# A Chromosome-level Genome Assembly of the Highly Heterozygous Sea Urchin *Echinometra* sp. EZ Reveals Adaptation in the Regulatory Regions of Stress Response Genes

Remi N. Ketchum[1,2,*], Phillip L. Davidson[3], Edward G. Smith[1], Gregory A. Wray[3], John A. Burt[4], Joseph F. Ryan[2], and Adam M. Reitzel[1]

[1]Department of Biological Sciences, University of North Carolina at Charlotte, Charlotte, North Carolina, USA

[2]Whitney Laboratory for Marine Bioscience, University of Florida, Marineland, Florida, USA

[3]Department of Biology, Duke University, Durham, North Carolina, USA

[4]Water Research Center & Center for Genomics and Systems Biology, New York University Abu Dhabi, Abu Dhabi, UAE

*Corresponding author: E-mail: remiketchum@gmail.com.

Accepted: 20 September 2022

## Abstract

*Echinometra* is the most widespread genus of sea urchin and has been the focus of a wide range of studies in ecology, speciation, and reproduction. However, available genetic data for this genus are generally limited to a few select loci. Here, we present a chromosome-level genome assembly based on 10x Genomics, PacBio, and Hi-C sequencing for *Echinometra* sp. EZ from the Persian/Arabian Gulf. The genome is assembled into 210 scaffolds totaling 817.8 Mb with an N50 of 39.5 Mb. From this assembly, we determined that the *E.* sp. EZ genome consists of $2n = 42$ chromosomes. BUSCO analysis showed that 95.3% of BUSCO genes were complete. Ab initio and transcript-informed gene modeling and annotation identified 29,405 genes, including a conserved Hox cluster. *E.* sp. EZ can be found in high-temperature and high-salinity environments, and we therefore compared *E.* sp. EZ gene families and transcription factors associated with environmental stress response ("defensome") with other echinoid species with similar high-quality genomic resources. While the number of defensome genes was broadly similar for all species, we identified strong signatures of positive selection in *E.* sp. EZ noncoding elements near genes involved in environmental response pathways as well as losses of transcription factors important for environmental response. These data provide key insights into the biology of *E.* sp. EZ as well as the diversification of *Echinometra* more widely and will serve as a useful tool for the community to explore questions in this taxonomic group and beyond.

**Key words:** *Echinometra* sp. EZ, sea urchin, echinoderm, Persian/Arabian Gulf, chemical defensome, positive selection.

## Introduction

The expansion of sequencing technologies, particularly long-read sequencing, has dramatically improved the assembly of genomes for all species, particularly those with high heterozygosity and/or repeat content. These advances are particularly impactful for species where no or few closely related reference genomes are available and for species with large population sizes where nucleotide diversity is typically high. Both of these factors are common for marine invertebrates. Within the phylum Echinodermata, the purple sea urchin, *Strongylocentrotus purpuratus*, was the first genome to be sequenced (Sodergren et al. 2006). This genome not only revealed a number of critical expansions of gene families and conservation of core genes in developmental regulatory networks, but also provided an essential tool for future studies of *cis*-regulation and genetic adaptation (Nègre et al. 2011; Technau and Schwaiger 2015;

## Significance

We present the first chromosome-level genome assembly for the sea urchin *Echinometra* sp. EZ from the environmentally extreme Persian/Arabian Gulf. There are currently no high-quality sea urchin genomes available for urchins that inhabit extreme environments, despite the fact that expanding the taxonomic representation of genome assemblies would help make inferences about evolutionary processes, especially where it pertains to adaptation. Here, we compared the *E.* sp. EZ genome to several other sea urchin genomes to provide insight into the evolution of key developmental and environmental stress response genes. *Echinometra* is a well-studied and widespread genus of sea urchin that plays critical ecological roles in coral reef communities, and this genome represents a crucial tool for future studies across multiple fields.

Clucas et al. 2019). Since the release of the *S. purpuratus* genome, the genomes for other echinoderms have been published (Long et al. 2016; Hall et al. 2017; Kinjo et al. 2018) and genomic resources such as SpBase (Cameron et al. 2009), Echinobase (Kudtarkar and Cameron 2017), and EchinoDB (Janies et al. 2016) have been established for data sharing and analyses. More recently, the genomes of *Lytechinus variegatus* (Davidson et al. 2020) and *Lytechinus pictus* (Warner et al. 2021) have been assembled to chromosome-level. To date, there are no sea urchin genome assemblies available from urchins that inhabit extreme environments such as the Persian/Arabian Gulf (herein the PAG), yet such efforts to expand the taxonomic representation of existing assemblies would help make inferences about evolutionary processes.

*Echinometra* is a widespread genus of pantropical sea urchins that have been the focus of many genetic (Matsuoka and Toshihiko 1991; Palumbi et al. 1997; McCartney et al. 2000; McCartney and Lessios 2004; Bronstein and Loya 2013), ecological (Khamala 1971; McClanahan and Muthiga 2007; Sangil and Guzman 2016), and reproductive studies (Metz et al. 1994; Aslan and Uehara 1997; Rahman et al. 2000, 2001; Mita et al. 2004). *Echinometra* are found in the Indo-Pacific, Caribbean, and Atlantic Ocean and they are often the most prevalent urchins on the reefs they inhabit (McClanahan and Muthiga 2007; Bronstein and Loya 2013). Although widely studied, the species-level taxonomy for this genus is incomplete and while some studies have referenced eight species, there is more likely to be at least nine species of *Echinometra* (Bronstein and Loya 2013; Ketchum et al. 2018). Bronstein and Loya (2013) were the first to describe *Echinometra* species EE and ZE (these abbreviations reference their collection site: Eilat and Zanzibar, respectively), which had been historically misidentified as *Echinometra mathaei*. Following the authors' analyses, these sea urchins were then inferred to be a single species (hence the combined name *Echinometra* sp. EZ). The misidentification for this new species also occurred in the PAG and Gulf of Oman but has since been corrected (Ketchum et al. 2020, 2021). The PAG populations are of particular interest as they inhabit the world's warmest reefs with daily mean summer temperatures regularly >35 °C and daily extremes exceeding 37 °C (Smith et al. 2017; Burt et al. 2019). Additionally, the PAG is highly saline (salinities often >45 ppt; Burt et al. 2008) and frequently experiences hypoxia in the summer (De Verneil et al. 2021). The *E.* sp. EZ populations inhabiting this sea live in an extreme environment with warming episodes that are predicted to increase in severity and frequency (Sheppard et al. 2010).

Here we present the chromosome-level assembly for *E.* sp. EZ from the PAG and the associated gene annotation set. This new assembly is a marked improvement on the previous draft genome assembled (Ketchum et al. 2020). With the genome in hand, we provide insight into the evolution of several gene families related to environmental stress responses and identify instances of strong positive selection in the putative regulatory regions for transcription factors involved in regulating molecular pathways for homeostasis.

## Results and Discussion

### Genome Assembly and Annotation

The initial *Echinometra* sp. EZ draft genome (Ketchum et al. 2020) was assembled from Illumina short (150 bp) paired-end reads, had an assembly size of 1,589 Mb, comprised 4,487,317 scaffolds, and had a BUSCO complete score of 34.8% (table 1). The improved assembly generated with 10x Genomics and PacBio long reads produced a genome assembly size of 817.5 Mb with 3,800 scaffolds and an overall BUSCO completeness score of 94.4%. We used Hi-C reads to assemble these 3,800 scaffolds into an assembly that was 817.8 Mb in length with 210 scaffolds. The longest scaffold was 65.8 Mb and the scaffold N50 length was 39.5 Mb. The contact map is indicative of the presence of 21 chromosomes in the haploid *E.* sp. EZ genome (supplementary fig. 1, Supplementary Material online). This is consistent with estimates for other *Echinometra* species based on karyotyping (Uehara et al. 2020). The assembly is highly complete with very little duplicated or missing content and has a BUSCO complete score of 95.3%. The high contiguity and BUSCO scores are in line with other recently published sea urchin genomes (Davidson et al. 2020;

Warner et al. 2021). Finally, we performed gene prediction in BRAKER v2.1.5 (Brůna et al. 2021) using *E.* sp. EZ RNA-seq data and *S. purpuratus* v5.0 protein models as inputs, and identified 29,405 genes in this assembly (Gene Model BUSCO Results: Complete 82.2%, Partial 4.6%, Missing 13.2%).

A notable challenge in assembling this genome was the high heterozygosity (supplementary fig. 2, Supplementary Material online). The estimated heterozygosity was 4.4% which is extremely high (e.g., heterozygosity of 101 drosophilids ranged from 0.00035% to 2.1% in Kim et al. 2021) but is consistent with the draft genome estimates (4.5%). These levels of heterozygosity are comparable with the *S. purpuratus* genome (4–5%) but are higher than the values for *L. variegatus* (2.9%). Prior to running PurgeHaplotigs, the percentage of core duplicated genes as predicted by BUSCO was 71.1%, indicating that regions of the genome where allelic variation exceeded CANU thresholds were assembled into multiple haplotigs rather than a single haplotype-fused representation of the haploid genome. With the application of PurgeHaplotigs, we reduced the duplication levels to 2.8%. Despite the higher heterozygosity, these duplication levels are within the ranges observed in the *Lytechinus* assemblies (0.6% and 1.36% in *L. variegatus* [Davidson et al. 2020] and *L. pictus* [Warner et al. 2021], respectively). As such, these results indicate that we have generated a high-quality single pseudo-haploid reference assembly containing sequences from both parental haplotypes.

The availability of a high-quality *E.* sp. EZ genome allows for a comparison of key developmental genes between other sequenced echinoid genomes. The Hox cluster of *E.* sp. EZ consists of 11 Hox genes as well as Hox-associated homeobox gene Evx. The 11 *E.* sp. EZ Hox genes span 736 kb and the cluster is organized in the same order and orientation as the genes in the Hox clusters of *S. purpuratus* (Cameron et al. 2006) and *L. variegatus* (Davidson et al. 2020) (fig. 1 and supplementary fig. 3, Supplementary Material online).

## Chemical Defensome

Adaptive evolution can occur rapidly in species as they respond to changing local environments or as they encounter environmental challenges during range expansion (Chen et al. 2018). Gene family expansions offer one potential mechanism for rapid adaptation to novel environments (Kondrashov 2012; Meerupati et al. 2013). The *E.* sp. EZ individual sampled for this genome is hypothesized to be a member of a population likely to have undergone rapid adaptation considering the young age of the PAG (modern coastlines formed 3,000–6,000 years ago in the wake of the marine transgression following the most recent ice age; Riegl and Purkis 2012) and the extreme thermal environment. Therefore, we hypothesized that key gene families underwent expansion events and would be evident in the *E.* sp. EZ genome when compared with four sea urchin species from other geographic regions.

We investigated a suite of gene families included in the chemical defensome; nuclear receptors, biotransformation enzymes, stress response, and transporter genes. Genomic analyses of five species of sea urchins showed that the number of chemical defensome genes ranged from 800 in *E.* sp. EZ to 944 in *Heliocidaris erythrogramma* (fig. 2). The variability in the number of genes could be a result of missing annotations, however, this is likely to be a random process so is unlikely to only effect one subfamily of genes. Each gene subfamily is represented in each species and although the number of genes in each subfamily varies across species, they are relatively consistent, with the exception of the nuclear receptors (*E.* sp. EZ had 21 compared with an average of about 45 from the other echinoid species). The nuclear receptor ROR-alpha-like and the nuclear receptor subfamily 2 group E member 1 (*nr2e1*) genes are missing from the *E.* sp. EZ genome (no hits were returned when using a local TBLASTN against the genome with default parameters). The nuclear receptor ROR-alpha-like gene performs a diverse array of biological functions which includes regulation of glucose and inflammation cytokines, and free fatty acid metabolism (Zueva et al. 2018). Nuclear receptor

**Table 1**

Comparison of Genome Assembly Benchmarks Between the Published *Echinometra* sp. EZ Draft Genome Assembly (Ketchum et al. 2020), the Current Genome Assembly Prior to Running HiRise, and the Final Chromosome-Level Genome Assembly

| | Draft Genome | Genome Assembly Prior to HiRise | Chromosome-Level Genome |
|---|---|---|---|
| Assembly size | 1,589 Mb | 817.5 Mb | 817.8 Mb |
| No. of scaffolds | 4,487,317 | 3,800 | 210 |
| N50 scaffold length | 1,006 bp | 352,130 bp | 39.5 Mb |
| Longest scaffold | 66,286 bp | 1.9 Mb | 65.8 Mb |
| BUSCO complete | 34.8% | 94.44% | 95.28% |
| BUSCO duplicated | 6.7% | 4.0% | 2.8% |
| BUSCO fragmented | 45.5% | 2.8% | 2.0% |
| BUSCO missing | 19.7% | 2.7% | 2.7% |

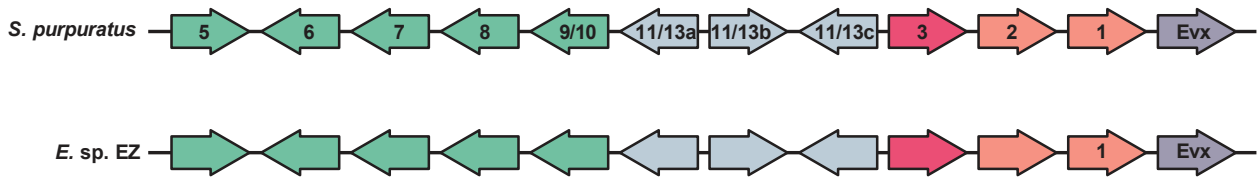


FIG. 1.—*Echinometra* sp. EZ Hox cluster organization is exactly the same as the organization of the *Strongylocentrotus purpuratus* Hox cluster in terms of membership and orientation. The names on the *S. purpuratus* Hox genes reflect the names assigned in an earlier study (Cameron et al. 2006). Except for Hox1 and Evx, we omitted names on the *E.* sp. EZ Hox genes to reflect the uncertainty of the orthology of echinoderm Hox genes relative to the vertebrate Hox genes of the same name (supplementary fig. 3, Supplementary Material online).

subfamily 2 group E member 1 is a gene involved in human retinal development and also regulates adult neural stem cell proliferation (O'Loghlen et al. 2015). Adaptive loss-of-function mutations or gene loss events are likely to occur more frequently than amino acid substitutions and have been documented in a wide variety of organisms (Wang et al. 2006; Borges et al. 2015). It is currently unclear what functions *nr2e1* performs in sea urchins or why these two nuclear receptors are missing from *E.* sp. EZ's genome. Overall, we did not observe any evidence for gene family expansions in *E.* sp. EZ that would be consistent with adaptation to a high temperature environment. However, changes in the coding regions or regulation of these genes may better account for *E.* sp. EZ's mechanism of adaptation.

## Positive Selection

We tested whether there is evidence of positive selection acting on noncoding elements near transcription factors commonly associated with maintaining homeostasis in order to assess whether putative regulatory elements of stress response genes may be evolving faster in *E.* sp. EZ. For this set of analyses we used urchins from the Family Echinometridae (*E.* sp. EZ, *H. erythrogramma*, and *H. tuberculata*) with *L. variegatus* as an outgroup. Transcription factors regulate the expression of large suites of downstream genes and thus are potentially important targets of positive selection for adaptation. Previous studies of selection for transcription factors involved in developmental gene regulatory networks have revealed particular cis-regulatory regions under positive selection, which may explain differences in developmental pathways (Erkenbrack and Davidson 2015).

We used *adaptiPhy* to identify elevated rates of sequence divergence associated with noncoding sequences 25 kb upstream and downstream of 25 key transcription factors that are broadly involved in mediating responses to environmental variation (supplementary table 1, Supplementary Material online). The divergence in these regions are compared with the background rates of sequence evolution across the phylogeny to identify evidence of branch-specific positive selection. Ten of 25 genes were discarded during filtering due to a lack of long syntenic fragments or possible duplications. *AdaptiPhy* returned 1,350 sites across the three species of sea urchins compared in this analysis. We detected evidence of positive selection in a median of 16.2% of sites in *E.* sp. EZ, 1.9% in *H. erythrogramma*, and 6.0% in *H. tuberculata*. Recent work on *Heliocidaris* (Davidson et al. 2022a) measured instances of positive selection in noncoding, open chromatin regions across the whole genome and found that when genes are categorized by function, a median of 1.18% and 1.74% of nearby noncoding sites show selection in *H. erythrogramma* and *H. tuberculata*, respectively. These values include sites near "Defensome" genes (1.90% and 1.22%, respectively) and sites near "Immune" genes (2.34% and 1.10%, respectively). While these data sets are not identical, the analyses are comparable and suggest that the enrichment of selection near these genes in *E.* sp. EZ may represent a significant signal of adaptation in *E.* sp. EZ relative to the other species. In particular, there were five genes (*hif1a*, *nfe2*, *nfkb*, *nr1h6b*, and *nr1h6c*) with nearby noncoding elements exhibiting significant signatures of positive selection predominantly in *E.* sp. EZ (fig. 3 and supplementary fig. 4, Supplementary Material online).

The first gene, *hif1a*, is a transcription factor that serves as a gene expression modulator in response to hypoxia and other environmental factors (Wenger 2002). A recent study showed that hypoxia occurs repeatedly on a southern PAG reef in the summer and can last for several hours at a time with occasional anoxic events (De Verneil et al. 2021). In the sea anemone *Exaiptasia pallida*, *hif1a* expression changes rapidly and significantly early on in the heat stress response (Cleves et al. 2020). A link between temperature and HIF-1 activity has also been documented in fishes (carp; Rissanen et al. (2006), and Atlantic salmon; Olsvik et al. (2013)) where temperature stress resulted in impaired binding activity of HIF-1. These findings point towards a critical relationship between *hif1a* and stress response. Further, our observation of positive selection in the noncoding sequences neighboring *hif1a* could be indicative of altered expression of this gene in *E.* sp. EZ, which in turn could potentially impact stress response genes under the regulatory control of this transcription factor. We also identified sites under selection for nuclear factor erythroid 2, which activates gene expression in response to oxidative and xenobiotic stress (Reitzel et al. 2008) and nuclear factor κB, a

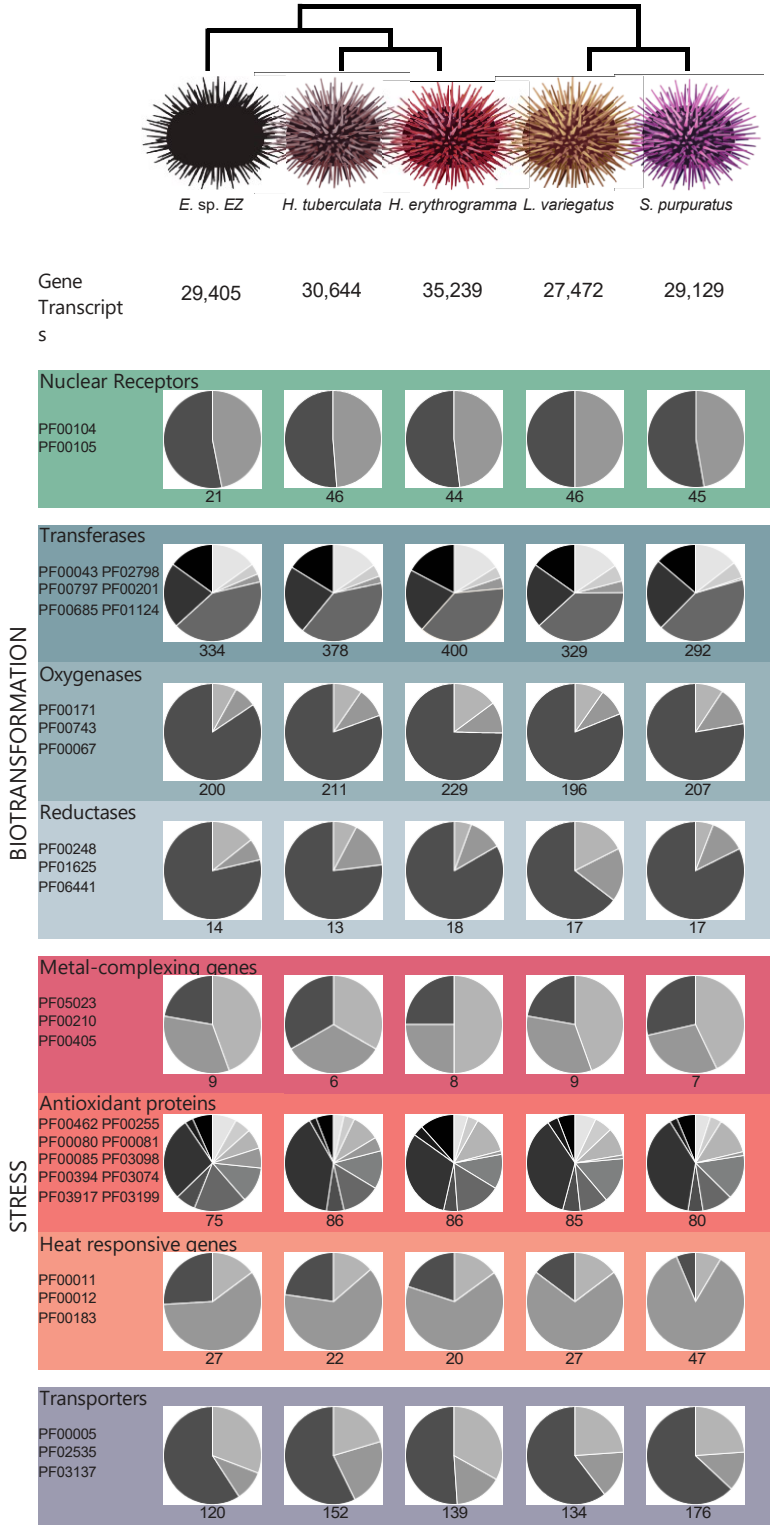


**Fig. 2.**—Chemical defensome genes in five species of sea urchin: *Echinometra* sp. EZ, *Lytechinus variegatus*, *Strongylocentrotus purpuratus*, *Heliocidaris erythrogramma*, and *Heliocidaris tuberculata*. Genes were identified by using HMMER searches with PFAM profiles. The gene families were then grouped by their role in the chemical defensome. The colors in the disk represent the PFAM ID and the size of the disk slice represents the number of genes in each respective gene family group. The total number of genes in each group family is represented underneath the disk. As the DNA-binding domain (PF00105) and the ligand-binding domain (PF00104) are likely to be picked up in most nuclear receptor genes, the values under each of the nuclear receptor disks represent the number of unique hits. PFAM groups included in this analysis were from Goldstone et al. (2006) and the grouping of PFAMs into categories followed Eide et al. (2021).

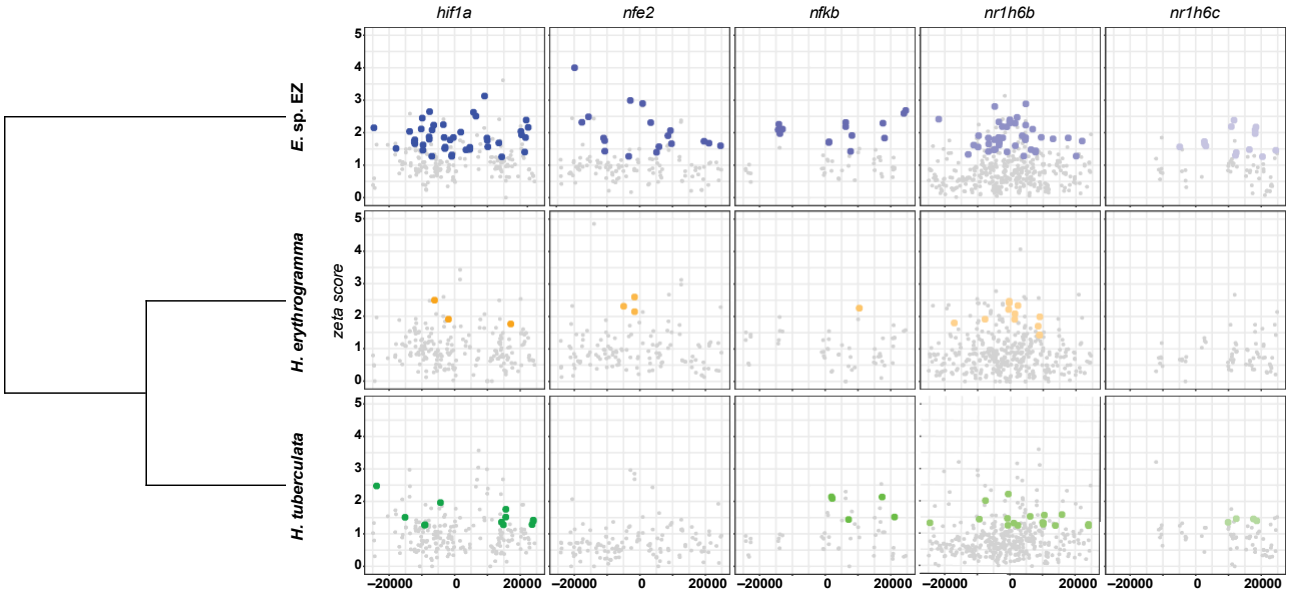


FIG. 3.—Positive selection results based on comparing noncoding elements of *Echinometra* sp. EZ to *Heliocidaris erythrogramma* and *Heliocidaris tuberculata* with *Lytechinus variegatus* as the outgroup. The plots presented include sites that show evidence of positive selection in all three sea urchin species and the title of each column is the corresponding gene name in the *S. purpuratus* genome. The *x*-axis is the distance of the site to the translation start site of the gene (negative is upstream and positive is downstream) and the *y*-axis is the zeta score (ratio of substitution rate of query to neutral substitution rate). Colored points are sites which have a zeta score ≥ 1.25 and a *P*-value ≤ 0.05.

regulator of innate immunity and linked to resistance to environmental stress (Gilmore and Wolenski 2012). The last two genes are nuclear receptors, *nr1h6b* and *nr1h6c*, whose biological function are not clearly defined but are predicted to be involved in the regulation of biotransformation enzymes and transporters based on the sequence similarity with vertebrate *nr1h* subfamily (e.g., FXR, LXR) (Goldstone et al. 2006). Although our analysis presented for this selection of transcription factors is relatively small compared with the nearly 30,000 annotated genes in the genome, we anticipate that many other loci will also reveal signatures of positive selection in future analyses of the *E.* sp. EZ genome.

## Conclusion

*Echinometra* is a widespread genus of sea urchin that plays important ecological roles in tropical benthic communities. This is the first chromosome-level genome assembly for a species from this genus and represents a crucial advance with implications across multiple fields. First, it will allow comparisons across multiple sea urchin genera to elucidate fundamental evolutionary patterns in genome evolution. Our initial analyses reported here suggest the genome size, number of genes, and gene family representation are fairly consistent across these echinoid species. Second, the completeness of the genome will facilitate ongoing research investigating the role of adaptation in these urchins, particularly with respect to identifying structural variation, positive selection, and gene arrangements along ecological gradients and contrasting habitats. We provide some examples of elevated signatures of positive selection for putative regulatory elements of orthologous transcription factors likely involved in environmental stress responses. Last, the *Echinometra* genus is a classically studied example of speciation and therefore this highly contiguous genome assembly will provide a new tool to elucidate the genomic underpinnings of speciation.

## Materials and Methods

### Tissue Collection and DNA/RNA Extraction

Samples were collected from the southern PAG which is characterized by extreme and highly variable temperatures and salinity (Vaughan et al. 2019). One *Echinometra* sp. EZ adult was collected from Dhabiya Reef (lat–long: 24.36549992, 54.10079998) in December 2017 for DNA isolation for genome assembly. Gonadal tissue was dissected from this individual and immediately placed on dry ice and sent to the Dovetail Genomics Center who performed the subsequent DNA extraction using the Qiagen Genomic-tip DNA kit. In order to generate a transcriptome, RNA was isolated from an additional *E.* sp. EZ adult that was collected from Saadiyat Reef (lat–long: 24.59899996, 54.42149998) in December 2017. Gonadal tissue was dissected from this individual and immediately placed on dry

ice. RNA was extracted using the RNAqueous Total RNA Isolation Kit.

## Sequencing and Genome Assembly

RNA was sent to DHMRI (Kannapolis, NC, USA) for library preparation and sequencing with an Illumina HiSeq2500. Total RNA was quantified using the Quant-iT RiboGreen RNA Assay Kit (Thermo Fisher) and RNA integrity assessed using an Agilent Bioanalyzer (Santa Clara, CA, USA). RNA sequencing libraries were generated using the Illumina TruSeq RNA Library Prep Kit following the manufacturer's protocol and quantitated using qPCR and fragments were visualized using an Agilent Bioanalyzer. The library was run on one flow cell for a 125 bp paired-end sequencing run.

Gonadal tissue was sent to the Dovetail Genomics Center who performed all the library construction and sequencing for the genome. The assembly and scaffolding was accomplished through a joint effort between the authors and Dovetail Genomics. Library construction, sequencing, assembly, and scaffolding are described in detail below.

### *10x Library Preparation and Sequencing*

Whole-genome sequencing libraries were prepared with 1.0 ng of genomic DNA using the Chromium Genome Library and Gel Bead Kit v.2 (10x Genomics, cat. 120262). Link-read based technology using 10x Genomics Chromium was performed according to the manufacturer's instructions with one modification. Briefly, in order to create Gel Bead-in-Emulsions (GEMs), gDNA was combined with Master Mix, a library of Genome Gel Beads, and partitioning oil on a Chromium Genome Chip. The GEMs were isothermally amplified with primers containing an Illumina Read 1 sequencing primer, a unique 16 bp 10x barcode and a 6 bp random primer sequence, and barcoded DNA fragments were recovered for Illumina library construction. The amount and fragment size of the post-GEM DNA was quantified prior to using a Bioanalyzer 2100 with an Agilent High sensitivity DNA kit (Agilent, cat. 5067-4626). The GEM amplification product was sheared on an E220 Focused Ultrasonicator (Covaris, Woburn, MA, USA) to ~350 bp (55 s at peak power = 175, duty factor = 10, and cycle/burst = 200). Next, the sheared GEMs were converted to a sequencing library following the 10x standard operating procedure and the library was quantified by qPCR with a Kapa Library Quant kit (Kapa Biosystems-Roche). Finally, the library was sequenced on a partial lane (472M reads) of the NovaSeq6000 sequencer (Illumina, San Diego, CA, USA) with paired-end 150 bp reads.

### *PacBio Library Preparation and Sequencing*

A SMRTbell library (~20 kb) for PacBio Sequel II was constructed using a SMRTbell Template Prep Kit 1.0 (PacBio, Menlo Park, CA, USA) according to the manufacturer's recommended protocol. The Sequel Binding Kit 2.0 (PacBio) was used to bind the pooled library to the polymerase and then loaded onto the PacBio Sequel using the MagBead Kit V2 (PacBio). Sequencing was performed on one PacBio Sequel SMRT cell (Instrument Control Software Version 5.0.0.6235, Primary Analysis Software Version 5.0.0.6236 and SMRT Link Version 5.0.0.6792).

### *Dovetail Hi-C Library Preparation and Sequencing*

A Dovetail Hi-C library was prepared in a similar manner as described in Lieberman-Aiden et al. (2009). For each library, chromatin was fixed with formaldehyde in the nucleus and then extracted. This was then digested with DpnII and the 5′ overhangs were filled with biotinylated nucleotides. After the free blunt ends were ligated, crosslinks were reversed and the DNA was purified from protein. Biotin that was not internal to ligated fragments was removed and DNA was then sheared to a mean fragment size of 350 bp. Sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before the PCR enrichment of each library. Finally, the libraries were sequenced on an Illumina HiSeq X to produce 200 million 2x150 bp paired-end reads.

### *Genome Assembly*

CANU v.2.0 (Koren et al. 2017) was used for genome assembly with the following parameters to generate a de novo assembly using the PacBio reads: genomeSize = 1,300, minReadLength = 3,000, corOutCoverage = 300, useGrid = False. 10x sequences were aligned to the CANU assembly using longRanger v.2.2.2 (Marks et al. 2019) and then Pilon was used to polish the CANU assembly with the 10x reads. In order to reduce the percentage of duplication (due to the high levels of heterozygosity), PurgeHaplotigs (Roach et al. 2018) was run multiple times with the percent cutoff for identifying a contig as a haplotig ranging from 35% to 55%. A cutoff of 40% yielded the most complete assembly and this genome assembly was subsequently run through the PurgeHaplotigs "clip" option that identifies and trims overlapping contig ends.

The purged genome assembly and the Dovetail Hi-C library reads were input into HiRise. HiRise is a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies (Putnam et al. 2016). The Dovetail Hi-C library sequences were aligned to the draft input assembly using the modified SNAP read mapper (http://snap.cs.berkeley.edu). HiRise analyzed the separations of Dovetail Hi-C read pairs mapped within the draft scaffolds to produce a likelihood model for genomic distance between read pairs. The model was used to identify and break

misjoins, to score potential joins, and make joins above a threshold.

Jellyfish v2.3.0 (Marçais and Kingsford 2011) and GenomeScope v1.0.0 (Vurture et al. 2017) were used to calculate genome size, heterozygosity, and repetitiveness by performing a nonlinear least-squares optimization to fit a mixture of four evenly spaced negative binomial distributions to the *k*-mer profile. The GenomeScope profile for *E.* sp. EZ was generated from linked-read sequencing data (10x Genomics) and a *k*-mer length of 21 (supplementary fig. 2, Supplementary Material online).

## Gene Prediction and Annotation

Prior to gene prediction and annotation, genomic repetitive elements were identified with RepeatModeler v1.0.11 (Smit and Hubley 2008–2015) to generate a *E.* sp. EZ-specific repeat element library. Repetitive regions were soft-masked prior to gene prediction and annotation using RepeatMasker v4.0.8 (Smit et al. 2013–2015) with the -s and -xsmall flags. Transcriptome sequences were cleaned and trimmed using Trimmomatic v0.38 (Bolger et al. 2014) with a phred quality score of 33. Leading and trailing bases with a quality score <3 were removed, a four-base wide sliding window was used to cut where the average quality per base dropped below 15, and reads that were <36 bp long were removed. BWA-mem v0.7.17 (Li 2013) was used to align the transcriptome sequences to the genome with default parameters. Aligned RNA-seq data and *S. purpuratus* v5.0 protein models (https://www.ncbi.nlm.nih.gov/assembly/GCF_000002235.5, accessed February 23, 2021) were input into the BRAKER v2.1.5 (Brůna et al. 2021) pipeline, which relies on GeneMark v4.62 (Lomsadze et al. 2005) and Augustus v3.4.0 (Stanke et al. 2006) to generate gene models.

To identify transposable elements (TEs) in the *E.* sp. EZ genome, a master TE file with 26,470 TEs was generated by combining TEs from RepeatModeler and the default set of TEs in Maker (Campbell et al. 2014). Potential TEs were inspected with BlastP v2.5.0+ (Altschul et al. 1990) against the TE master file and the *S. purpuratus* v5.0 gene models. Gene models were filtered by removing those that 1) had a hit to the TE master file but no hit to the *S. purpuratus* gene models, 2) had a highly similar hit (BLASTX *E*-value <1e−20) to a TE but a weak hit to the *S. purpuratus* gene models (BLASTX *E*-value > 1), and 3) had a match to the *S. purpuratus* gene models and were labeled as retrotransposons. Further TE analysis was performed in extensive de novo TE annotator (Ou et al. 2019) using default parameters (supplementary table 2, Supplementary Material online).

Assembled protein models were functionally annotated using BLASTP v2.5.0+ with three protein databases: *S. purpuratus* v4.2, UniProt Knowledgebase Swiss-Prot protein models v2021-03, and RefSeq invertebrate protein models with *S. purpuratus* excluded. Lastly, BUSCO v5 (Manni et al. 2021) was used to measure the completeness of the genome assembly and the protein models using the metazoan database.

## Hox Phylogeny

To maximize transparency and minimize bias, we generated and posted to GitHub (https://github.com/josephryan/SpezHox) a phylotocol (DeBiasse and Ryan 2018), which specified our planned phylogenetic analyses prior to running these analyses. The alignment, command lines, and treefiles are also available at this GitHub repository.

We used hmm2aln.pl version 0.05 (https://github.com/josephryan/hmm2aln.pl), which uses hmmsearch from HMMer version 3.3 (Potter et al. 2018) to identify target domains in protein sequences and aligns resulting sequences to the specified hidden Markov model (HMM). We used the Hox HMM (hd60.hmm) from Zwarycz et al. (2015) to search transcriptomes from the following species: *E.* sp. EZ, *S. purpuratus*, *L. variegatus*, *Patiria miniata*, *Parastichopus parvimensis*, and *Holothuria glaberrima*. To this we added the curated set of homeodomains of the HOXL subclass from HomeoDB (Zhong and Holland 2011) as well as the published Hox/ParaHox homeodomains from *Crassostrea gigas* (Paps et al. 2015) and *Capitella teleta* (Zwarycz et al. 2015). From this set, we removed the three Hox3-related genes of *Drosophila melanogaster* (zen1, zen2, and bicoid), as these have been shown to be highly derived relative to the ancestral Hox3 sequence (Stauber et al. 1999; Hughes et al. 2004) as was done in Ryan et al (2007).

As expected, our HMM approach led to the inclusion of non-HOXL homeodomains. To create a dataset consisting only of HOXL class homeodomains, we generated an initial tree using IQTREE version 1.6.12 (Nguyen et al. 2015) under the LG model with otherwise default parameters. We then used the make_subalignment script version 0.06 (https://github.com/josephryan/make_subalignment) to retain (in our alignment) only sequences that were included in the smallest clade that included all of the human HOXL homeodomains in our preliminary tree, essentially pruning all non-HOXL homeodomains. At this point we removed three *S. purpuratus* sequences (XP_011675355, XP_011682234, and XP_011682243), which were copies of three other sequences (NP_999816, NP_999774, and XP_793141, respectively) that were retained in the final alignment. We then ran a final tree using IQTREE with the LG + G4 model and otherwise default parameters on this pruned data set.

## PFAM Analysis

We conducted a comparison of gene family representation in the *E.* sp. EZ assembly with other echinoids to compare the gene annotations and identify potential expansions and contractions across species. We focused this comparison on

genes and gene families related to environmental responses. The list of genes included for these comparisons were from the chemical defensome developed from *S. purpuratus* (Goldstone et al. 2006). We performed hidden Markov model searches using HMMER (Finn et al. 2011) and PFAM profiles to compare gene family representation in five species of sea urchins: *E.* sp. EZ, *L. variegatus* (Davidson et al. 2020), *S. purpuratus* (v5.0 gene models), *H. erythrogramma*, and *H. tuberculata* (Davidson et al. 2022b).

### Positive Selection

We used *adaptiPhy* (Berrio et al. 2020), which infers regions of the genome that are targets of branch-specific positive selection, to identify specific noncoding regions of the *E.* sp. EZ genome under selection (see https://github.com/wodanaz/adaptiPhy for command lines). This method controls for branch length when comparing substitution rates of query and neutral sequences between species and therefore should not be effected by comparing two relatively closely related species to a more distantly related one. Therefore, our analyses compared *E.* sp. EZ to the *H. erythrogramma* and *H. tuberculata* genomes with *L. variegatus* used as the outgroup.

First, we curated a list of 25 transcription factors based on their roles in the defensome, general stress response, as well as metabolism, cell growth, immunity, and wound healing (supplementary table 1, Supplementary Material online). Using a whole-genome alignment between all four species, we extracted 25 kb upstream and downstream of the transcription start site for each of the 25 genes (median and mean intergenic distance for *E.* sp. EZ is 18.1 and 27.6 kb, respectively). We used the sliding window approach in bedtools to split these regions into 300 bp fragments and excluded exons, UTRs, and repetitive elements. We then identified one-to-one orthologous sequences shared across all urchins and ran *adaptiPhy* to test for selection. After filtering for gaps and alignment lengths of at least 75 bp in each species, a *P*-value was calculated based on a likelihood ratio test comparing models of branch-specific positive selection and neutrally evolving sequence (Davidson et al. 2022b), and then adjusted using the false discovery rate estimation. Finally, we calculated the zeta statistic (ratio of substitution rate of query to neutral substitution rate) to visualize the strength of selection. The branch substitution rates for the zeta scores were calculated separately in *phyloFit* (Hubisz et al. 2011) for the query and reference sequences. A zeta statistic >1.25 and supported by a *P*-value <10% for a noncoding element in one lineage indicates evidence of positive selection.

## Supplementary material

Supplementary material is available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

## Acknowledgments

This work was supported by NSF Award 1924498 to A.M.R. and J.A.B. and a NSF Graduate Research Fellowship (GRF) to R.N.K. J.A.B. was also supported by grant CGSB5 from Tamkeen. The fieldwork was carried out using the Marine Biology Core Technology Platform resources at New York University Abu Dhabi, and this support is greatly appreciated. We thank the Environmental Agency—Abu Dhabi for permits (TMBS/L/281 and 13097) and Dibba Municipality (unnumbered permit, June 2017).

## Data Availability

The data sets supporting the results of this article are available in a Dryad repository, https://doi.org/10.5061/dryad.mw6m90602, the NCBI Sequence Read Archive Database under BioProject ID PRJNA846775, and the code used to generate our findings are available at https://github.com/remiketchum/GBE_Ketchum_et_al_2022, https://github.com/josephryan/SpezHox, and https://github.com/wodanaz/adaptiPhy.

## Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Aslan LM, Uehara T. 1997. Hybridization and F1 backcrosses between two closely related tropical species of sea urchins (genus *Echinometra*) in Okinawa. Invertebr Reprod Dev. 31:319–324.

Berrio A, Haygood R, Wray GA. 2020. Identifying branch-specific positive selection throughout the regulatory genome using an appropriate proxy neutral. BMC Genomics 21:1–16.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Borges R, et al. 2015. Gene loss, adaptive evolution and the co-evolution of plumage coloration genes with opsins in birds. BMC Genomics 16:1–14.

Bronstein O, Loya Y. 2013. The taxonomy and phylogeny of *Echinometra* (Camarodonta: Echinometridae) from the Red Sea and Western Indian Ocean. PLoS One 8:e77374.

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genomics Bioinf. 3(1):lqaa108. https://doi.org/10.1093/nargab/lqaa108.

Burt J, Bartholomew A, Usseglio P. 2008. Recovery of corals a decade after a bleaching event in Dubai, United Arab Emirates. Mar Biol. 154:27–36.

Burt JA, Paparella F, Al-Mansoori N, Al-Mansoori A, Al-Jailani H. 2019. Causes and consequences of the 2017 coral bleaching event in the southern Persian/Arabian Gulf. Coral Reefs 38:567–589.

Cameron RA, et al. 2006. Unusual gene order and organization of the sea urchin hox cluster. J Exp Zool Part B Mol Dev Evol. 306:45–58.

Cameron RA, Samanta M, Yuan A, He D, Davidson E. 2009. SpBase: the sea urchin genome database and web site. Nucleic Acids Res. 37:D750–D754.

Campbell MS, Holt C, Moore B, Yandell M. 2014. Genome annotation and curation using MAKER and MAKER-P. Curr Protoc Bioinformatics. 48:4.11.11–14.11.39.

Chen Y, et al. 2018. Rapid microevolution during recent range expansion to harsh environments. BMC Evol Biol. 18:1–13.
Cleves PA, Krediet CJ, Lehnert EM, Onishi M, Pringle JR. 2020. Insights into coral bleaching under heat stress from analysis of gene expression in a sea anemone model system. Proc Natl Acad Sci U S A. 117: 28906–28917.
Clucas GV, Lou RN, Therkildsen NO, Kovach AI. 2019. Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole-genome sequencing. Evol Appl. 12:1971–1987.
Davidson PL, et al. 2020. Chromosomal-level genome assembly of the sea urchin *Lytechinus variegatus* substantially improves functional genomic analyses. Genome Biol Evol. 12:1080–1086.
Davidson PL, Byrne M, Wray GA. 2022a. Evolutionary changes in the chromatin landscape contribute to reorganization of a developmental gene network during rapid life history evolution in sea urchins. Mol Biol Evol. 39:msac172.
Davidson PL, et al. 2022b. Recent reconfiguration of an ancient developmental gene regulatory network in Heliocidaris sea urchins. bioRxiv. https://doi.org/10.1101/2022.03.03.482868.
DeBiasse MB, Ryan JF. 2018. Phylotocol: promoting transparency and overcoming bias in phylogenetics. Syst Biol. 68:672–678.
De Verneil A, Burt JA, Mitchell M, Paparella F. 2021. Summer oxygen dynamics on a southern Arabian Gulf coral reef. Front Mar Sci. 1676. https://doi.org/10.3389/fmars.2021.781428.
Eide M, et al. 2021. The chemical defensome of five model teleost fish. Sci Rep. 11:1–13.
Erkenbrack EM, Davidson EH. 2015. Evolutionary rewiring of gene regulatory network linkages at divergence of the echinoid subclasses. Proc Natl Acad Sci U S A. 112:E4075–E4084.
Finn RD, Clements J, Eddy SR. 2011. HMMER Web server: interactive sequence similarity searching. Nucleic Acids Res. 39: W29–W37.
Gilmore TD, Wolenski FS. 2012. NF-Kb: where did it come from and why? Immunol Rev. 246:14–35.
Goldstone J, et al. 2006. The chemical defensome: environmental sensing and response genes in the *Strongylocentrotus purpuratus* genome. Dev Biol. 300:366–384.
Hall MR, et al. 2017. The crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest. Nature 544: 231–234.
Hubisz MJ, Pollard KS, Siepel A. 2011. PHAST and RPHAST: phylogenetic analysis with space/time models. Brief Bioinform. 12:41–51.
Hughes CL, Liu PZ, Kaufman TC. 2004. Expression patterns of the rogue Hox genes Hox3/zen and fushi tarazu in the apterygote insect *Thermobia domestica*. Evol Dev. 6:393–401.
Janies DA, et al. 2016. EchinoDB, an application for comparative transcriptomics of deeply-sampled clades of echinoderms. BMC Bioinform. 17:1–6.
Ketchum RN, et al. 2020. Population genomic analyses of the sea urchin *Echinometra* sp. *EZ* across an extreme environmental gradient. Genome Biol Evol. 12:1819–1829.
Ketchum RN, et al. 2021. Unraveling the predictive role of temperature in the gut microbiota of the sea urchin *Echinometra* sp. EZ across spatial and temporal gradients. Mol Ecol. 30:3869–3881.
Ketchum RN, DeBiasse MB, Ryan JF, Burt JA, Reitzel AM. 2018. The complete mitochondrial genome of the sea urchin, *Echinometra* sp. *EZ*. Mitochondrial DNA Part B 3:1225–1227.
Khamala C. 1971. Ecology of *Echinometra mathaei* (Echinoidea: Echinodermata) at Diani Beach, Kenya. Mar Biol. 11:167–172.
Kim BY, et al. 2021. Highly contiguous assemblies of 101 drosophilid genomes. Elife 10:e66405.
Kinjo S, Kiyomoto M, Yamamoto T, Ikeo K, Yaguchi S. 2018. Hpbase: a genome database of a sea urchin, *Hemicentrotus pulcherrimus*. Dev Growth Differ. 60:174–182.
Kondrashov FA. 2012. Gene duplication as a mechanism of genomic adaptation to a changing environment. Proc R Soc B Biol Sci. 279:5048–5057.
Koren S, et al. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27:722–736.
Kudtarkar P, Cameron RA. 2017. Echinobase: an expanding resource for echinoderm genomic information. Database 2017:bax074.
Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.
Lieberman-Aiden E, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293.
Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 33:6494–6506.
Long KA, Nossa CW, Sewell MA, Putnam NH, Ryan JF. 2016. Low coverage sequencing of three echinoderm genomes: the brittle star *Ophionereis fasciata*, the sea star *Patiriella regularis*, and the sea cucumber *Australostichopus mollis*. GigaScience 5:20.
Manni M, Berkeley MR, Seppey M, Simao FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. arXiv preprint arXiv:2106.11799.
Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27: 764–770.
Marks P, et al. 2019. Resolving the full spectrum of human genome variation using Linked-Reads. Genome Res. 29:635–645.
Matsuoka N, Toshihiko H. 1991. Molecular evidence for the existence of four sibling species with the sea-urchin, *Echinometra mathaei* in Japanese waters and their evolutionary relationships. Zool Sci. 8: 121–133.
McCartney MA, Keller G, Lessios HA. 2000. Dispersal barriers in tropical oceans and speciation in Atlantic and eastern Pacific sea urchins of the genus *Echinometra*. Mol Ecol. 9:1391–1400.
McCartney MA, Lessios HA. 2004. Adaptive evolution of sperm bindin tracks egg incompatibility in neotropical sea urchins of the genus *Echinometra*. Mol Biol Evol. 21:732–745.
McClanahan TR, Muthiga NA. 2007. Chapter 15 Ecology of *Echinometra*. In: Lawrence JM, editor. Developments in aquaculture and fisheries science. Elsevier. p. 297–317.
Meerupati T, et al. 2013. Genomic mechanisms accounting for the adaptation to parasitism in nematode-trapping fungi. PLoS Genet. 9:e1003909.
Metz EC, Kane RE, Yanagimachi H, Palumbi SR. 1994. Fertilization between closely related sea urchins is blocked by incompatibilities during sperm-egg attachment and early stages of fusion. Biol Bull. 187:23–34.
Mita M, Uehara T, Nakamura M. 2004. Speciation in four closely related species of sea urchins (genus *Echinometra*) with special reference to the acrosome reaction. Invertebr Reprod Dev. 45:169–174.
Nègre N, et al. 2011. A cis-regulatory map of the *Drosophila* genome. Nature 471:527–531.
Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32:268–274.
O'Loghlen A, et al. 2015. The nuclear receptor NR2E1/TLX controls senescence. Oncogene 34:4069–4077.
Olsvik PA, Vikeså V, Lie KK, Hevrøy EM. 2013. Transcriptional responses to temperature and low oxygen stress in Atlantic salmon studied with next-generation sequencing technology. BMC Genomics 14:1–21.

Ou S, et al. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 20:1–18.
Palumbi SR, Grabowsky G, Duda T, Geyer L, Tachino N. 1997. Speciation and population genetic structure in tropical Pacific sea urchins. Evolution 51:1506–1517.
Paps J, Xu F, Zhang G, Holland PW. 2015. Reinforcing the egg-timer: recruitment of novel lophotrochozoa homeobox genes to early and late development in the pacific oyster. Genome Biol Evol. 7: 677–688.
Potter SC, et al. 2018. HMMER Web server: 2018 update. Nucleic Acids Res. 46:W200–W204.
Putnam NH, et al. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 26: 342–350.
Rahman MA, Uehara T, Aslan LM. 2000. Comparative viability and growth of hybrids between two sympatric species of sea urchins (Genus *Echinometra*) in Okinawa. Aquaculture 183:45–56.
Rahman MA, Uehara T, Pearse JS. 2001. Hybrids of two closely related tropical sea urchins (genus *Echinometra*): evidence against postzygotic isolating mechanisms. Biol Bull. 200:97–106.
Reitzel AM, Sullivan JC, Traylor-Knowles N, Finnerty JR. 2008. Genomic survey of candidate stress-response genes in the estuarine anemone *Nematostella vectensis*. Biol Bull. 214:233–254.
Riegl BM, Purkis SJ. 2012. Coral reefs of the Gulf: adaptation to climatic extremes in the world's Hottest sea. In: Coral reefs of the Gulf. Springer. p. 1–4.
Rissanen E, Tranberg HK, Sollid J, Nilsson GE, Nikinmaa M. 2006. Temperature regulates hypoxia-inducible factor-1 (HIF-1) in a poikilothermic vertebrate, crucian carp (*Carassius carassius*). J Exp Biol. 209:994–1003.
Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinform. 19:1–10.
Ryan JF, et al. 2007. Pre-bilaterian origins of the Hox cluster and the Hox code: evidence from the sea anemone, *Nematostella vectensis*. PLoS One 2:e153.
Sangil C, Guzman HM. 2016. Assessing the herbivore role of the sea-urchin *Echinometra viridis*: keys to determine the structure of communities in disturbed coral reefs. Mar Environ Res. 120: 202–213.
Sheppard C, et al. 2010. The Gulf: a young sea in decline. Mar Pollut Bull. 60:13–38.
Smit AFA, Hubley R. 2008–2015. RepeatModeler Open-1.0.
Smit AFA, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. Available from: http://www.repeatmasker.org.
Smith EG, Vaughan GO, Ketchum RN, McParland D, Burt JA. 2017. Symbiont community stability through severe coral bleaching in a thermally extreme lagoon. Sci Rep. 7:2428.
Sodergren E, et al. 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. Science 314:941–952.
Stanke M, et al. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34:W435–W439.
Stauber M, Jäckle H, Schmidt-Ott U. 1999. The anterior determinant bicoid of Drosophila is a derived Hox class 3 gene. Proc Natl Acad Sci U S A. 96:3786–3789.
Technau U, Schwaiger M. 2015. Recent advances in genomics and transcriptomics of cnidarians. Mar Genomics. 24:131–138.
Uehara T, Shingaki M, Taira K, Arakaki Y, Nakatomi H. 2020. Chromosome studies in eleven Okinawan species of sea urchins, with special reference to four species of the Indo-Pacific Echinometra. In: Biology of echinodermata. CRC Press. p. 119–129.
Vaughan GO, Al-Mansoori N, Burt JA. 2019. The Arabian gulf. In World seas: an environmental evaluation. Elsevier. p. 1–23.
Vurture GW, et al. 2017. Genomescope: fast reference-free genome profiling from short reads. Bioinformatics 33:2202–2204.
Wang X, Grus WE, Zhang J. 2006. Gene losses during human origins. PLoS Biol. 4:e52.
Warner JF, et al. 2021. Chromosomal-level genome assembly of the painted sea Urchin *Lytechinus pictus*: a genetically enabled model system for cell biology and embryonic development. Genome Biol Evol. 13:evab061.
Wenger RH. 2002. Cellular adaptation to hypoxia: $O_2$-sensing protein hydroxylases, hypoxia-inducible transcription factors, and $O_2$-regulated gene expression. FASEB J. 16:1151–1162.
Zhong Y-F, Holland PW. 2011. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. Evol Dev. 13:567–568.
Zueva KJ, Lumme J, Veselov AE, Kent MP, Primmer CR. 2018. Genomic signatures of parasite-driven natural selection in north European Atlantic salmon (*Salmo salar*). Mar Genomics. 39:26–38.
Zwarycz AS, Nossa CW, Putnam NH, Ryan JF. 2015. Timing and scope of genomic expansion within Annelida: evidence from homeoboxes in the genome of the earthworm *Eisenia fetida*. Genome Biol Evol. 8:271–281.

Associate editor: Rachel O'Neill