



Near-Chromosomal-Level Genome Assembly of the Sea Urchin *Echinometra lucunter*, a Model for Speciation in the Sea

Phillip L. Davidson ^{1,2,*}, Harilaos A. Lessios³, Gregory A. Wray ¹, W. Owen McMillan³, and Carlos Prada^{3,4}

¹Department of Biology, Duke University, Durham, North Carolina, USA

²Department of Biology, Indiana University, Bloomington, Indiana, USA

³Smithsonian Tropical Research Institute, Gamboa, Panama

⁴Department of Biological Sciences, University of Rhode Island, Kingston, Rhode Island, USA

*Corresponding author: E-mail: phidavid@iu.edu.

Accepted: 19 May 2023

Abstract

Echinometra lucunter, the rock-boring sea urchin, is a widely distributed echinoid and a model for ecological studies of reproduction, responses to climate change, and speciation. We present a near chromosome-level genome assembly of *E. lucunter*, including 21 scaffolds larger than 10 Mb predicted to represent each of the chromosomes of the species. The 760.4 Mb assembly includes a scaffold N50 of 30.0 Mb and BUSCO (benchmarking universal single-copy orthologue) single copy and a duplicated score of 95.8% and 1.4%, respectively. Ab-initio gene model prediction and annotation with transcriptomic data constructed 33,989 gene models composing 50.4% of the assembly, including 37,036 transcripts. Repetitive elements make up approximately 39.6% of the assembly, and unresolved gap sequences are estimated to be 0.65%. Whole genome alignment with *Echinometra* sp. *EZ* revealed high synteny and conservation between the two species, further bolstering *Echinometra* as an emerging genus for comparative genomics studies. This genome assembly represents a high-quality genomic resource for future evolutionary and developmental studies of this species and more broadly of echinoderms.

Significance

Sea urchins have been model systems for developmental biology and ecology, and more recently have emerged as a leading group of organisms for comparative genomics. *Echinometra lucunter* is an established model for studies of speciation, population dynamics, and reproduction, yet no genomic reference for this species is available. We present a near chromosome-scale genome assembly of *E. lucunter*, which will serve as a high-quality resource for future research of this species' ecology, development, and more broadly for comparative genomic work of the Echinodermata.

Key words: genome, assembly, sea urchin, *Echinometra lucunter*, echinoderm, marine speciation.

Introduction

Sea urchins have been a model to study development for more than 100 years (Harvey 1956). This is partly because of the ease with which their gametes are obtained and manipulated, but primarily because, as deuterostomes, they provide insights relevant to the development of vertebrates, including humans (Adonin et al. 2021). The assembly of the first sea urchin genome, that of *Strongylocentrotus purpuratus* (Sodergren et al. 2006), has assisted in the study of gene function, expression during development, and the understanding of gene regulatory networks underlying skeletogenic, ectoderm, and endoderm development (Davidson et al. 2002; McIntyre et al. 2013). Its publication has facilitated the annotation of genomes of other species of sea urchins (Kinjo et al. 2018; Davidson et al. 2020; Warner et al. 2021; Ketchum et al. 2022). The additions of entire genomes of closely related species have permitted the reconstruction of the genealogy of genes, which in turn has allowed testing for the effects of different kinds of selection on their evolution (Davidson et al. 2022).

The genus *Echinometra* has figured prominently in studies of speciation (Palumbi and Metz 1991; Landry et al. 2003; Palumbi and Lessios 2005; Lessios 2007, 2011) and the evolution of gamete recognition proteins (Palumbi and Metz 1991; Palumbi 1994, 1998, 2009; Metz and Palumbi 1996; Palumbi et al. 1997; McCartney and Lessios 2004; Geyer and Palumbi 2005; Zigler et al. 2005; Geyer and Lessios 2009). *Echinometra* species inhabit shallow water of all tropical and subtropical seas. Their most distinctive morphological feature is an oval, rather than circular, shape, a characteristic shared with few other regular echinoids. The genus contains five accepted species, but molecular and artificial hybridization studies have revealed the existence of five additional species in the Indo-Pacific (Kroh and Mooi 2022). A reference genome of one of these species, *Echinometra* sp. *EZ*, was recently published (Ketchum et al. 2022). In *E. sp. EZ*, an exceedingly high level of heterozygosity was estimated (4.4%), yet a high-quality genome assembly was achieved including 21 chromosome-length scaffolds with an N50 length of 39.5 Mb and BUSCO (benchmarking universal single-copy orthologue) single-copy score of 92.5%.

Echinometra lucunter is a member of the neotropical clade of the genus, which split from the Indian Ocean–West Pacific clade 3.3–6.6 million years ago (Ma) (Palumbi 1996; McCartney et al. 2000) (fig. 1A). It is found in the tropical Atlantic and in the Caribbean, where it coexists with *E. viridis*, from which it split approximately 1.5 Ma after their common ancestor was isolated from the eastern Pacific *E. vanbrunti* by the completion of the Isthmus of Panama 3 Ma (McCartney et al. 2000). It develops echinopluteus planktonic larvae, which in the laboratory settle after 19 days (Mortensen 1921). It shows signs of strong positive selection in the

evolution of its sperm gamete recognition molecule, binding (McCartney and Lessios 2004), reflecting barriers developed by its eggs toward heterospecific fertilization (Lessios and Cunningham 1990; McCartney and Lessios 2002).

A high-quality genomic reference for this species is currently lacking. Here, we present a near chromosome-scale genome assembly of *E. lucunter* (*Eluc_v1.0*), including annotations of coding and repetitive sequence, chromosome-length scaffolds, and a genomic synteny analysis with *E. sp. EZ*.

Results and Discussion

Genome Assembly

The *Eluc_v1.0* assembly is 760.4 Mb in length, including a scaffold and contig N50 of 30.0 Mb and 128.0 Kb, respectively (table 1). A combination of four sequencing strategies was implemented to generate the final assembly, including PacBio continuous long reads (CLR) to build the primary contigs, Illumina paired-end short reads for polishing and error correction, and CHICAGO linked reads followed by Dovetail HiC for scaffolding the error-corrected contigs. The assembly includes an estimated 39.6% repetitive element composition and 37.3% GC content. Unresolved gap sequences (“N”) make up approximately 0.65% of the assembly. The assembly has a BUSCO completeness score of 97.2%, including a single copy and duplication score of 95.8% and 1.4%, respectively, placing it among the most complete echinoderm genome assemblies currently available in terms of single-copy ortholog representation (database tested: metazoan odb10; 954 BUSCOs). While the full assembly is composed of 6,704 scaffolds and does not have as high contiguity as the genome assembly of the sister species *E. sp. EZ* (see above), much of the predicted primary assembly is encompassed by the largest 21 scaffolds which cumulatively make up 674.5 Mb of sequence (fig. 1B).

Gene Annotation

Ab-initio gene model prediction was carried out using the BRAKER2 (Bruna et al. 2021) genome annotation pipeline, aided by alignments of *E. lucunter* developmental transcriptomic data and the *S. purpuratus* v. 5.0 protein models for generating the training gene sets. Combined, these analyses identified 33,989 gene models (including 37,036 transcripts) across the entire genome assembly, 28,619 (84.2%) of which reside on the largest 21 scaffolds (fig. 1B), with an average length of 12.53 Kb. The 5,370 gene models on the remaining smaller scaffolds have an average gene length of 4.54 Kb, suggesting much of the coding content in *E. lucunter* is contained in the 21 chromosome-length scaffolds. The number of genes modeled in the *E. lucunter* genome is greater than the number of predicted genes in the genome of the related congener *Echinometra* sp. *EZ*, (29,405) (Ketchum et al. 2022), but within the range

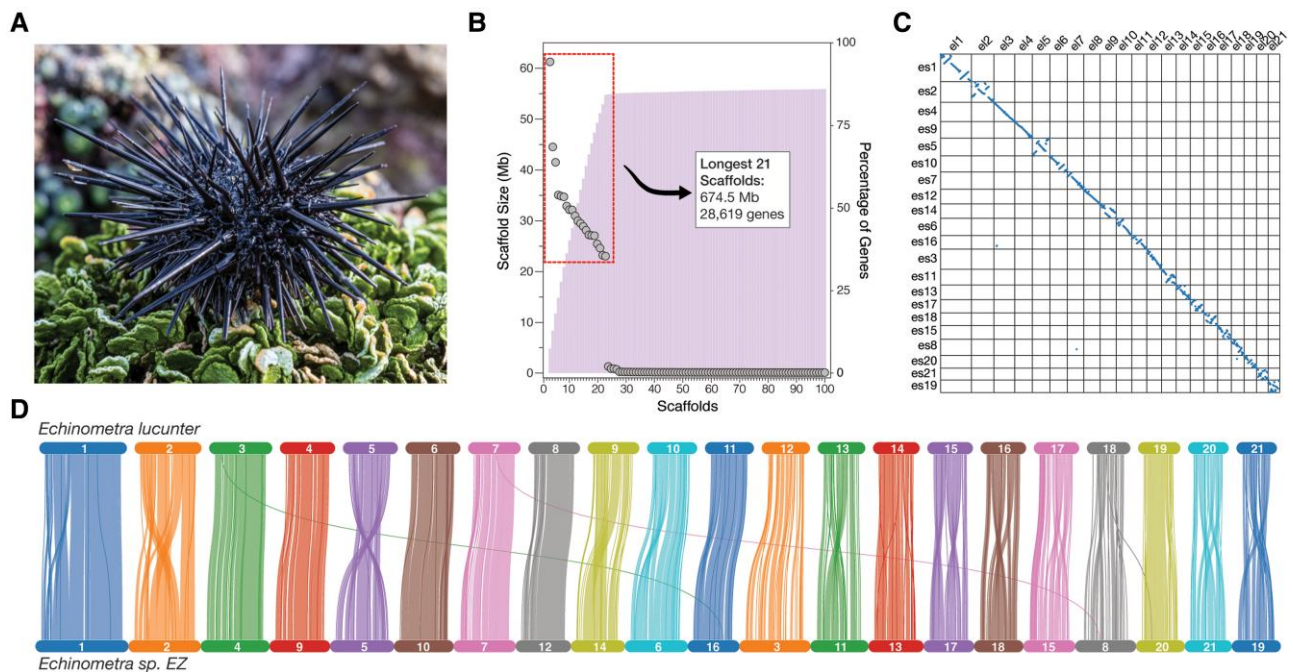


FIG. 1.—A near chromosome-level genome of *Echinometra lucunter* has been assembled. (A) An adult *Echinometra lucunter*. (B) Scaffold size distribution and cumulative percentage of gene models identified within the largest 100 scaffolds in the *Eluc_v1.0* genome assembly. The 21 largest scaffolds compose most of the primary assembly. (C) Dot plot of synteny relationship between the largest 21 scaffolds of the *Eluc_v1.0* assembly and the genome assembly of *Echinometra sp. EZ* (Ketchum et al. 2022). (D) Link plot comparing chromosome-scale scaffolds of *E. lucunter* and *E. sp. EZ* shows a high-degree of synteny between the two species.

of identified gene models in other published echinoid genomes (27,232–35,242).

Transcripts were annotated using three preexisting protein databases to putatively assign gene identification and function to each model, including the *S. purpuratus* v 4.2 protein models, the RefSeq nonredundant invertebrate protein database, and UniProt (Swiss-Prot only) metazoan proteins (supplementary Data 1, Supplementary Material online). Protein models from the annotation's set of longest isoforms have a BUSCO completion score of 95.0%, including 94.0% and 1.9% single copy and duplication scores, respectively. That we detected a BUSCO single-copy score of 95.8% across the entire genome suggests the annotation pipeline was largely successful in predicting the location and structure for an initial set of gene models.

To access how the gene models identified in the *E. lucunter* genome compare to annotation in other sea urchin genomes, orthologous gene sets ("orthogroups") were calculated for this species, three other Echinometrid species (*E. sp. EZ*, *Heliocidaris erythrogramma*, *Heliocidaris tuberculata*), and a more distantly related species, *Lytechinus variegatus*. We found 88.5% of all gene models in *E. lucunter* were assigned to an orthogroup, a percent comparable to other species (86.4–94.6%). Furthermore, 78.9% of all orthogroups contained at least one gene model from *E. lucunter*, which also falls in the range calculated in other sea urchin species (69.7–79.2%). Taken

together, these results suggest the gene set presented in this report is qualitatively comparable to other published sea urchin genomes.

Synteny Analysis

We carried out a gene-anchored synteny analysis between this novel assembly and the previously published *E. sp. EZ* genome (Ketchum et al. 2022) to assess the conservation and relatedness of the largest 21 scaffolds of the *E. lucunter* genome to a closely related species. While we identified a number of transversion events and three putative translocations (see chromosomes 3, 7, and 18), at a global scale we found a high-degree of synteny between the two genomes including a 36.8% share of collinear genes (fig. 1C and D). Interestingly, computational analysis has determined the *E. sp. EZ* genome also has 21 chromosomes along with two other sea urchin species in the family Echinometridae (*H. tuberculata* and *H. erythrogramma*) (Davidson et al. 2022). Nevertheless, future karyotyping work will still be necessary to confirm that these 21 scaffolds represent the chromosomes in the *E. lucunter* genome.

Materials and Methods

Tissue Collection and Sequencing Strategy

To assemble the nuclear genome of *E. lucunter*, we collected on March 24, 2017 muscle and gonad tissues of

Table 1
Genome Assembly Summary Statistics Following Each Assembly or Polishing Step

	Eluc_v.0.1	Eluc_v.0.2	Eluc_v.0.3	Eluc_v.0.4	Eluc_v.0.5	Eluc_v1.0
Step	Assembly	Polishing 1	Polishing 2	De-duplication	Scaffolding 1	Scaffolding 2
Method	<i>Canu</i>	<i>Arrow</i>	<i>Pilon</i>	<i>Purge_Dups</i>	<i>3d-DNA</i>	<i>3d-DNA</i>
Sequencing Strategy	PacBio CLR	PacBio CLR	Illumina PE		Chicago	HIC
Size	1414.8 Mb	1417.3 Mb	1417.2 Mb	755.5 Mb	759.1 Mb	760.4 Mb
# Scaffolds					5,854	6704
Scaffold N50					1.98 Mb	30.01 Mb
Largest Scaffold					12.40 Mb	61.26 Mb
# Scaffolds > 10 MB					3	21
# Contigs	19,645	19,645	19,645	7,563	13,078	16,653
Contig N50	116.5 Kb	116.7 Kb	116.7 Kb	167.8 Kb	132.4 Kb	128.0 Kb
Largest Contig	0.94 MB	0.94 Mb	0.94 Mb	0.94 Mb	0.83 Mb	0.83 Mb
% GC	37.3	37.3	37.3	37.3	37.3	37.3
% Ns	0.0	0.0	0.0	0.0	0.47	0.65
BUSCO (%)						
Total = 954						
<i>metazoan_odb10</i>						
Complete	95.7	96.3	96.2	95.5	96.7	97.2
Single Copy	40.1	32.1	25.9	93.5	94.9	95.8
Duplicated	55.6	64.2	70.3	2.0	1.8	1.4
Fragmented	1.5	1.0	1.2	1.7	1.2	0.4
Missing	2.8	2.7	2.6	2.8	2.1	2.4

Eluc_v1.0 (bolded) represents the final assembly presented in this report.

two adult individuals at Galeta, Panama (9°24'18.0"N 79°52'08.0"W) under permit SEX/A-21–17. The first individual was used to generate the primary contigs of the genome assembly using a hybrid Illumina-PacBio approach, and the second individual was used to produce Chicago and Dovetail HiC sequencing libraries to improve the accuracy of the initial assembly and place the contigs into chromosome-length scaffolds.

Tissues were flash-frozen in liquid nitrogen and shipped over dry ice to Duke's Center for Genomic and Computational Biology for library preparation and sequencing. High molecular weight (HMW) DNA was extracted using the Beckman Coulter Genfind v2 kit with bead-bound DNA from lysate. Two attempts were made to obtain HMW DNA. The first extraction had two peaks, one at 30 Kb and one at 60 Kb, and the second extraction had a broad peak with most of the DNA fragments at 52 Kb. The PacBio Sequel Sequencing CLR (continuous long read) platform was used to sequence four SMRT (single-molecule real-time) cells (4.01 million reads; ~39.0× coverage, N50: 12.7 Kb), and the Illumina library was sequenced on a NextSeq 500 High Output 150 bp PE (paired-end) lane (~97.8× coverage).

Genome Assembly

We used PacBio CLR reads to construct an initial primary contig assembly with *canu* v2.1.1 under default parameters (Koren et al. 2017). The resulting assembly was then subjected to two rounds of error correction: 1) with the same PacBio CLR reads using the "arrow" algorithm within *gcpp* v2.0.2 and 2) with trimmed Illumina NextSeq 500 short read paired-end data using *Pilon* v1.23 (Walker et al. 2014). After error correction, the contig assembly was 1417.2 Mb in size including 19,645 scaffolds with an N50 length of 116.7 Kb (table 1).

This version of the assembly was near twice the predicted *E. lucunter* genome size (750–800 Mb), suggesting both parental haplotypes were represented in the assembly. Therefore, we used *purge_dups* v1.2.5 (-a 75) (Guan et al. 2020) to collapse the genome into a single haploid assembly resulting in a 755.5 Mb genome assembly and BUSCO duplication score of 2.0% (table 1).

Genome Scaffolding With Proximity Ligation

To improve the contiguity of the genome assembly, we shipped flash-frozen tissue of the second individual from Galeta to Dovetail Genomics for HMW DNA library preparation (average fragment size ~50 Kb). A combination of two proximity ligation libraries, Chicago long-range linkage, and Dovetail HiC, was generated to further improve the contiguity and accuracy of the initial genome assembly. Both libraries were checked for quality by sequencing ~2 M PE 75 bp reads on a MiSeq instrument and mapped to the draft assembly described above. Each library was

sequenced on a single HiSeq 4000 Illumina lane with 150 bp PE sequencing. The target sequencing depth was 100× physical coverage (in pairs spanning 1–50Kbp) of the Chicago library and ~100 M read pairs for the Hi-C library. For both proximity ligation libraries, scaffolding was carried out via the Juicer and 3d-DNA scaffolding pipeline (Durand et al. 2016a; Dudchenko et al. 2017). The first scaffolding procedure was carried out with the CHICAGO sequencing data, and given the broad distribution and relatively short insert sizes of these libraries, no manual correction of the scaffold structure was attempted. This step resulted in an assembly with a scaffold N50 of 1.98 Mb and modestly improved BUSCO scores (table 1), likely from resolving the orientation and location of contig fragments. Lastly, this assembly was further scaffolded using the Dovetail HiC library, and three scaffold pairs were manually joined using the Juicebox software v1.11.08 (Durand et al. 2016b) (see supplementary fig. S1, Supplementary Material online for a before-and-after comparison of HiC contact maps). The resulting and final version of the assembly had a size of 760.4 Mb and scaffold N50 of 30.01 Mb (table 1).

Repeat and Gene Annotation

A library of repetitive elements was generated for the *Eluc_v1.0* genome via RepeatModeler v2.0.1 (Flynn et al. 2020). Next, this filtered repeat library was input into RepeatMasker v. 4.1.1 to identify the genomic location of each repetitive element and soft-mask the *Eluc_v1.0* assembly using the most sensitive preset parameters (-s). The resulting soft-masked genome assembly was then used as input for proceeding with gene modeling and annotation.

The BRAKER2 (Bruna et al. 2021) pipeline (-etpmode), which uses a suite of additional software including Augustus (Stanke et al. 2006), Genmark-EP+ (Bruna et al. 2020), Genemark-ET (Lomsadze et al. 2014), DIAMOND (double index alignment of next-generation sequencing data) (Buchfink et al. 2015), and SAMtools (Li et al. 2009), was used to generate an initial set of gene models. *S. purpuratus* v. 5.0 protein models and RNA-seq reads from developmental transcriptomic data of *E. lucunter* were included in this analysis to aid in predicting gene location and structure. These gene models were annotated by aligning at the peptide level to three separate protein databases: 1) the *S. purpuratus* v. 4.2 protein models; 2) the RefSeq nonredundant invertebrate protein database (O'Leary et al. 2016), and 3) the metazoan Uniprot Swiss-Prot peptide database (Bateman 2019). Due to the exploratory nature of *de-novo* gene annotation, an e-value cutoff of 1e-5 was required to assign a significant hit to each model. OrthoFinder v2.5.4 (Emms and Kelly 2019) was implemented to identify orthogroup and associated comparative statistics across sea urchin gene annotation sets using the longest isoform from each species' gene

set (supplementary Data 1, Supplementary Material online). For *E. lucunter*, this analysis included 33,989 protein models from a complete set of 37,036 transcripts. The final set of annotations and gene models are available as supplementary Data 2, Supplementary Material online.

Synteny Analysis

A gene-anchored synteny analysis was performed using MCSanX (Wang et al. 2012) under default parameters between *E. lucunter* (this assembly) and a related echinometrid *E. sp. EZ* following an all-to-all BLASTP (Camacho et al. 2009) similarity search between each species' protein models (`-max_target_seqs 5`, `-evalue 1e-10`, `-outfmt 6`). Dot and link plots were produced from this synteny analysis using SynVisio (Bandi and Gutwin 2020) by inputting a combined GFF (general feature format) file of each species' genome annotations and the collinearity output file from the MCSanX analysis (supplementary Data 3, Supplementary Material online).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

P.L.D. is supported by an NSF Postdoctoral Research Fellowship in Biology (2208912). Research costs were covered by general funds from the Smithsonian Tropical Research Institute. C.P. is supported by the USDA National Institute of Food and Agriculture (1017848) and National Science Foundation IOS (2032919).

Data Availability

Gene models, protein models, repetitive elements, and associated annotation files generated for this genome assembly are available in supplementary Data 2, Supplementary Material online and on FigShare at doi.org/10.6084/m9.figshare.22776386.v1. All genomic data generated during this project including PacBio CLR reads used for contig assembly, Illumina paired-end short reads used for error correction, HiC and CHiCAGO linked reads used for scaffolding, RNA-seq reads used for gene annotation, and the *Eluc_v1.0* genome assembly have deposited on NCBI under BioProject ID PRJNA914403.

Literature Cited

Adonin L, Drozdov A, Barlev NA. 2021. Sea urchin as a universal model for studies of gene networks. *Front Genet.* 11:627259.
Bandi V, Gutwin C. 2020. Interactive exploration of genomic conservation. Proceedings of the 46th graphics interface conference on proceedings of graphics interface 2020 (GI'20). Waterloo, Canada: Canadian Human-Computer Communications Society.

Bateman A. 2019. Uniprot: a universal hub of protein knowledge. *Protein Sci.* 28:32–32.
Bruna T, et al. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform.* 3:lqaa108.
Bruna T, Lomsadze A, Borodovsky M. 2020. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom Bioinform.* 2:lqaa026.
Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 12:59–60.
Camacho C, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
Davidson EH, et al. 2002. A genomic regulatory network for development. *Science* 295:1669–1678.
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, et al. 2022. Recent reconfiguration of an ancient developmental gene regulatory network in *Heliocidaris* sea urchins. *Nat Ecol Evol.* 6:1907–1920.
Dudchenko O, et al. 2017. De novo assembly of the *Aedes aegypti* genome using hi-C yields chromosome-length scaffolds. *Science* 356:92–95.
Durand NC, et al. 2016a. Juicer provides a one-click system for analyzing loop-resolution hi-C experiments. *Cell Syst.* 3:95–98.
Durand NC, et al. 2016b. Juicebox provides a visualization system for hi-C contact maps with unlimited zoom. *Cell Syst.* 3:99–101.
Emms DM, Kelly S. 2019. Orthofinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:1–4.
Flynn JM, et al. 2020. Repeatmodeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A.* 117:9451–9457.
Geyer LB, Lessios H. 2009. Lack of character displacement in the male recognition molecule, bindin, in atlantic sea urchins of the genus *Echinometra*. *Mol Biol Evol.* 26:2135–2146.
Geyer LB, Palumbi SR. 2005. Conspecific sperm precedence in two species of tropical sea urchins. *Evolution* 59:97–105.
Guan D, et al. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36:2896–2898.
Harvey EB. 1956. The American arbia and other sea urchins. Princeton (NJ): Princeton University Press.
Ketchum RN, et al. 2022. A chromosome-level genome assembly of the highly heterozygous sea urchin *Echinometra* sp. EZ reveals adaptation in the regulatory regions of stress response genes. *Genome Biol Evol.* 14:11.
Kinjo SK, et al. 2018. Hpbase: a genome database of a sea urchin, *Hemicentrotus pulcherrimus*. *Dev Growth Differ.* 60:174–182.
Koren S, Walenz BP, Berlin K, Miller JR, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27:722–736.
Kroh, A, Mooi, R. 2022. World Echinoidea database. *Echinometra* Accessed through: World Register of Marine Species at: <https://www.marinespecies.org/aphia.php?p=taxdetails&id=213383> on 2022-12-05) in December 2022.
Landry C, et al. 2003. Recent speciation in the indo-west pacific: rapid evolution of gamete recognition and sperm morphology in cryptic species of sea urchin. *Proc R Soc London Ser B Biol Sci.* 270:1839–1847.
Lessios HA. 2007. Reproductive isolation between species of sea urchins. *Bull Mar Sci.* 81:191–208.
Lessios HA. 2011. Speciation genes in free-spawning marine invertebrates. *Integr Comp Biol.* 51:456–465.

- Lessios HA, Cunningham CW. 1990. Gametic incompatibility between species of the sea urchin *Echinometra* on the two sides of the isthmus of Panama. *Evolution* 44:933–941.
- Li H, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Lomsadze A, Burns PD, Borodovsky M. 2014. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.* 42:e119.
- McCartney MA, Keller G, Lessios HA. 2000. Dispersal barriers in tropical oceans and speciation of atlantic and eastern pacific *Echinometra* sea urchins. *Mol Ecol.* 9:1391–1400.
- McCartney MA, Lessios HA. 2002. Quantitative analysis of gametic incompatibility between closely related species of neotropical sea urchins. *Biol Bull.* 202:166–181.
- 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.
- McIntyre DC, et al. 2013. Short-range wnt5 signaling initiates specification of sea urchin posterior ectoderm. *Development* 140:4881–4889.
- Metz EC, Palumbi SR. 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Mol Biol Evol.* 13:397–406.
- Mortensen T. 1921. Studies of the development and larval forms of echinoderms. Copenhagen: G.E.C. Gad.
- O’Leary NA, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44:D733–D745.
- Palumbi SR. 1994. Genetic divergence, reproductive isolation, and marine speciation. *Annu Rev Ecol Syst.* 25:547–572.
- Palumbi SR. 1996. What can molecular genetics contribute to marine biogeography? An urchin’s Tale. *J Exp Mar Biol Ecol.* 203:75–92.
- Palumbi SR, et al. 1997. Speciation and population genetic structure in tropical pacific sea urchins. *Evolution* 51:1506–1517.
- Palumbi SR. 1998. Species formation and the evolution of gamete recognition loci. In: Howard DJ and Berlocher SH, editors. *Endless forms: species and speciation*. Oxford: Oxford University Press. p. 271–278.
- Palumbi SR. 2009. Speciation and the evolution of gamete recognition genes: pattern and process. *Heredity (Edinburg)* 102:66–76.
- Palumbi SR, Lessios HA. 2005. Evolutionary animation: how do molecular phylogenies compare to Mayr’s Reconstruction of speciation patterns in the sea? *Proc Natl Acad Sci U S A.* 102(Suppl. 1):6566–6572.
- Palumbi SR, Metz EC. 1991. Strong reproductive isolation between closely related tropical sea urchins (genus *Echinometra*). *Mol Biol Evol.* 8:227–239.
- Sodergren E, et al. 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 314:941–952.
- Stanke M, Schoffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden markov model that uses hints from external sources. *BMC Bioinformatics* 7:62.
- Walker BJ, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963.
- Wang Y, et al. 2012. MCSscanx: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40(7):e49.
- 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.* 13:7.
- Zigler KS, et al. 2005. Sea urchin bindin divergence predicts gamete compatibility. *Evolution* 59:2399–2404.

Associate editor: Prof. Christopher Wheat