Title

A reference genome for ecological restoration of the sunflower sea star, Pycnopodia helianthoides.

Journal

Journal of Heredity, 115(1)

Authors

Schiebelhut, Lauren

DeBiasse, Melissa

Gabriel, Lars

et al.

Publication Date

2024-02-03

DOI

10.1093/jhered/esad054

Peer reviewed

Journal of Heredity, 2024, **115**, 86–93
https://doi.org/10.1093/jhered/esad054
Advance access publication 20 September 2023
Genome Resources

American Genetic Association

OXFORD

Genome Resources

A reference genome for ecological restoration of the sunflower sea star, *Pycnopodia helianthoides*

Lauren M. Schiebelhut[1], Melissa B. DeBiasse[1,2], Lars Gabriel[3], Katharina J. Hoff[3], Michael N. Dawson[1,*]

[1]Life & Environmental Sciences, University of California, Merced, CA, United States
[2]Department of Biology, Radford University, Radford, VA, United States
[3]Institute for Mathematics and Computer Science & Center for Functional Genomics of Microbes, University of Greifswald, Greifswald, Germany

[*]Corresponding author: Life & Environmental Sciences, University of California, Merced, CA, United States. Email: mdawson@ucmerced.edu

Corresponding Editor: Rachel Meyer

Abstract

Wildlife diseases, such as the sea star wasting (SSW) epizootic that outbroke in the mid-2010s, appear to be associated with acute and/or chronic abiotic environmental change; dissociating the effects of different drivers can be difficult. The sunflower sea star, *Pycnopodia helianthoides*, was the species most severely impacted during the SSW outbreak, which overlapped with periods of anomalous atmospheric and oceanographic conditions, and there is not yet a consensus on the cause(s). Genomic data may reveal underlying molecular signatures that implicate a subset of factors and, thus, clarify past events while also setting the scene for effective restoration efforts. To advance this goal, we used Pacific Biosciences HiFi long sequencing reads and Dovetail Omni-C proximity reads to generate a highly contiguous genome assembly that was then annotated using RNA-seq-informed gene prediction. The genome assembly is 484 Mb long, with contig N50 of 1.9 Mb, scaffold N50 of 21.8 Mb, BUSCO completeness score of 96.1%, and 22 major scaffolds consistent with prior evidence that sea star genomes comprise 22 autosomes. These statistics generally fall between those of other recently assembled chromosome-scale assemblies for two species in the distantly related asteroid genus *Pisaster*. These novel genomic resources for *P. helianthoides* will underwrite population genomic, comparative genomic, and phylogenomic analyses—as well as their integration across scales—of SSW and environmental stressors.

Key words: Asteroidea, climate change, conservation, kelp forest, ocean health, sea star wasting

Introduction

In 2013, an outbreak of sea star wasting (SSW) disease caused measurable population-level impacts in over a dozen species of sea stars (Dawson et al. in press), with approximately another dozen species being documented as susceptible to the disease (Eisenlord et al. 2016). The most severely impacted was the sunflower sea star (*Pycnopodia helianthoides*), a keystone predator of urchins in kelp forests. The sunflower star suffered 87.8% mortality across its northern range (Aleutian Islands, Alaska, to Cape Flattery, Washington, United States) and was eradicated from the ~2,700 km southern half of its range (i.e. 99% to 100% mortality; Cape Flattery, Washington, United States, to Baja California, Mexico; Fig. 1), leading to the species being listed as *Critically Endangered* by the International Union for Conservation of Nature (IUCN; Gravem et al. 2021; Hamilton et al. 2021; Heady et al. 2022). The ecological consequences have been devastating: explosion of urchin populations, decimation of kelp forests, and limited signs of rebound of kelp ecosystems (McPherson et al. 2021) including the sunflower star itself (Gravem et al. 2021; Hamilton et al. 2021).

Despite the consequences and magnitude of the SSW epizootic of the 2010s—it was one of the most widespread marine mass mortality events ever documented—a decade later we are still struggling to understand its causes and implications and to develop strategies for redress. One of many outstanding questions is: why did different syntopic species of sea stars show such different levels of susceptibility? For example, while the sunflower sea star was singularly the most severely affected, being driven to functional extinction across most of its range, other sea star species (e.g. *Dermasterias imbricata*) appear to have been relatively resilient to the outbreak (Dawson et al. in press). Uncovering the genomic factors that influence the differential risk and consequences of SSW may provide insight into the causes and also information important for conservation efforts. Genomic information is an essential component, not only of understanding the impacts of disease and other stressors, but also for monitoring diversity in the wild, informing captive breeding and rearing efforts, and guiding translocation and/or outplanting decisions (Heady et al. 2022). Therefore, developing genomic resources that convey resistance to SSW disease and associated risk factors can be seen as critical infrastructure development.

Received June 10, 2023; Accepted September 29, 2023

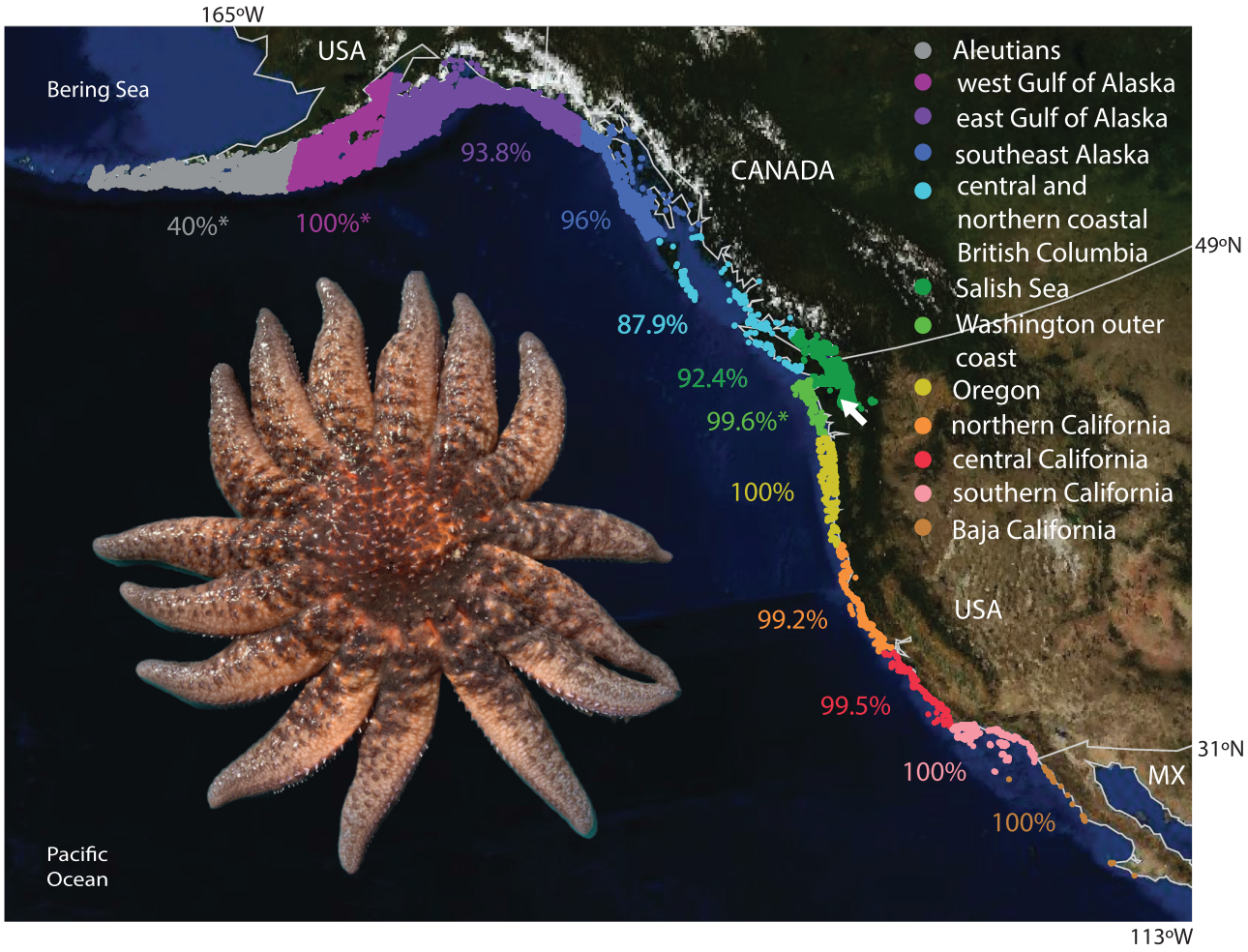

Fig. 1. The sunflower sea star, *Pycnopodia helianthoides*, historically occurred across ~6,000 km of coastline from Alaska, United States, to Baja California, Mexico (MX), but exceptional mortality during an outbreak of SSW beginning in 2013 halved its range, restricting the species to Alaska, British Columbia, and northern Washington. Colored regions show estimated percent declines in sunflower sea star population density due to SSW between 2013 and 2017; * = low confidence due to small sample sizes; based on Hamilton et al. (2021). Map modified from Heady et al. (2022) and Gravem et al. (2021) with permission. White arrow = the sampling locality of the animal used in this study. (inset) Photograph of *P. helianthoides* from the Salish Sea, courtesy of Taylor Frierson, Washington Department of Fish and Wildlife.

To facilitate conservation actions for the sunflower star, we here report the first annotated chromosome-scale reference genome assembly for *P. helianthoides*. We then compare basic descriptive statistics with other recently published sea star genomes (Ruiz-Ramos et al. 2020; DeBiasse et al. 2022).

Methods

Collection and preparation of sample

One sunflower sea star was collected at a depth of approximately 10 to 13 m from Octopus Hole in Hood Canal, Washington, United States (47.445750, −123.113709) on 10 January 2020 by the Washington Department of Fish & Wildlife. The specimen was biopsied and dissected tissues were flash frozen in liquid nitrogen then stored at −80 °C. A voucher specimen is archived in the Royal BC Museum, Victoria, BC, Canada (specimen ID# RBCM 023-00008-001; M0D057908R). Frozen tissue samples were shipped on dry ice to Dovetail Genomics (Scotts Valley, California) for high molecular weight DNA extraction, library preparation, proximity ligation using Omni-C, and sequencing.

Genome sequencing and assembly

DNA was extracted using the Qiagen blood and cell culture DNA midi kit following the manufacturer instructions. DNA samples were quantified on a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, California, United States). The PacBio SMRTbell library (~20 kb) for PacBio Sequel was constructed using SMRTbell Express Template Prep Kit 2.0 (PacBio, Menlo Park, California, United States) using the manufacturer-recommended protocol. The library was bound to polymerase using the Sequel II Binding Kit 2.0 (PacBio). Sequencing was performed on a PacBio Sequel II using 8M SMRT cells.

Wtdbg2 (Ruan and Li 2020) was used to assemble the PacBio CLR sequence reads. Blobtools v.1.1.1 (Laetsch and Blaxter 2017) was used to identify potential contamination in the assembly based on blast (v.2.9) results of the assembly against the NT database. A fraction of the scaffolds was identified as contaminant and removed from the assembly. The filtered assembly was then used as an input for purge_dups v.1.1.2 (Guan et al. 2020) which filtered potential haplotypic duplications from the assembly, resulting in the purged PacBio assembly (Table 1).

Table 1. Assembly pipeline and programs used.

Task	Software	Version
PacBio CLR assembly	wtdbg2	2.5
Contamination screen	blobtools	1.1.1
	Blast	2.9
Duplication filter	purge_dups	1.1.2
Omni-C read alignment	bwa	0.7.17
Scaffolding	HiRise	1.0
RNA-Seq adapter trimming	Trimmomatic	0.39
Mitochondrial genome assembly	MitoFinder	4.1
Repeat library generation	RepeatModeler2	1.0.11
Repeat masking	RepeatMasker2	4.0.7
RNA-Seq read mapping	HSAT2	2.1.0
Gene prediction	BRAKER3	3.0.2
	GeneMark-ETP	1.0
	AUGUSTUS	3.4.0
Protein sequence annotation	Interproscan	5.61-93.0
Genome size and heterozygosity	GenomeScope	2.0
	Jellyfish	2.3.0
Genome quality control	gVolante	2.0
	BUSCO	5.0
Genome–genome alignment	MUMmer	4.0
	Dot	1.0

Software citations are listed in the text.

The purged PacBio assembly and Dovetail Omni-C library reads were used as input data for HiRise v1.0, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies (Putnam et al. 2016). Dovetail Omni-C library sequences were aligned to the draft input assembly using bwa (https://github.com/lh3/bwa). The separations of Dovetail Omni-C read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, to score prospective joins, and make joins above a threshold.

Transcriptome sequencing

Three different *P. helianthoides* specimens were used to generate RNA sequence for genome annotation (M0D057909S, M0D059933O, and M0D060390D; archived in the Dawson Lab collection at the University of California, Merced). The first specimen was a male and included tissues from the dermis, pyloric cecum, and tube feet; the second specimen was a female and included tissues from dermis, pyloric cecum, tube feet, and gonad; the third specimen was a male and included tissue from the gonad. All three stars were collected from the same locality on the same date as the genome specimen and tissues were preserved in RNA later prior to RNA extraction. RNA was extracted following the "Purification of Total RNA from Animal Tissues" protocol for the Qiagen RNeasy Mini Kit with the optional on-column DNase digestion. Tissue disruption and homogenization were achieved using two 2 mm diameter chrome steel beads per sample, shaken at 1,500 rpm in a Mini G 1600 tissue homogenizer for a duration of 40 to 50 s.

Total RNA was submitted to Novogene (Sacramento, California, United States) for quality control, library preparation, and sequencing. RNA integrity and quantitation were assessed on an Agilent 2100. Library preparation consisted of mRNA enrichment using oligo(dT) beads and rRNA removal using the Ribo-Zero kit, then mRNA fragmented randomly with fragmentation buffer. cDNA was synthesized using the mRNA template and random primers of hexamers, followed by second-strand synthesis. After terminal repair, A-ligation, and sequencing adapter ligation, the double-stranded cDNA was size-selected and enriched using PCR to generate the final library. A Qubit 2.0 was then used to assess the library concentration, Agilent 2100 to assess the insert size, and qPCR to quantify the library effective concentration. Paired-end 150 bp sequences were generated for the libraries on a NovaSeq 6000 targeting 12 Gb of raw data per sample.

Mitochondrial genome assembly

We used the program MitoFinder v1.4.1 (Allio et al. 2020) to assemble and annotate a mitochondrial genome sequence for *P. helianthoides*. First, we trimmed adapters and low-quality bases from the RNA-Seq data using Trimmomatic v0.39 (Bolger et al. 2014). Trimmed RNA-Seq reads were assembled in MitoFinder using the mitochondrial genome for *Pisaster ochraceus* (GenBank accession number: NC_042741) as a reference (Table 1).

Nuclear genome annotation

To prepare the input data—genomic sequences, short-read RNA-Seq, and database of protein sequences—for gene prediction, we performed preprocessing as follows (Table 1): We used RepeatModeler2 v-open-1.0.11 (Flynn et al. 2020) to generate a species-specific repeat library, and RepeatMasker2 v-open-4.0.7 (2013–2015Smit et al. 2015) to mask the genomic sequences for repeats. We mapped eight paired RNA-Seq libraries to the genome using HISAT2 v2.1.0 (Kim et al. 2019) and sorted them with Samtools v1.13 (Li et al. 2009). We obtained a protein sequence database from OrthoDB 11 (Kuznetsov et al. 2023) consisting of 15,257,394 sequences from species of the clade Metazoa, using the orthodb-clades pipeline (Bruna et al. 2023a).

For the prediction of gene loci and structures, we used the BRAKER3 pipeline v3.0.2 (Gabriel et al. 2023). It uses genome sequence and short-read RNA-Seq data as input to subsequently run the gene prediction tool GeneMark-ETP v1.00 (Bruna et al. 2023b) and then AUGUSTUS v3.4.0 (Stanke et al. 2006) for its annotation process. In its last step, BRAKER3 combines and filters the result of both tools into a final prediction. We skipped this last step as the gene count and BUSCO v5.4.4 (Manni et al. 2021) with metazoa_odb10 score indicated that in this step too many gene models were filtered out due to a comparably low amount of extrinsic evidence. Instead, we used the AUGUSTUS prediction with hints as the final gene set.

Using the resulting data—genome assembly, short-read RNA-seq, genome annotation—we created a genome browser hub for the UCSC genome browser (Kent et al. 2002). We used the MakeHub software for the automated generation of the assembly hub. For functional annotation of the protein sequences, we used Interproscan v5.61-93.0 (Jones et al. 2014).

Nuclear genome-summary statistics

We generated k-mer counts ($k = 21$) for Omni-C Illumina reads with jellyfish v2.3.0 (Marçais and Kingsford 2011) using default parameters. We used this k-mer dataset to estimate genome size and heterozygosity with GenomeScope v2.0 (Ranallo-Benavidez et al. 2020) using default parameters. We used RepeatModeler v4.1.2-pl and RepeatMasker v2.0.1 implemented in Dfam TETools v1.3.1 (http://www.repeatmasker.org) to characterize repetitive transposable elements in each genome. We used BUSCO v5.0 (Simão et al. 2015) implemented in gVolante v2.0 (Nishimura et al. 2017) to measure the quality, contiguity, and completeness of the genome assembly by searching it against the metazoan ortholog database (metazoa_odb10) which contains 954 core genes. To test if a drop-off in scaffold size corresponded to the number of predicted sea star chromosomes (Saotome and Komatsu 2002), we performed a clustering analysis on the lengths of the longest 30 *P. helianthoides* scaffolds using the k-means++ method (Arthur and Vassilvitskii 2006) implemented in the SciStatCalc webserver (https://scistatcalc.blogspot.com/). The expectation is that longer scaffolds, which represent putative chromosomes, will cluster in one group while shorter scaffolds will cluster in a second group based on a significant change in size between the last putative chromosome scaffold and the first non-chromosome.

Comparison to previously published sea star genomes

We compared the *P. helianthoides* genome to chromosome-level assemblies available for *P. ochraceus* (Ruiz-Ramos et al. 2020) and *Pisaster brevispinus* (DeBiasse et al. 2022). We generated BUSCO metrics for the *Pisaster* genomes as described above. To assess homology of the *P. helianthoides* scaffolds with the 22 chromosomes identified in the *P. ochraceus* genome, we aligned the 22 largest *P. helianthoides* scaffolds to the *P. ochraceus* chromosomes using the program NUCMER in the MUMmer v4.0 package (Marçais et al. 2018) and visualized the alignment using the program Dot (github.com/marianattestad/dot).

Results

Genome assemblies

The nuclear genome sequence was assembled into 1,610 scaffolds and was 483,884,608 bp long with an N50 of 21,765,409 bp (Table 2). Results of the BUSCO analysis showed the genome was high quality, containing 96.1% complete and partial single-copy core genes and low numbers of missing (3.9%), fragmented (1.4%), and duplicated (0.5%) core genes (Table 2). The genome was also highly contiguous with 94.6% of the total sequence length contained in the largest 22 scaffolds. The RepeatModeler and GenomeScope analyses estimated the genome contained 43% and 34% repetitive sequence, respectively.

The mitochondrial genome assembly was 16,326 bp long. The mitochondrion contained 13 protein-coding genes, 12 ribosomal genes, and 22 transfer RNAs. The annotated mitochondrial genome assembly is available on GenBank under accession number OR345354.

Several lines of evidence supported the hypothesis that, like other sea stars, the *P. helianthoides* nuclear genome is arranged into 22 chromosomes. The contact map showed that the Omni-C reads aligned to the genome were organized into 22 major bins (Fig. 2A). Cluster analysis of scaffold length grouped scaffolds 1 to 22 together in one group versus scaffolds 23 to 30 in a second group (Fig. 2B). Alignments showed the largest 22 scaffolds in *P. helianthoides* corresponded well to the *P. ochraceus* chromosomes, with areas of putative sequence inversion and sequence gaps across the alignment (Fig. 2C).

Transcriptome sequencing and annotation

RNA sequencing yielded 57.2 to 85.5 million reads per individual (median 83.3 M), after filtering reads of low quality,

Table 2. Summary statistics and BUSCO scores for sea star genomes.

Pycnopodia helianthoides	*Pisaster brevispinus*	*Pisaster ochraceus*	
Citation	This paper	DeBiasse et al. (2022)	Ruiz-Ramos et al. (2020)
NCBI BioProject	PRJNA980115	PRJNA810506	PRJNA532896
Sequence methods (genomic, proximity)	PacBio CLR, Omni-C	PacBio CCS, Omni-C	Illumina, Hi-C
Assembly length (bp)	483,884,608	505,343,882	401,943,971
Sequences	1610	127	1,844
Length in predicted chromosomes (%)	95	84	99
GC content (%)	39.1	39.5	39.0
Contig N50 (Mb)	1.9	4.6	0.009
Longest contig (Mb)	32.5	13.8	13.0
Scaffold N50 (Mb)	21.8	21.4	21.9
Longest scaffold (Mb)	36.9	31.2	31.5
Gaps ≥5N's	675	144	104,686
Complete single-copy genes	904 (94.2%)	937 (98.2%)	818 (85.7%)
Complete + partial single-copy genes	917 (96.1%)	951 (99.7%)	914 (95.8%)
Duplicated genes	5 (0.5%)	5 (0.5%)	1 (0.1%)
Fragmented genes	13 (1.4%)	9 (0.9%)	96 (10.1%)
Missing genes	37 (3.9%)	4 (0.4%)	40 (4.2%)

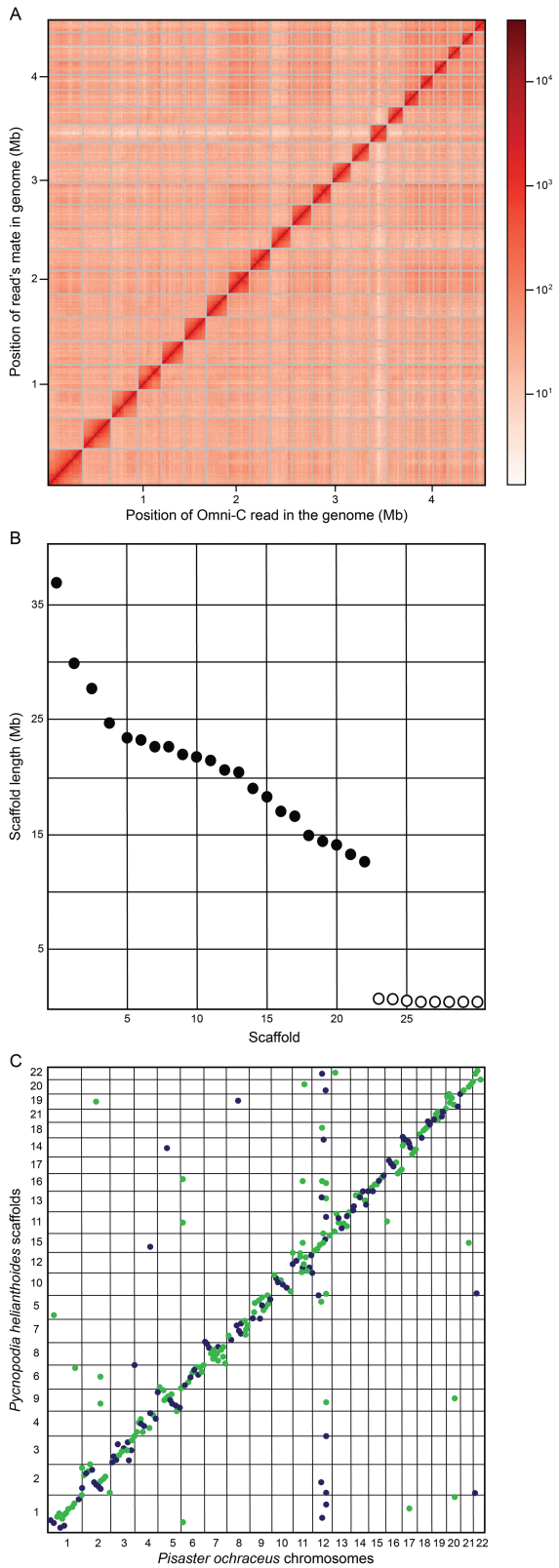

Fig. 2. (A) Contact map of Illumina Omni-C reads mapped to the genome assembly. The position of each R1 read is indicated along the *x*-axis and the position of each R1 read's mate is indicated on the *y*-axis. Each bin in the contact map corresponds to the sequence data supporting the physical linkage (i.e. contact) between two genomic regions. The color scale indicates the number of read pairs in each bin. (B) Graph of k-means clustering performed on lengths of largest 30 *Pycnopodia* genomic scaffolds. Closed circles indicate scaffolds placed in group 1 and open circles represent scaffolds placed in group 2. (C) Whole genome alignment between the predicted *Pisaster ochraceus* 22 chromosomes (*x*-axis) and 22 longest *Pycnopodia helianthoides* scaffolds (*y*-axis). Green dots represent areas of sequence alignment in the same direction and purple dots represent areas of inverted sequence alignment in *P. helianthoides* (the query) relative to *P. ochraceus* (the reference). Light gray lines indicate chromosome and scaffold boundaries. Each scaffold-to-chromosome alignment block is scaled by the length of the *P. ochraceus* chromosome (*x*-axis) and the *P. helianthoides* scaffold (*y*-axis).

those that contained adapter, and those that contained more than 10% N's. Augustus predicted 24,184 protein-coding genes and a total of 26,581 transcripts for these. The ratio of mono-exonic to multi-exonic genes was 0.29. The median number of exons was 4, while the largest number of observed exons in a single transcript was 142. BUSCO reported a completeness of 92.3% in the protein-coding genes, 3.6% fragmented BUSCOs, and 4.1% missing BUSCOs.

Discussion

This first annotated genome assembly for *P. helianthoides* provides a highly contiguous and largely complete chromosome-scale resource to underwrite conservation actions for a kelp forest keystone predator. As the eighth chromosome-scale genome assembly for sea stars, it corroborates existing evidence for a highly conserved number of chromosomes across Asteroidea ($n = 22$: Ruiz-Ramos et al. 2020; DeBiasse et al. 2022; see also Saotome and Komatsu 2002; *Asterias rubens PRJEB33974; Luidia sarsii PRJEB61567; Marthasterias glacialis PRJEB46624; Patiria pectinifera*, PRJNA882565; *Plazaster borealis* PRJNA776097). This high degree of conservation of the large-scale architecture of sea star genomes suggests they will be amenable to phylogenomic comparisons of the impacts of wasting and, with variation in the form of inversions and insertion–deletions, valuable resources for explaining differences in SSW susceptibility and impact among species (Dawson et al. in press).

Considering the chromosome-scale assemblies for three northeastern Pacific sea stars differentially affected by SSW—*P. helianthoides*, *P. ochraceus* (Ruiz-Ramos et al. 2020), and *P. brevispinus* (DeBiasse et al. 2022)—interpreting genomic information may be complicated by the sequencing approach, which likely contributes to differences in BUSCO scores (Lee et al. 2023) and assessment of repeats (Logsdon et al. 2020). The *P. helianthoides* assembly used continuous long reads (CLR) which impact alignment due to the high sequencing error rate, while *P. brevispinus* used more accurate circular consensus sequencing (CCS); *P. ochraceus* employed only short reads. When comparing the proportion of repeats, the expected increase in the percentage of repetitive regions when using long-read technology, which has the ability to sequence through long repeats, does occur: 24% in *P. ochraceus* (Ruiz-Ramos et al. 2020) versus 43.1% in *P. helianthoides* and 43.3% in *P. brevispinus*. However, despite the differences in accuracy in CLR relative to CCS long-read technology, *P. helianthoides* and *P. brevispinus* shared a similar proportion of repetitive element content. Thus, differences in sequencing approach can contribute to the differences seen in summary statistics and are important but not prohibitive considerations for comparative genomic analyses.

Having a new, genomic, perspective on SSW is important because the rapidity and severity of the 2013 SSW outbreak meant that early ecological studies tended toward post hoc analyses and were unable to find strong evidence of correlates with disease prevalence or morbidity (Hewson et al. 2018; Moritsch 2018). The sudden and unpredictable emergence and rapid progression of wasting, or wasting-like, symptoms in captive animals also has confounded laboratory experiments (Hewson et al. 2018). Given the elusive etiology of SSW using standard procedures (e.g. Hewson et al. 2018), we have been developing the strategy of "genomic autopsies" to infer attributes of the disease and to identify candidate loci associated with survival of SSW (Ruiz-Ramos et al. 2020). Initial comparisons of preexisting ddRAD and RNA-seq datasets from *P. helianthoides* and *P. ochraceus* mapped to the *P. ochraceus* genome identified positional similarity of differentially expressed genes and outlier alleles associated with wasting and/or with temperature stress (Ruiz-Ramos et al. 2020). By continuing to grow the genomic resources available within and across species and for individuals at different wasting stages, a more complete and nuanced understanding of the molecular and environmental underpinnings of SSW should emerge. Moreover, by comparing historical samples from the SSW outbreak versus more recent specimens that show symptoms of wasting in different circumstances, we aim to identify genomic profiles that illuminate whether there is a single SSW disease or multiple SSW-like syndromes that are being conflated and should be considered separately.

Identifying genomic characteristics of SSW or SSW-like symptoms may also lead to differentiating causes, consequences, and responses. Hypothesized cause(s) of the 2013 outbreak of SSW include viral pathogen (Hewson et al. 2014; Fuess et al., 2015; Bucci et al. 2017), warmer temperature (Eisenlord et al. 2016; Kohl et al. 2016; Harvell et al. 2019), cooler temperature (Menge et al. 2016), reduced precipitation (Hewson et al. 2018), dysoxia-induced dysbiosis (Aquino et al. 2021), and/or interactions thereof (Hewson 2021), mediated by oceanographic and/or weather conditions (Aalto et al. 2020; Aquino et al. 2021; Hewson 2021; Dawson et al. in press), and moderated by individual size (Eisenlord et al. 2016; Menge et al. 2016) and genotype (Schiebelhut et al. 2018). Each of these putative causes could feasibly leave different genomic signatures in surviving individuals and, therefore, require different considerations (e.g. rearing conditions or reproductive crosses) when undertaking captive breeding.

Such differences may have a phylogenetic pattern. SSW-associated mortality was not synchronous across species despite broad sympatry; rather, mortality progressed taxonomically, emerging first in one species and then in another, progressing through time. Though there was some geographic heterogeneity, in general, very early and high mortality was observed in Order Forcipulatida, *P. helianthoides* (Hamilton et al. 2021). Mortality was subsequent and less severe in another forcipulatid, *P. ochraceus* (Eisenlord et al. 2016; Montecino-Latorre et al. 2016; Dawson et al. in press). Other taxa were affected even later and/or much less (e.g. orders Valvatida and Spinulosida; Montecino-Latorre et al. 2016; Dawson et al. in press). These observations of serial outbreak suggest a phylogenomic component to susceptibility, and that comparative genomic analyses could clarify why species were more or less susceptible to SSW.

Thus the annotated reference genome developed here will provide a foundational resource that can inform genomic autopsies of *P. helianthoides* and, with similar resources for other sea star species, another step toward diagnosing shared causes. Emerging commonalities will help fulfill the goals of restoration efforts, to efficiently and effectively enhance population viability by increasing population size *and* genetic diversity (Gomes Destro et al. 2018) and adaptive potential via genetic rescue (Hohenlohe et al. 2019). Importantly, timely genomic analyses can ensure that we *complement* and *augment*, not replace, the natural resilience and recovery potential in the subset of populations of sunflower sea stars that persisted through the outbreak.

Lessons learned from conservation genomics emphasize the importance of harnessing adaptive loci, while preserving adaptive potential, without compromising the natural selective process of wild survivors through outbreeding depression (Hohenlohe et al. 2019). Recent selective breeding programs in mollusks with inbreeding control achieved improved disease resistance by 15% per generation (Hollenbeck and Johnston 2018). Thus, the genomic resources generated here can empower managers to use genotyping to repopulate the seascape with resistant individuals while simultaneously avoiding a genetic bottleneck, which would be the likely outcome if natural recolonization occurred stepwise (Slatkin and Excoffier 2012) or if reintroduction used standard selective breeding to establish disease resistance (Leberg and Firmin 2008).

Acknowledgments

The project was made possible by the vision and partnership of all at Revive & Restore. We thank the Washington Department of Fish & Wildlife (WDFW) subtidal shellfish dive team—Taylor Frierson, Emily Loose, and Katie Sowul—for collecting the genome specimen and Jason Hodin at Friday Harbor Laboratories for sampling the tissues of the specimen for sequencing. A permit was not required for specimen collection by WDFW. The team at Dovetail Genomics (Cantata Bio) were yet again considerate, helpful, and professional. University of California Merced Cyberinfrastructure and Research Technologies (CIRT) staff members Robert Romero and Sarvani Chadalapaka provided computational support. Portions of this research were conducted using the CIRT MERCED (NSF-MRI, #1429783) and Pinnacles (NSF-MRI, #2019144) clusters.

Funding

This work was supported primarily by Revive & Restore's Wild Genomes Program (grant #2020_009) and in part by the US National Science Foundation's Biological Oceanography program (OCE-1737381). LG is funded by the US National Institute of Health grant GM128145.

Data availability

We have deposited the primary data underlying these analyses with NCBI under BioProject PRJNA980115 as follows: genome assembly (JASTWB000000000) which incorporated PacBio reads and Omni-C reads; raw RNA-seq reads (SAMN35712975-SAMN35712982), and mitochondrial genome assembly (OR345354). The doi for the gene predictions is: https://doi.org/10.5061/dryad.51c59zwfd. *Pycnopodia helianthoides* genome is available for browsing on the UCSC

genome browser, hub info: https://bioinf.uni-greifswald.de/hubs/sunflower_sea_star/hub.txt

References

Aalto EA, Lafferty KD, Sokolow SH, Grewelle RE, Ben-Horin T, Boch CA, Raimondi PT, Bograd SJ, Hazen EL, Jacox MG, et al. Models with environmental drivers offer a plausible mechanism for the rapid spread of infectious disease outbreaks in marine organisms. Sci Rep. 2020:10:5975. doi:10.1038/s41598-020-62118-4.

Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz B, Delsuc F. MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. Mol Ecol Resour. 2020:20:892–905. doi:10.1111/1755-0998.13160

Aquino CA, Besemer RM, DeRito CM, Kocian J, Porter IR, Raimondi PT, Rede JE, Schiebelhut LM, Sparks JP, Wares JP, et al. Evidence that microorganisms at the animal-water interface drive sea star wasting disease. Front Microbiol. 2021:11:113278.

Arthur D, Vassilvitskii S. k-means++: the advantages of careful seeding. SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. 2006:1027–1035.

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

Bruna T. orthodb-clades. 2023a [accessed 2023 April 21]. https://github.com/tomasbruna/orthodb-clades.

Bruna T, Lomsadze A, Borodovsky M. GeneMark-ETP: automatic gene finding in eukaryotic genomes in consistence with extrinsic data. bioRxiv. 2023b:2023–01. doi:10.1101/2023.01.13.524024.

Bucci C, Francoeur M, McGreal J, Smolowitz R, Zazueta-Novoa V, Wessel GM. Sea Star Wasting Disease in Asterias forbesi along the Atlantic Coast of North America. PLoS ONE. 2017:12:e0188523. doi:10.1371/journal.pone.0188523.

Dawson MN, Duffin P, Giakoumis M, Schiebelhut LM, Beas-Luna R, Bosley KL, Castilho R, Ewers-Saucedo C, Gavenus KA, Keller A, et al. A decade of death and other dynamics: deepening perspectives on the diversity and distribution of sea stars and wasting. Biol Bull. in press.

DeBiasse MB, Schiebelhut LM, Escalona M, Beraut E, Fairbairn C, Marimuthu MP, Nguyen O, Sahasrabudhe R, Dawson MN. A chromosome-level reference genome for the giant pink sea star, *Pisaster brevispinus*, a species severely impacted by wasting. J Heredity. 2022:113:689–698.

Eisenlord ME, Groner ML, Yoshioka RM, Elliott J, Maynard J, Fradkin S, Turner M, Pyne K, Rivlin N, van Hooidonk R, et al. Ochre star mortality during the 2014 wasting disease epizootic: role of population size structure and temperature. Philos Trans R Soc London B Biol Sci. 2016:371:20150212.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA. 2020:117:9451–9457.

Fuess LE, Eisenlord ME, Closek CJ, Tracy AM, Mauntz R, Gignoux-Wolfsohn S, Moritsch MM, Yoshioka R, Burge CA, Harvell CD, et al. Up in arms: immune and nervous system response to sea star wasting disease. PLoS One. 2015:10:e0133053.

Gabriel L, Bruna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. BRAKER3: Fully Automated Genome Annotation Using RNA-Seq and Protein Evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. bioRxiv. 2023.

Gomes Destro GF, De Marco P, Terribile LC. Threats for bird population restoration: a systematic review. Perspect Ecol Conserv. 2018:16:68–73.

Gravem SA, Heady WN, Saccomanno VR, Alvstad KF, Gehman ALM, Frierson TN, Hamilton SL. Pycnopodia helianthoides. IUCN red list of threatened species. 2021 [accessed 2023 August 26]. https://www.iucnredlist.org/species/178290276/197818455

Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 2020:36:2896–2898.

Hamilton SL, Saccomanno VR, Heady WN, Gehman AL, Lonhart SI, Beas-Luna R, Francis FT, Lee L, Rogers-Bennett L, Salomon AK, et al. Disease-driven mass mortality event leads to widespread extirpation and variable recovery potential of a marine predator across the eastern Pacific. Proc R Soc Lond B Biol Sci. 2021:288:20211195.

Harvell CD, Montecino-Latorre D, Caldwell JM, Burt JM, Bosley K, Keller A, Heron SF, Salomon AK, Lee L, Pontier O, et al. Disease epidemic and a marine heat wave are associated with the continental-scale collapse of a pivotal predator (*Pycnopodia helianthoides*). Sci Adv. 2019:5:eaau7042.

Heady WN, Beas-Luna R, Dawson MN, Eddy N, Elsmore K, Francis FT, Frierson T, Gehman AL, Gotthardt T, Gravem SA, et al. Roadmap to recovery for the sunflower sea star (*Pycnopodia helianthoides*) along the west coast of North America. Sacramento, CA: The Nature Conservancy; 2022:44.

Hewson I. Microbial respiration in the asteroid diffusive boundary layer influenced sea star wasting disease during the 2013-2014 northeast Pacific Ocean mass mortality event. Mar Ecol Prog Ser. 2021:668:231–237.

Hewson I, Bistolas KS, Quijano Cardé EM, Button JB, Foster PJ, Flanzenbaum JM, Kocian J, Lewis CK. Investigating the complex association between viral ecology, environment, and northeast Pacific sea star wasting. Front Mar Sci. 2018:5:77.

Hewson I, Button JB, Gudenkauf BM, Miner BG, Newton AL, Gaydos JK, Wynne J, Groves CL, Hendler G, Murray M, et al. Densovirus associated with sea-star wasting disease and mass mortality. Proc Natl Acad Sci USA. 2014:111:17278–17283.

Hohenlohe PA, McCallum HI, Jones ME, Lawrance MF, Hamede RK, Storfer A. Conserving adaptive potential: lessons from Tasmanian devils and their transmissible cancer. Conserv Genet. 2019:20:81–87.

Hollenbeck CM, Johnston IA. Genomic tools and selective breeding in molluscs. Front Genet. 2018:9:253.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014:30:1236–1240. doi:10.1093/bioinformatics/btu031.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002:12:996–1006.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019:37:907–915.

Kohl WT, Mcclure TI, Miner BG. Decreased temperature facilitates short-term sea star wasting disease survival in the keystone intertidal sea star *Pisaster ochraceus*. PLoS One. 2016:11:e0153670.

Kuznetsov D, Tegenfeldt F, Manni M, Seppey M, Berkeley M, Kriventseva EV, Zdobnov EM. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. Nucleic Acids Res. 2023:51:D445–D451.

Laetsch DR, Blaxter ML. BlobTools: interrogation of genome assemblies. F1000Research. 2017:6:1287–1287.

Leberg PL, Firmin BD. Role of inbreeding depression and purging in captive breeding and restoration programmes. Mol Ecol. 2008:17:334–343. doi:10.1111/j.1365-294X.2007.03433.x.

Lee H, Kim J, Lee J. Benchmarking datasets for assembly-based variant calling using high-fidelity long reads. BMC Genomics. 2023:24:148.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. Bioinformatics. 2009:25:2078–2079.

Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. Nat Rev Genet. 2020:21:597–614.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021:38:4647–4654.

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. PLoS Comput Biol. 2018:14:e1005944.

Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011:27:764–770. doi:10.1093/bioinformatics/btr011.

McPherson ML, Finger DJ, Houskeeper HF, Bell TW, Carr MH, Rogers-Bennett L, Kudela RM. Large-scale shift in the structure of a kelp forest ecosystem co-occurs with an epizootic and marine heatwave. Comm Biol. 2021:4:298.

Menge BA, Cerny-Chipman EB, Johnson A, Sullivan J, Gravem S, Chan F. Sea star wasting disease in the keystone predator *Pisaster ochraceus* in Oregon: insights into differential population impacts, recovery, predation rate, and temperature effects from long-term research. PLoS One. 2016:11:e0153994.

Montecino-Latorre D, Eisenlord ME, Turner M, Yoshioka R, Harvell CD, Pattengill-Semmens CV, Nichols JD, Gaydos JK. Devastating transboundary impacts of sea star wasting disease on subtidal asteroids. PLoS One. 2016:11:e0163190.

Moritsch M. Ecological causes and consequences of Sea Star Wasting Syndrome on the Pacific coast [Doctoral dissertation], UC Santa Cruz; 2018. https://escholarship.org/uc/item/9dr8t5kq

Nishimura O, Hara Y, Kuraku S. gVolante for standardizing completeness assessment of genome and transcriptome assemblies. Bioinformatics. 2017:33:3635–3637. doi:10.1093/bioinformatics/btx445.

Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 2016:26:342–350.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020:11:1432. doi:10.1038/s41467-020-14998-3.

Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. Nat Methods. 2020:17:155–158.

Ruiz-Ramos DV, Schiebelhut LM, Hoff KJ, Wares JP, Dawson MN. An initial comparative genomic autopsy of wasting disease in sea stars. Mol Ecol. 2020:29:1087–1102.

Saotome K, Komatsu M. Chromosomes of Japanese starfishes. Zool Sci. 2002:19:1095–1103.

Schiebelhut LM, Puritz JB, Dawson MN. Decimation by sea star wasting disease and rapid genetic change in a keystone species, *Pisaster ochraceus*. Proc Natl Acad Sci USA. 2018:115:7069–7074.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015:31:3210–3212. doi:10.1093/bioinformatics/btv351.

Slatkin M, Excoffier L. Serial founder effects during range expansion: a spatial analog of genetic drift. Genetics. 2012:191:171–181.

Smit AFA, Hubley R., Green P. RepeatMasker Open-4.0. 2013–2015. www.repeatmasker.org

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 2006:34:W435–W439.