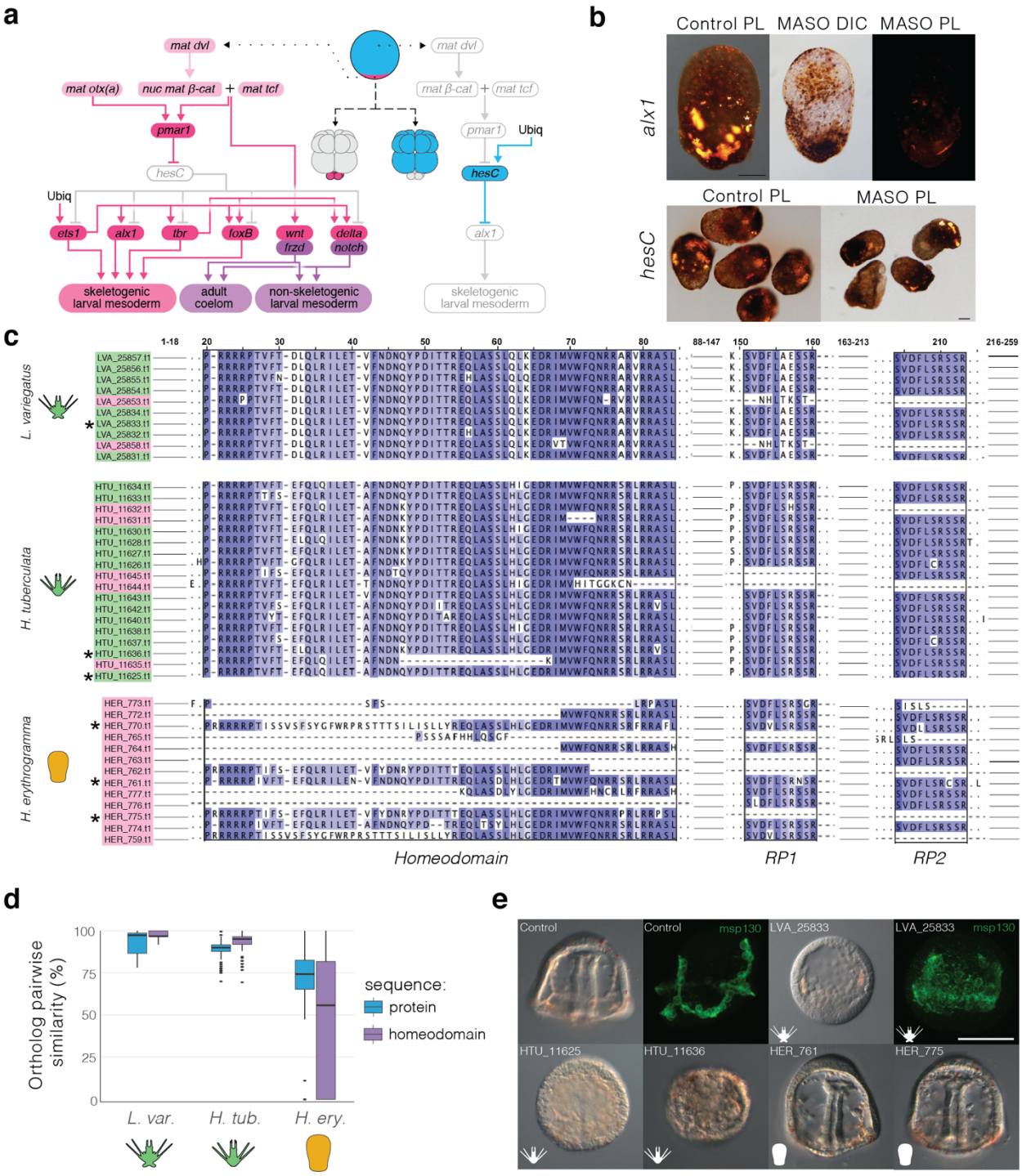


Figure 4. Evolutionary change at the "top" of a conserved developmental dGRN. a, Schematic of the ancestral dGRN that specifies skeletogenic cell fate: HesC suppresses this fate in most of the embryo (blue) but Pmar1 suppresses *hesC* in the precursors of the skeletogenic

899 cells (magenta), where Alx1 then activates a differentiation program. **b,** Control and MASO
900 knock-down of Alx1 and HesC in *H. erythrogramma* (early larva; scale bar 100μm). Polarized
901 light (PL) illuminates skeletal elements. Knockdown of *alx1* expression eliminates skeleton
902 formation in *H. erythrogramma*, as in the ancestral dGRN[34]. In contrast, HesC knockdown
903 shows no phenotype, a dramatic change from the ancestral dGRN[37]. **c,** Alignment of
904 homeodomain (DNA-binding) and RP domains (protein:protein interaction) from *pmar1*
905 orthologs. Green = likely functional copy, red = predicted non-functional copy, asterisks =
906 orthologs whose function was experimentally validated. **d,** Within-species pairwise sequence
907 similarity of *pmar1* orthologs. Note rapid sequence divergence among paralogs in *H.*
908 *erythrogramma*, and particularly within the homeodomain. Center line = median pairwise
909 similarity; dots = outliers **e,** Overexpression assays of control and *pmar1* mRNA (prism stage;
910 skeletogenic cells labeled with antibody that recognizes cell surface protein MSP130). DIC and
911 fluorescent images demonstrate that mRNA of *Pmar1* orthologs from *L. variegatus* and *H.*
912 *tuberculata* convert most of the embryo to skeletogenic cells, whereas even the most intact *H.*
913 *erythrogramma* orthologs show no such phenotype, indicating loss of function. dGRN:
914 developmental gene regulatory network; MASO: morpholino antisense oligonucleotide; DIC:
915 differential interference contrast; RP: repeated peptides.