

# Twinkle twinkle brittle star, how I wonder what your genes are: *Ophioderma brevispinum* as a genomic resource for regeneration

**Vladimir Mashanov**

Wake Forest Institute for Regenerative Medicine

**Denis Jacob Machado** (✉ [denisjacobmachado@gmail.com](mailto:denisjacobmachado@gmail.com))

University of North Carolina at Charlotte Hill Louis Round Wilson Special Collections Library: The University of North Carolina at Chapel Hill <https://orcid.org/0000-0001-9858-4515>

**Robert Reid**

University of North Carolina at Charlotte

**Cory Brouwer**

University of North Carolina at Charlotte

**Janice Kofsky**

University of North Carolina at Charlotte

**Daniel A. Janies**

University of North Carolina at Charlotte

---

## Research Article

**Keywords:** Echinodermata, Ophiuroidea, brittle star, tissue regeneration, comparative genomics, Notch signaling pathway

**Posted Date:** February 23rd, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1148623/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

## RESEARCH

# Twinkle twinkle brittle star, how I wonder what your genes are: *Ophioderma brevispinum* as a genomic resource for regeneration

Vladimir Mashanov<sup>1,2\*</sup>, Denis Jacob Machado<sup>3</sup>, Robert Reid<sup>4</sup>, Cory Brouwer<sup>4</sup>, Janice Kofsky<sup>3</sup> and Daniel A Janies<sup>3</sup>

\*Correspondence:

[vmashano@wakehealth.edu](mailto:vmashano@wakehealth.edu)

<sup>1</sup>Wake Forest Institute for Regenerative Medicine, 391 Technology Way, 27101 Winston-Salem, NC, USA  
Full list of author information is available at the end of the article

## Abstract

**Background:** Echinoderms are established models in experimental and developmental biology however genomic resources are still lacking for many species. Here, we present the draft genome of *Ophioderma brevispinum*, an emerging model organism in the field of regenerative biology. This new genomic resource provides a reference experimental studies of regenerative mechanisms.

**Results:** We report a *de novo* nuclear genome assembly for the brittle star *O. brevispinum* and annotation facilitated by the transcriptome assembly. The final assembly is 2.68 Gb in length and contains 146,833 predicted protein-coding gene models. We also report a mitochondrial genome for this species, which is 15,831 bp in length, and contains 13 protein-coding, 22 tRNAs, and 2 rRNAs genes, respectively. In addition, 29 genes of the Notch signaling pathway are identified to illustrate the practical utility of the assembly for studies of regeneration.

**Conclusions:** The sequenced and annotated genome of *O. brevispinum* presented here provides the first such resource for an ophiuroid model species. Considering the remarkable regenerative capacity of this species, this genome will be an essential resource in future research efforts on molecular mechanisms regulating regeneration.

**Keywords:** Echinodermata; Ophiuroidea; brittle star; tissue regeneration; comparative genomics; Notch signaling pathway

## Background

Echinoderms are a phylum of marine invertebrates, which together with hemichordates constitute the group Ambulacraria. In turn, the Ambulacraria form a sister clade to the phylum Chordata within the monophyletic superphylum Deuterostomia. Echinoderms have attracted much attention from scholars from various disciplines of biology (e.g., ecology, evolution, and developmental biology [1]). One of the fascinating aspects of the biology of many echinoderm species is the ability to regenerate all the tissues of large body parts and organ systems completely [2]. Thus far, attempts to understand the molecular underpinnings of echinoderm regeneration have been driven by transcriptome and gene expression studies [3–7]. Although these approaches are valuable in identifying the sets of genes that are involved in the regeneration program, gene expression studies do not by themselves lead to a comprehensive mechanistic understanding of regeneration. A mechanistic understanding of regeneration will be achieved through the reconstruction of the cause-and-effect

relationships between regulatory and effector genes in functional genomics studies. The availability of genomic data is one of the essential prerequisites for these studies and it shall allow for the identification of *cis*-regulatory modules, reconstruction of chromatin accessibility maps, and the design of the genome editing experiments to probe for the function of candidate genes.

The efforts in echinoderm genomics were pioneered by the sequencing of the genome of the purple sea urchin, *Strongylocentrotus purpuratus* (Stimpson, 1857) [8]. Since then, the availability and affordability of the new generation high-throughput sequencing technologies have expanded genome sequencing and annotation projects to other species representing all five classes within the phylum. Table 1 lists the 21 echinoderm genomes that are currently available on public databases. These genomes, sequenced and annotated to a varying degree of completeness, have been submitted to the National Center for Biotechnology Information (NCBI) or other databases (e.g., [9–15]). These sequencing efforts, however, have rarely been driven by regeneration research. Only seven of the currently sequenced species have ever been studied in terms of their regenerative capacities (i.e., there is at least one published study on the topic). Only three of the sequenced species, the sea cucumbers *Holothuria glaberrima* and *Aposichopus japonicus* and the sea star *Patiria miniata*, are established model organisms in echinoderm regenerative biology (i.e., there has been continuous effort to characterize cellular and molecular events underlying regeneration). Two brittle star genomes are available at present, *Ophionereis fasciata* and *Ophiothrix spiculata*, but neither of those species has been used in studies of regenerative biology.

The primary aim of this contribution is to provide genomic data and tools for a highly regenerative brittle star (ophiuroid echinoderm) *Ophioderma brevispinum* [16], an emerging model organism in the field of regenerative biology [3] (Fig. 1). This species is capable of autotomizing and quickly regenerating its arms, the segmented body appendages. Arm regeneration is a classic example of "epimorphic" regeneration, which involves extensive unidirectional terminal growth through rapid generation of new tissues distal to the plane of the injury [3] (Fig. 1). As arms are often exposed to predators in natural habitats, brittle stars are known for their ability to sustain arm loss followed by remarkably high rates of arm regeneration. It has been estimated that regenerated tissues may account for as much as half of the total body weight in a brittle star individual [17].

The rationale for choosing this particular brittle star species for our research is as practical as it is fundamental. *O. brevispinum* is common in the Western Atlantic in shallow water and can be easily collected in numbers sufficient for molecular biology studies. During experiments, these animals are easily maintained in indoor aquaria and as research stock with minimal maintenance for extended periods (e.g., months). In addition, live individuals of *O. brevispinum* are available to all interested researchers, as they can be ordered from several commercial suppliers at a moderate price.

Our previous work on *O. brevispinum* demonstrated a critical role of the the Notch signaling pathway (Fig. 2) in ophiuroid arm regeneration [3]. This pathway is a key node of a complex hypernetwork of interconnected signaling pathways that mediate cell-cell interactions in various developmental contexts [18]. Thus a specific goal of

this paper includes describing the genomic composition of the key components of the Notch pathway [18]. A report of these genes in *O. brevispinum* is provided to serve as a fundamental toolkit for future studies on regeneration. This work will facilitate further research to unravel the functions of signaling pathways in regeneration and identify their target genes through functional genomics approaches.

Here, we report a *de novo* genome assembly and annotation of *Ophioderma brevispinum*. This is the third ophiuroid for which genome sequencing has been applied following low coverage sequencing of *Ophiothrix spiculata* [12] (NCBI's BioProject accession number PRJNA182997) and *Ophionereis fasciata* [15]. Our *Ophioderma brevispinum* genome and transcriptome [3] allow us to describe regulatory gene families to further explore the molecular bases of echinoderm regeneration. This use case is an example of one of the several applications for these genomic data that researchers will find.

## Results

### Sequencing

Three different genomic DNA library preparation and sequencing strategies were employed. First, PCR-free library preparation (“short” libraries) followed by sequencing on an Illumina HiSeq 25000 machine yielded  $2 \times 163,310,307$  250 nt paired-end reads with an overall GC content of 37% (NCBI's SRA accession number **will be added here**). Second, mate-paired libraries (“long” libraries) with approximately 3 Kbp insert size were sequenced as above and resulted in  $2 \times 103,843,354$  250 nt paired-end reads with an overall GC content of 41%. Third, PacBio long-read sequencing generated approximately 23 million reads with a total yield of 159 billion bp ( $51.3 \times$  coverage).

In addition, in order to facilitate the annotation efforts, we took advantage of the earlier RNA-Seq study [3] that generated 17,318,775 MiSeq and 832,245,006 HiSeq quality filtered and adapter trimmed reads used in a *de novo* transcriptome assembly.

### Nuclear DNA assembly statistics

The draft assembly of the *O. brevispinum* genome generated 88,538 scaffolds with the total length of the assembly of 2,684,874,461 bp (approximately 2.68 Gb) (Table 2). This value is close to the haploid genome size of 2.89 Gb independently determined by a densitometry assay. Scaffolds range from 2,035 bp to 61,2917 bp, with a mean scaffold size of 30,325 bp. The N50 scaffold length and L50 scaffold size are 48,505 bp and 15,677, respectively. The scaffold nucleotide content is 30.77%, 19.22%, 19.18%, and 30.83% for A, C, G, and T, respectively.

An independent *de novo* assembly of repetitive DNA elements with REPdenovo resulted in 92,505 individual sequences with a total length of 134,023,983 bp. The average length and N50 of the repetitive DNA segments assembled this way were 1,448.83 bp and 2,838 bp (14,228 sequences), respectively. The sequences assembled with REPdenovo were used to aid in our draft genome assembly's repeat identification and masking. A total of 1,410,344,530 bp (52.53%) of the draft assembly were classified as repetitive DNA and masked (see the summary of results in Table 3). Most DNA repeats (42.91% of the repetitive DNA sequence length) were classified as interspersed elements. However, a significant number of repeats (49.91%

of the repetitive DNA) were marked as unclassified. The most common (5.27% of the repetitive DNA) transposable element (TE) in the classification of repeats corresponds to long interspersed nuclear elements (LINEs). Short interspersed nuclear elements (SINEs), repetitive DNA elements, and long terminal repeats (LTRs) amounted to 0.11%, 0.73%, and 1.12% of the total sequence length, respectively.

In the resulting GTF files, we found 373,707 exons in 146,833 genes, each corresponding to a different transcript in a total of 53,658 nuclear DNA scaffolds. According to position information, at least 3,394 of those 146,833 genes could represent gene isoforms. These putative isoforms are found in 1,311 scaffolds.

The completeness of the draft genome assembly was also evaluated with BUSCO [19,20], a commonly used tool to assess the representation of marker genes in newly generated genomic and transcriptomic datasets. BUSCO results are summarized in Table 4.

### Notch signaling pathway

To demonstrate the utility of the genome, we performed a case study in which we assessed the genomic representation of the main components of the Notch signaling pathway. This pathway is highly conserved across all multicellular animals and is known to coordinate a multitude of diverse cellular events, including: proliferation, differentiation, cell fate specification, and cell death [18, 21–26]. In the context of echinoderm regeneration, we have recently demonstrated that the proper function of the Notch pathway is crucial for the arm regeneration in *O. brevispinum* [3]. Here, we searched the draft genome for 29 genes involved in the pathway (Fig. 2, Table 5).

All genes were retrieved by BLAST [27] search. In addition, for all selected genes, except *Mesp2*, *Presenilin 1*, and *NACK*, we also recovered the same putative coding regions from exonerate alignments. In addition, in 18 genes, we also identified the expected conserved protein domains. Taken together, the newly assembled draft genome of *O. brevispinum* allowed us to retrieve the sequences of the Notch pathway components that will be subsequently used to design functional genomic studies to further probe into the mechanistic role of the pathway in brittle star regeneration.

### Mitogenome assembly

The *O. brevispinum* mitochondrial genome (mitogenome) is 15,831 bp long and has a GC content of 32.4% (Figure 5). These values are similar to those of the previously published [28] reference mitogenome of another brittle star species *Ophiarachnella gorgonia* (NCBI accession number NC\_046053), which has a length of 15,948 bp and a GC content of 36.7%. Likewise, mitochondrial genome features of *O. brevispinum* showed the same gene order reported for *O. gorgonia*, and their putative control regions are of similar length (488 and 474 bp, respectively).

There are also differences between these two brittle star mitogenomes that are worth noting. For example, the size difference between the mitogenomes of *O. brevispinum* and *O. gorgonia* is mostly due to deletions in non-coding intergenic regions. However, deletions in tRNA, rRNA, and protein-coding genes are also observed. Furthermore, unlike in *O. gorgonia*, the ND4 coding sequence in *O. brevispinum* is complete and does not add 3' adenine residues to the mRNA.

## Discussion

Here, we present a draft genome assembly for the highly regenerative brittle star species *O. brevispinum*. Due to its availability, ease of maintenance, and amenability to experimental manipulations, *O. brevispinum* has become an emerging model organism in echinoderm regenerative biology [3]. We previously performed transcriptome-wide gene expression studies in this species and identified a range of candidate regeneration-associated genes for further experiments. However, without a fully sequenced genome, including non-coding and regulatory regions, it was not previously feasible to delve into the molecular mechanism of regeneration with functional genomics tools for purposes such as reconstructing gene regulatory networks that underlie regenerative events. This draft genome of *O. brevispinum* provides the first of such resource in ongoing and future molecular studies

The new genome has immediately allowed analysis of protein-coding genes. To demonstrate the utility of the genome, we aimed to retrieve 29 select components of the Notch signaling pathway, including the ligands, receptors, transcription factors, regulators, and target genes. All 29 genes of interest were identified in the assembly. The identity and predicted function of the proteins can be inferred by the presence of the conserved domains.

One of the limitations of our new draft genome assembly is its fragmented state. Ideally, the ultimate goal of any genome sequencing and annotation project is to reconstruct continuous chromosome-size sequences with the fully preserved order of the genes and non-coding sequence elements. Like many first-effort sequencing projects, our assembly will require subsequent efforts to reach that level. Even at its current state these data provide a valuable resource for the ongoing and future studies. This research will not be limited to regenerative biology, but can also benefit other areas, such as evolution of the echinoderm body plan, animal phylogeny, and history of gene families, to name a few.

## Conclusion

Here we presented the first draft nuclear genome and a complete mitochondrial genome of the brittle star *Ophioderma brevispinum* (Say, 1825) (Echinodermata: Ophiuroidea: Ophiacanthida: Ophiodermatidae), a rising model for regenerative studies (e.g., [3, 29–31]). The mitochondrial genome of this brittle star has 15,831 bp (with a mean depth of 1,658.7 and GC content of 32.4%) with 13 protein-coding genes, 22 tRNAs, and 2 rRNAs. The draft nuclear DNA assembly has 88,538 scaffolds summing up to 2.7 bp, corresponding to ~93% of the expected haploid genome size independently determined by a densitometry assay. Despite the high degree of fragmentation of the assembly, which is partially caused by a high frequency of repetitive DNA elements (~52.5% of the assembly), we demonstrated the usefulness of these data for biological investigation by identifying 29 key genes of its Notch-signaling, which is essential to tissue regeneration (e.g., [3, 23, 26, 32, 33]). We predict that the resources we are making available in this publication will be fundamental towards assembling its entire genome at the chromosomal level and establishing *O. brevispinum* as a model for studies of regeneration and other fields.

## Methods

### Supporting genomic resources

Comparative genomic analysis between our original data and other echinoderms relied on different genomic resources that are publicly available, including the complete genome of the purple sea urchin *Strongylocentrotus purpuratus* [8,9] and other genomes available before July 15, 2019 (Table 1). These genomes represent 19 species and 17 genera of the classes Asterozoa (orders Forcipulatida and Valvatida), Crinozoa (order Comatulida), Echinozoa (orders Camarodonta and Cidarozoa), Holothurozoa (orders Holothuriida and Synallactida), and Ophiurozoa (order Amphilepidida).

Additional interrogation and exploration of echinoderm genomic data leverage on resources made available through Echinobase ([www.echinobase.org](http://www.echinobase.org)) [34]. Recent reviews of genomic resources for the study of echinoderm development and evolution are available elsewhere [1,35].

### Computational resources

All analysis were performed using the University of North Carolina at Charlotte's computer clusters (Red Hat Enterprise Linux 7.5 with 64 CPUs and 512 GB to 1.5 TB of memory) as well as high-memory machines (Red Hat Enterprise Linux 7.5 with 16 CPUs and 512 GB to 4 TB of memory).

### Animal collection

Adult individuals of the brittle star *O. brevispinum* were obtained from the Marine Biological Laboratory (Woods Hole, MA, USA). Specimens (catalog no. 1970) were received on April 13, 2016. Immediately after delivery, the package was opened and left overnight to slowly allow the seawater to warm up to room temperature. The animals were then kept in aquaria with aerated artificial seawater.

### RNA-Seq

Complete RNA-Seq analysis (from RNA sampling and isolation until sequencing, *de novo* transcriptome assembly, and gene expression analysis) was described in [3]. Results correspond to BioProject number PRJNA596798 and SRA accession number SRP238266, and were deposited to NCBI's Gene Expression Omnibus (GSE142391, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE142391>).

### DNA isolation and evaluation

A total of 100 mg of tissue was collected from the arms of a single non-regenerating adult individual. The tissue samples were washed in filter-sterilized (0.2  $\mu\text{m}$ ) seawater, cut into small pieces with a sterile blade, and put into the lysis buffer. The high molecular weight nuclear genome DNA was then extracted using the Qiagen Mag-Attract HMW kit according to the manufacturer's instructions with the following modifications that were found to increase the yield and the molecular weight of the resulting DNA: 1) the vortexing speed with magnetic speed was reduced from 1,400 rpm to 1,200 rpm; and 2) after the first vortexing step with the magnetic beads the samples were allowed to sit for 5 min at room temperature before being placed into the magnetic rack. The concentration of the extracted DNA was assessed using the

Qubit dsDNA Broad Range Kit (ThermoFisher). The total amount of DNA was determined at  $\sim 50 \mu\text{g}$ . The integrity of the genomic DNA sample was verified by agarose (0.6%) gel electrophoresis (at 2 V/cm, 4 hours).

#### DNA library preparation and sequencing

All DNA sequencing was performed at the David H. Murdock Research Institute (DHMRI; Kannapolis NC, USA). Two different technologies were employed to obtain short and long sequence reads from the high molecular weight genomic DNA (HWM gDNA) sample extracted as described above.

For sequencing on the Illumina HiSeq2500 platform, subsamples of  $\geq 8 \mu\text{g}$  of HWM gDNA were used to produce short sequence reads. Two complementary strategies were used to generate "short" and "long" insert libraries, respectively. First, three PCR-free paired-end read libraries ("short" insert size) with a  $\sim 450$  bp fragment size were constructed using the TruSeq DNA PCR-free library preparation kit (catalog no. FC-121-3001; Illumina, USA). Second, three mate paired-end read libraries ("long" insert size) were generated using the Illumina Nextera Mate Pair Sample Preparation Kit with the insert size of  $\sim 3$  Kbp. The short and long libraries were combined onto their respective pools and sequenced in the Rapid Mode to produce  $2 \times 251$  bp reads.

To generate long sequence reads ( $\sim 10$  Kbp), we used the Pacific Biosciences Single Molecule Real-Time (SMRT) platform [36]. Assuming a low rate of heterozygosity and a genome size of  $\sim 2.9$  Gbp (see "Genome size estimate," below), we aimed to generate a  $\sim 50 \times$  coverage in an effort to improve the assembly by reducing and closing gaps. Over  $10 \mu\text{g}$  of HWM gDNA with an average fragment length of  $\geq 60$  Kbp was used to produce four SMRTbell libraries. The libraries were generated using the SMRTbell Template Prep Kit 1.0 following the PacBio ">20 Kbp Template Preparation Using BluePippin Size-Selection System (15-20 Kbp) for Sequel Systems" procedures and checklist (catalog no. 100-286-000-07; Pacific Biosciences, USA). Libraries

#### DNA assembly and descriptive statistics

We performed several quality preprocessing steps on the raw DNA sequence reads before the assembly. The overall quality of sequence data before and after each step was determined by FastQC. See the supplementary information for additional details on the DNA assembly, including the commands used to execute the computational analyses described below (Additional file 1: File S1).

The Illumina reads from the short insert library were processed with Trimmomatic v0.38 [37] to remove the adapters and low-quality bases at the ends of each read. Trimmomatic also scanned each read with a 4-base sliding window, cutting when the average quality per base dropped below 15. Reads shorter than 36 bp were discarded. The long insert library Illumina reads were processed with NxTrim v0.3.0-alpha [38] using default parameters to separate reads into four different categories according to the adapter position: mate pairs, unknown (which are mostly mate pairs), paired end, and single end sequence reads.

All the sequence files produced by Trimmomatic and NxTrim were then evaluated with the HTQC toolkit v0.90.8 [39] to produce quality stats per file (using `ht-stat`)

and perform the final read trimming and filtering (with `ht-trim` and `ht-filter`, respectively). Finally, FastUniq v1.1 [40] was used to remove duplicates introduced by PCR amplification from paired short reads.

All cleaned Illumina reads were used as input for the *de novo* assembly with ABySS v2.11 [41] using *k*-mers ranging from 23 to 61 (with a steps of 2). The individual assemblies generated at each *k*-mer value were ranked using several metrics, including the number of sequences, total assembly length, L50, and N50 [42]. We then polished the resulting best assembly in Pilon v1.2.3 [43] to improve base calling and detect sequence variation.

The long PacBio sequence reads we assembled with MaSuRCA v3.2.7 [44]. The contigs generated this way were polished using Arrow v2.3.3 [45] and merged with the ABySS assembly using quickmerge v3be7287 [46].

Assembly stats were calculated using the “assembly-stats” [47] (developed at the Wellcome Sanger Institute) and “assemblathon\_stats.pl” [48] (developed at the UC Davis Bioinformatics Core) tools. The completeness of protein-coding gene representation in the transcriptome was assessed with BUSCO v4.0.6 [19] run in the “genome mode” against the evolutionary conserved metazoan gene set (metazoa\_odb10, creation date: 2021-02-17, number of species: 65, number of BUSCOs: 954) and the conserved eukaryota gene set (eukaryota\_odb10, creation date: 2020-09-10, number of species: 70, number of BUSCOs: 255). Since *O. brevispinum* is not listed among the available species available for Augustus training, we tested other three species: *Homo sapiens*, *Drosophila melanogaster*, and *S. purpuratus*.

#### Genome size estimate

The Animal Genome Size database has only one entry for species of *Ophioderma*, *O. panamensis* [49]. That entry indicates that the expected *C*-value for *O. panamensis* is ~3.3 pg (~3.23 Gbp) based on bulk fluorometric assay [50]. The genome size of *O. brevispinum* have never been determined before. We, therefore, estimated it using two complementary approaches using Feulgen densitometry (FD) assay [51–53] and also from the sequence data.

For the FD assay, soft uncalcified tissues (stomach wall and podia) from a single individual were finely minced with a razor blade, fixed in methanol:acetic acid (3:1) for 10 min and squashed in a drop of 45% acetic acid on a gelatin-coated slide. The samples were then air-dried and post-fixed in methanol:formalin:acetic acid (85:10:15) for 24 hours. After rinsing in tap and distilled water, the samples were hydrolyzed in 5N HCl and stained for 2 hours in Schiff reagent. After brief washes in a 0.5% sodium metabisulfite solution and then in water, the slides were dehydrated in an ethanol series, air dried and mounted in the immersion oil. Microscopic images were then taken at a consistent light intensity in the green monochromatic channel. The optical density of the stained nuclei was measured in the Fiji/ImageJ software [54]. To convert the optical density relative units to the absolute values of DNA mass per nucleus, the following control samples with known DNA content were processed and quantified along with the *O. brevispinum* specimens: chicken erythrocyte nuclei, trout erythrocyte nuclei, triploid trout nuclei, and human (male) cheek epithelial cell nuclei.

In addition to the FD assay, we also estimated the haploid genome size of *O. brevispinum* from the paired-end sequence data in Jellyfish v2.2.4 [55] using a *k*-mer-based statistical approach. The histograms produced by Jellyfish were used to estimate the haploid genome size in GenomeScope [56].

Detailed protocols for genome size estimation based on FD and on *k*-mer-based statistics are provided in Additional file 2: File S2. The estimated genome size was evaluated considering the variation in haploid genome size among echinoderms (see Additional file 3: File S3).

#### Assembly and classification of repetitive DNA elements

Previous to this study, the reports of observed haploid genome size in echinoderms varied over 8-fold. Haploid genome size ranged from 0.53 Gbp in the sea star *Dermasterias imbricata* to 4.3 Gbp in the sea cucumber *Thyonella gemmata* [49]. The largest haploid genomes in the subphylum Asterozoa belong to the order Ophiurida, *Ophioderma panamensis*, with 3.3 Gbp [49].

In addition to the whole-genome assembly described above, we also performed a stand-alone *de novo* assembly of repetitive DNA elements in the genome of *O. brevispinum* with REPdenovo v0.0 [57] following the protocol described in the supplementary information (Additional file 4: File S4).

In short, we used REPdenovo to assemble repeats directly from the cleaned paired-end and single-end short sequence reads that resulted from the quality control steps described above, using different *k*-mer sizes ranging from 25 to 50 with a step of 2.

The contigs assembled with REPdenovo were used as input to RepeatModeler v1.0.11 [58] to build a library of repetitive genomic elements in the genome of *O. brevispinum*. This resulting brittle star repeat library was then combined with repeat libraries from the 2018 version of Repbase [59–62] and RepeatMasker v4.0.8 [63]. Only unique entries were kept to generate a final custom repeat library. This custom repeat library was then used to screen the draft genome of *O. brevispinum* with RepeatMasker to identify interspersed repeats and low complexity DNA sequences. Finally, the RepeatMasker output was manually curated and written into a General Feature Format version 3 (GFF3) file [64]. The details of the repeat library preparation and repeat masking are provided in Additional file 5: File S5.

#### Gene prediction and annotation

The gene prediction and annotation workflow is summarised in Fig. 4. We also provide template scripts listing the parameters used to execute each program listed below in Additional file 6: File S6.

Full gene structure annotations were generated with BRAKER v2.1.2 [65, 66], which integrates GeneMark-ET/EP+ v4.38 [67] and AUGUSTUS v.3.3.2 [68, 69] and allows for fully automated training from RNA-Seq or protein homology information. We also conducted an independent run with AUGUSTUS on selected scaffolds.

The BRAKER annotation pipeline used the genome of the purple sea urchin *S. purpuratus* (assembly Spur\_5.0) as a reference [8, 70] and also the *de novo* assembled transcriptome of *O. brevispinum* [3].

The predicted gene models were aligned with BLAST v.2.9.0+ [27] against the following databases (each downloaded on October 25, 2019): the UniProt Archive

(UniParc; <https://www.uniprot.org/help/uniparc>), the NCBI's non-redundant nucleotide database ("nt"; <https://ftp.ncbi.nlm.nih.gov/>), and the complete Echinodb database of protein coding genes (<https://echinodb.uncc.edu/>).

In addition, the genomic scaffolds were also aligned to the transcriptome of *O. brevispinum* and the cDNA sequences from *S. purpuratus* (assembly Spur\_5.0) using exonerate v2.4.0ls [71], GMAP v2021.03.08 [72], and BLAT v36x2 [73].

#### Annotation of genes associated with the Notch signaling pathway

As a case study, to demonstrate the practical utility of our draft genome assembly, we annotated selected core components and modifiers of the Notch signaling pathway (Fig. 2) using reference sequences from the UniProt and Echinobase ([www.echinobase.org](http://www.echinobase.org)) [34] databases. The sequences from this query reference database were aligned to target exons from the BRAKER annotation using BLAST with the E-value, bit score, and percentage identity cutoff thresholds of  $1.0E-5$ , 30.0, and 23%, respectively. In parallel, we also aligned the query sequences to all assembled scaffolds using exonerate to test if its exon predictions match BLAST results.

We used NCBI's Conserved Domain Search ([www.ncbi.nlm.nih.gov/Structure/cdd](http://www.ncbi.nlm.nih.gov/Structure/cdd)) to identify conserved protein domains in the brittle star Notch pathway genes returned by BLAST and/or exonerate. The conserved domains were searched against the CDD v3.19 database, with an E-value threshold of 0.01 and compositional-based statistics adjustment. We stored the best 500 hits for each gene sequence and then manually inspected the output for the presence or absence of diagnostic domains.

The complete list of genes related to the Notch signaling pathway we searched is provided in the Results section.

#### Mitochondrial genome

The mitochondrial genome (mitogenome) of *O. brevispinum* was contained in a single scaffold generated during the whole-genome assembly. It was identified via sequence alignments using BLAST v.2.9.0+ [27] and a reference sequence from an ophiuroid of the same family (NCBI's accession number NC\_046053.1), *Ophiarachnella gorgonia* (Müller & Troschel, 1842) (Echinodermata: Ophiuroidea: Ophiacanthida: Ophiodermatidae).

The putative circular sequence was extracted from the selected scaffold using AWA (available from <https://gitlab.com/MachadoDJ/awa>; accessed on July 22, 2021) [74]. Next, we remapped filtered short paired-end reads back to AWA's putative mitogenome using Bowtie2 to review base calling. Finally, we used MITOS Web-Server (version 2; available from <http://mitos2.bioinf.uni-leipzig.de/index.py>) [75] to predict genes and an independent analysis with tRNAscan-SE 2.0 [76, 77] to confirm the annotation of tRNAs.

#### Abbreviations

BAM: binary alignment map; BLAST: Basic Local Alignment Search Tool; bp: base pair; dsDNA: Double stranded DNA; EBI: European Bioinformatics Institute; EMBL: European Molecular Biology Laboratory; EMBOSS: European Molecular Biology Open Software Suite; ENCODE: Encyclopedia of DNA Elements; FASTA: it is pronounced "fast A" and stands for "Fast-All"; FASTP: it is pronounced "fast P" and stands for "Fast-Protein"; FD: Feulgen densitometry; FDR: false discovery rate; FPKM: fragments per kilobase of exon per million fragments mapped; GB: gigabyte (approx. 1024 MB); Gb: same as Gbp; Gbp: giga base pairs (1,000,000,000 bp); gDNA: genomic DNA;

GFF3: general feature format or gene-finding format, version 3; GO: gene ontology; HTS: high-throughput sequencing; HMW: high molecular weight; InDel: insertion or deletion; Kb: same as Kbp; Kbp: kilo base pairs (1,000 bp); L50: If we sort sequences by size and sum their sizes in succession from the shortest sequence, the L50 will be the number of sequences needed to achieve 50% of the total size; MB: a unit of information equal to  $2^{20}$  bytes or, loosely, one million bytes; Mb: same as Mbp; Mbp: mega base pairs (1,000,000 bp); mRNA: messenger RNA; MSA: multiple sequence alignment; mtDNA: mitochondrial DNA; nt: nucleotide; nucDNA: nuclear DNA; N50: If we sort sequences by size and sum their sizes in succession from the shortest sequence, the N50 will be the last sequenced added to achieve 50% of the total size; ORFs: open reading frames; PacBio: Pacific Biosciences; PE: paired-end (sequence of both ends of a fragment); rRNA: ribosomal RNA; SMRT: Single Molecule Real-Time; SAM: sequence alignment/map; SE: standard errors; SR: single-read (sequencing from only one end); ssDNA: single stranded DNA tRNA: transfer RNA; VCF: variant call file.

#### Acknowledgements

No comments.

#### Competing interests

The authors declare that they have no competing interests.

#### Author's contributions

VM and DJM share first authorship. VM, RR, CB, and DAJ: initial conceptualization and funding acquisition. VM, DAJ, RR, DJM, and JK: writing (original draft, review, and editing). RR and DJM: methodology, formal analysis, investigation, resources, software, validation, data curation, and visualization. DJ, VM, CB, DJM and JK: project administration.

#### Funding

Research reported in this publication was supported by the National Institute for General Medical Sciences of the National Institutes of Health under award number 1R15GM128066-01. We also acknowledge funding from the University of North Florida. We acknowledge funding and logistical support from several entities of the University of North Carolina at Charlotte including: The Bioinformatics Services Division, the Department of Bioinformatics and Genomics, the Bioinformatics Research Center, University Research Computing, the College of Computing and Informatics. We are grateful for funding from the Belk Family.

#### Availability of data and materials

The data sets supporting the conclusions of this article are included within the article and its additional files. Supplementary information accompanies this paper is available at Zenodo, DOI: 10.5281/zenodo.4247706. Data corresponding to our draft genome assembly of *O. brevispina* can be found at NCBI's databases under BioProject number PRJNA779014 and BioSample number SAMN23008116. The NCBI's submission identifier for this draft genome is SUB10570566. The mitochondrial genome sequence and annotations have been submitted to NCBI Bankit (accession number pending NCBI approval).

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Author details

<sup>1</sup>Wake Forest Institute for Regenerative Medicine, 391 Technology Way, 27101 Winston-Salem, NC, USA.

<sup>2</sup>University of North Florida, Department of Biology, university, etc 1 UNF Drive, 3224 Jacksonville, FL, USA.

<sup>3</sup>University of North Carolina at Charlotte, College of Computing and Informatics, Department of Bioinformatics and Genomics, 9201 University City Blvd, 28223 Charlotte, NC, USA. <sup>4</sup>University of North Carolina at Charlotte, College of Computing and Informatics, North Carolina Research Campus, 150 Research Campus Drive, 28081 Kannapolis, NC, USA.

#### References

- Cary GA, Cameron RA, Hinman VF. Genomic resources for the study of echinoderm development and evolution. In: Methods Cell Biol. vol. 151. New York, NY: Elsevier; 2019. p. 65–88.
- Carnevali MC. Regeneration in Echinoderms: repair, regrowth, cloning. *Invert Surviv J.* 2006;3(1):64–76.
- Mashanov V, Akiona J, Khoury M, Ferrier J, Reid R, Machado DJ, et al. Active Notch signaling is required for arm regeneration in a brittle star. *PLoS ONE.* 2020;15(5):e0232981. Available from: <https://doi.org/10.1371/journal.pone.0232981>.
- Mashanov VS, Zueva OR, García-Arrarás JE. Transcriptomic changes during regeneration of the central nervous system in an echinoderm. *BMC Genomics.* 2014;15(1):1–21. Available from: <https://doi.org/10.1186/1471-2164-15-357>.
- Quispe-Parra DJ, Medina-Feliciano JG, Cruz-González S, Ortiz-Zuazaga H, García-Arrarás JE. Transcriptomic analysis of early stages of intestinal regeneration in *Holothuria glaberrima*. *Sci Rep.* 2021;11(1):1–14. Available from: <https://doi.org/10.1038/s41598-020-79436-2>.
- Purushothaman S, Saxena S, Meghah V, Swamy CVB, Ortega-Martinez O, Dupont S, et al. Transcriptomic and proteomic analyses of *Amphiura filiformis* arm tissue-undergoing regeneration. *J Proteomics.* 2015;112:113–124. Available from: <https://doi.org/10.1016/j.jprot.2014.08.011>.
- Czarkwiani A, Dylus DV, Carballo L, Oliveri P. FGF signalling plays similar roles in development and regeneration of the skeleton in the brittle star *Amphiura filiformis*. *Development.* 2021;148(10):dev180760. Available from: <https://doi.org/10.1242/dev.180760>.

8. Sodergren E, Weinstock GM, Davidson EH, Cameron RA, Gibbs RA, Angerer RC, et al. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science*. 2006;314(5801):941–952. Available from: <https://doi.org/10.1126/science.1133609>.
9. Cameron RA, Kudtarkar P, Gordon SM, Worley KC, Gibbs RA. Do echinoderm genomes measure up? *Mar Genom*. 2015;22:1–9. Available from: <https://doi.org/10.1016/j.margen.2015.02.004>.
10. Kinjo S, Kiyomoto M, Yamamoto T, Ikeo K, Yaguchi S. HpBase: A genome database of a sea urchin, *Hemicentrotus pulcherrimus*. *Dev Growth Differ*. 2018;60(3):174–182. Available from: <https://doi.org/10.1111/dgd.12429>.
11. Sergiev PV, Artemov AA, Prokhortchouk EB, Dontsova OA, Berezkin GV. Genomes of *Strongylocentrotus franciscanus* and *Lytechinus variegatus*: are there any genomic explanations for the two order of magnitude difference in the lifespan of sea urchins? *Aging*. 2016;8(2):260. Available from: <https://doi.org/10.1371/10.18632/aging.100889>.
12. Kudtarkar P, Cameron RA. Echinobase: an expanding resource for echinoderm genomic information. *Database*. 2017;2017. Available from: <https://doi.org/10.1093/database/bax074>.
13. Zhang X, Sun L, Yuan J, Sun Y, Gao Y, Zhang L, et al. The sea cucumber genome provides insights into morphological evolution and visceral regeneration. *PLoS Biol*. 2017;15(10):e2003790. Available from: <https://doi.org/10.1371/journal.pbio.2003790>.
14. Hall MR, Kocot KM, Baughman KW, Fernandez-Valverde SL, Gauthier ME, Hatleberg WL, et al. The crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest. *Nature*. 2017;544(7649):231–234. Available from: <https://doi.org/10.1038/nature22033>.
15. Long KA, Nossa CW, Sewell MA, Putnam NH, Ryan JF. Low coverage sequencing of three echinoderm genomes: the brittle star *Ophionereis fasciata*, the sea star *Patiriella regularis*, and the sea cucumber *Australostichopus mollis*. *GigaScience*. 2016;5(1):s13742–016. Available from: <https://doi.org/10.1186/s13742-016-0125-6>.
16. Say T. On the species of the Linnean genus *Asterias* inhabiting the coast of the United States. *P Acad Nat Sci Phila*. 1825;5(1):151–154.
17. Bowmer T, Keegan B. Field survey of the occurrence and significance of regeneration in *Amphiura filiformis* (Echinodermata: Ophiuroidea) from Galway Bay, west coast of Ireland. *Mar Biol*. 1983;74(1):65–71. Available from: <https://doi.org/10.1007/BF00394276>.
18. Hurlbut GD, Kankel MW, Lake RJ, Artavanis-Tsakonas S. Crossing paths with Notch in the hyper-network. *Curr Opin Cell Biol*. 2007;19(2):166–175. Available from: <https://doi.org/10.1016/j.ceb.2007.02.012>.
19. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210–3212. Available from: <https://doi.org/10.1093/bioinformatics/btv351>.
20. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*. 2018;35(3):543–548. Available from: <https://doi.org/10.1093/molbev/msx319>.
21. Walton KD, Croce JC, Glenn TD, Wu SY, McClay DR. Genomics and expression profiles of the Hedgehog and Notch signaling pathways in sea urchin development. *Dev Biol*. 2006;300(1):153–164.
22. Marlow H, Roettinger E, Boekhout M, Martindale MQ. Functional roles of Notch signaling in the cnidarian *Nematostella vectensis*. *Dev Biol*. 2012;362(2):295–308.
23. Layden MJ, Martindale MQ. Non-canonical Notch signaling represents an ancestral mechanism to regulate neural differentiation. *EvoDevo*. 2014;5(1):1–14.
24. Erkenbrack EM. Notch-mediated lateral inhibition is an evolutionarily conserved mechanism patterning the ectoderm in echinoids. *Dev Genes Evol*. 2018;228(1):1–11.
25. Favarolo MB, López SL. Notch signaling in the division of germ layers in bilaterian embryos. *Mech Develop*. 2018;154:122–144.
26. Lloyd-Lewis B, Mourikis P, Fre S. Notch signalling: sensor and instructor of the microenvironment to coordinate cell fate and organ morphogenesis. *Curr Opin Cell Biol*. 2019;61:16–23.
27. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–410. Available from: [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
28. Lee T, Bae YJ, Shin S. Mitochondrial gene rearrangement and phylogenetic relationships in the Amphilepidida and Ophiacanthida (Echinodermata, Ophiuroidea). *Mar Biol Res*. 2019;15(1):26–35.
29. Zueva O, Khoury M, Heinzeller T, Mashanova D, Mashanov V. The complex simplicity of the brittle star nervous system. *Front in Zool*. 2018;15(1):1–26. Available from: <https://doi.org/10.1186/s12983-017-0247-4>.
30. Clark EG, Fezzaa K, Burke J, Racicot R, Shaw J, Westacott S, et al. A farewell to arms: using X-ray synchrotron imaging to investigate autotomy in brittle stars. *Zoomorphology*. 2019;138(3):419–424. Available from: <https://doi.org/10.1007/s00435-019-00451-7>.
31. Mashanov V, Zueva O. Radial glia in echinoderms. *Dev Neurobiol*. 2019;79(5):396–405. Available from: <https://doi.org/10.1002/dneu.22659>.
32. Ehebauer M, Hayward P, Martinez-Arias A. Notch signaling pathway. *Science's STKE*. 2006;2006(364):cm7. Available from: <https://doi.org/10.1126/stke.3642006cm7>.
33. Cormier S, Le Bras S, Souilhol C, Vandormael-Pournin S, Durand B, Babinet C, et al. The murine ortholog of notchless, a direct regulator of the notch pathway in *Drosophila melanogaster*, is essential for survival of inner cell mass cells. *Mol Cell Biol*. 2006;26(9):3541–3549. Available from: <https://doi.org/10.1128/MCB.26.9.3541-3549.2006>.
34. Cary GA, Cameron RA, Hinman VF. EchinoBase: tools for echinoderm genome analyses. In: Kollmar M, editor. *Eukaryotic genomic databases*. vol. 1757 of *Methods Mol Biol*; 2018. p. 349–369.
35. Kondo M, Akasaka K. Current status of echinoderm genome analysis - what do we know? *Curr Genomics*. 2012;13(2):134–143. Available from: <https://doi.org/10.2174/138920212799860643>.
36. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase

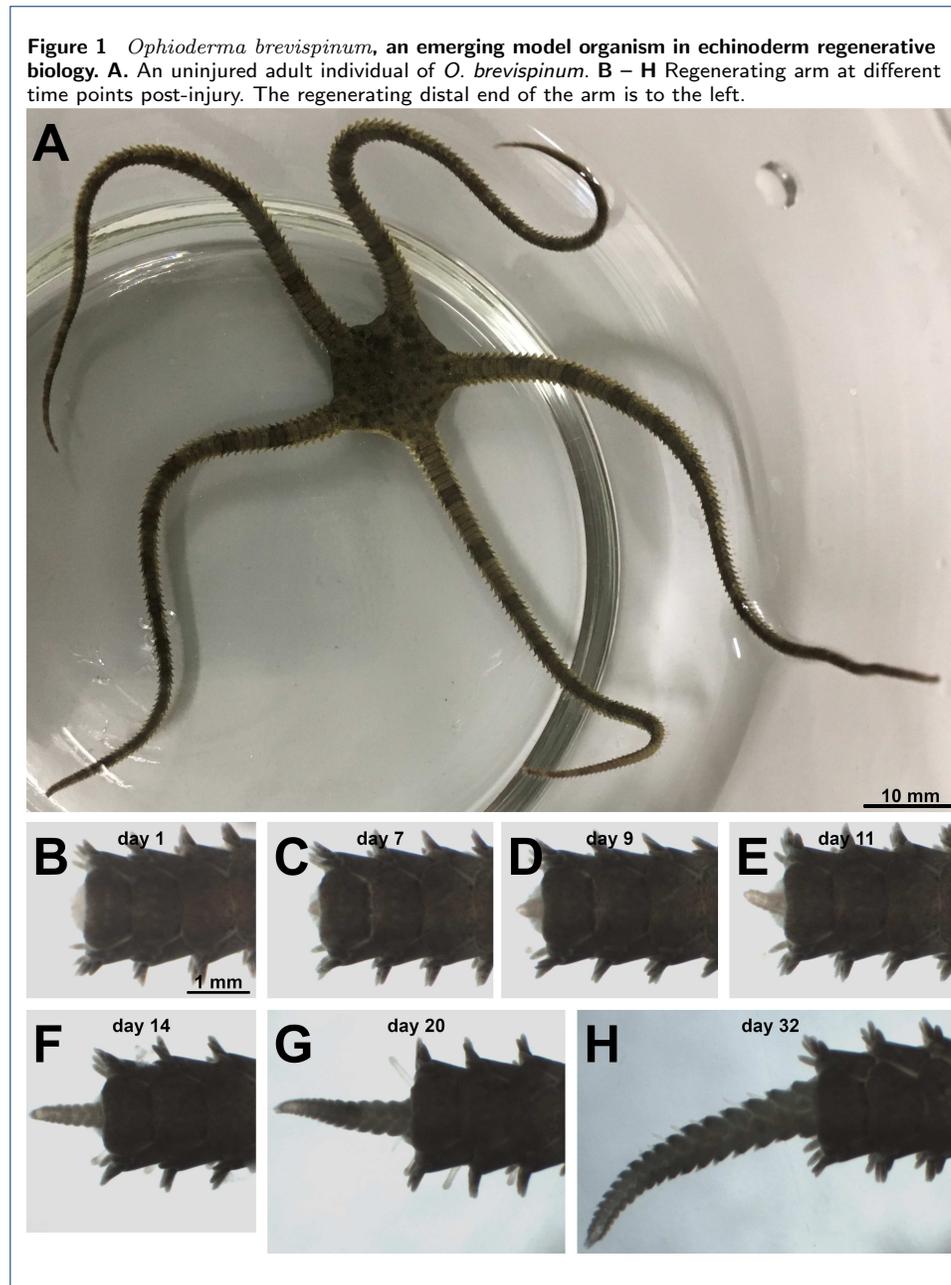
- molecules. *Science*. 2009;323(5910):133–138. Available from: <https://doi.org/10.1126/science.1162986>.
37. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–2120. Available from: <https://doi.org/10.1093/bioinformatics/btu170>.
  38. O'Connell J, Schulz-Trieglaff O, Carlson E, Hims MM, Gormley NA, Cox AJ. NxTrim: optimized trimming of Illumina mate pair reads. *Bioinformatics*. 2015;31(12):2035–2037. Available from: <https://doi.org/10.1093/bioinformatics/btv057>.
  39. Yang X, Liu D, Liu F, Wu J, Zou J, Xiao X, et al. HTQC: a fast quality control toolkit for Illumina sequencing data. *BMC Bioinformatics*. 2013;14(1):1–4. Available from: <https://doi.org/10.1186/1471-2105-14-33>.
  40. Xu H, Luo X, Qian J, Pang X, Song J, Qian G, et al. FastUniq: a fast *de novo* duplicates removal tool for paired short reads. *PLoS ONE*. 2012;7(12):e52249. Available from: <https://doi.org/10.1371/journal.pone.0052249>.
  41. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. ABySS: a parallel assembler for short read sequence data. *Genome Res*. 2009;19(6):1117–1123. Available from: <http://www.genome.org/cgi/doi/10.1101/gr.089532.108>.
  42. Thrash A, Hoffmann F, Perkins A. Toward a more holistic method of genome assembly assessment. *BMC Bioinformatics*. 2020;21(4):1–8. Available from: <https://doi.org/10.1186/s12859-020-3382-4>.
  43. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*. 2014;9(11):e112963. Available from: <https://doi.org/10.1371/journal.pone.0112963>.
  44. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. *Bioinformatics*. 2013;29(21):2669–2677. Available from: <https://doi.org/10.1093/bioinformatics/btt476>.
  45. Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science*. 2002;298(5591):129–149. Available from: <https://doi.org/10.1126/science.1076181>.
  46. Chakraborty M, Baldwin-Brown JG, Long AD, Emerson J. Contiguous and accurate *de novo* assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res*. 2016;44(19):e147–e147. Available from: <https://doi.org/10.1093/nar/gkw654>.
  47. Hunt M. assembly-stats; 2014–2021. Accessed 18 Feb 2019. Available from: <https://github.com/sanger-pathogens/assembly-stats>.
  48. UC Davis Bioinformatics Core. assemblathon2-analysis; 2012. Accessed 24 Mar 2019. Available from: <https://github.com/ucdavis-bioinformatics/assemblathon2-analysis>.
  49. Gregory TR. Animal genome size database.; 2020. Accessed 28 Oct 2020. Available from: <http://www.genomesize.com>.
  50. Hinegardner R. Cellular DNA content of the Echinodermata. *Comp Biochem Physiol*. 1974;49B:219–226.
  51. Hardie DC, Gregory TR, Hebert PD. From pixels to picograms: a beginners' guide to genome quantification by Feulgen image analysis densitometry. *J Histochem Cytochem*. 2002;50(6):735–749. Available from: <https://doi.org/10.1177/002215540205000601>.
  52. Rasch EM, Lee CE, Wyngaard GA. DNA–Feulgen cytophotometric determination of genome size for the freshwater-invading copepod *Eurytemora affinis*. *Genome*. 2004;47(3):559–564. Available from: <https://doi.org/10.1139/g04-014>.
  53. Donnenberg VS, Landreneau RJ, Pfeifer ME, Donnenberg AD. Flow cytometric determination of stem/progenitor content in epithelial tissues: an example from nonsmall lung cancer and normal lung. *Cytometry Part A*. 2013;83(1):141–149. Available from: <https://doi.org/10.1002/cyto.a.22156>.
  54. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9(7):676–682.
  55. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics*. 2011;27(6):764–770. Available from: <https://doi.org/10.1093/bioinformatics/btr011>.
  56. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*. 2017;33(14):2202–2204. Available from: <https://doi.org/10.1093/bioinformatics/btx153>.
  57. Chu C, Nielsen R, Wu Y. REPdenovo: inferring *de novo* repeat motifs from short sequence reads. *PLoS ONE*. 2016;11(3):e0150719. Available from: <https://doi.org/10.1371/journal.pone.0150719>.
  58. Smit AFA, Hubley R. RepeatModeler Open-1.0; 2008. Accessed 10 Nov 2020. Available from: <http://www.repeatmasker.org>.
  59. Jurka J. Repeats in genomic DNA: mining and meaning. *Curr Opin Struct Biol*. 1998;8(3):333–337. Available from: [https://doi.org/10.1016/S0959-440X\(98\)80067-5](https://doi.org/10.1016/S0959-440X(98)80067-5).
  60. Jurka J. Repbase update: a database and an electronic journal of repetitive elements. *Trends in Genetics*. 2000;16(9):418–420. Available from: [https://doi.org/10.1016/S0168-9525\(00\)02093-X](https://doi.org/10.1016/S0168-9525(00)02093-X).
  61. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res*. 2005;110(1-4):462–467. Available from: <https://doi.org/10.1159/000084979>.
  62. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*. 2015;6(1):11. Available from: <https://doi.org/10.1186/s13100-015-0041-9>.
  63. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0; 2013–2015. Accessed 10 Nov 2020. Available from: <http://www.repeatmasker.org>.
  64. Reese MG, Moore B, Batchelor C, Salas F, Cunningham F, Marth GT, et al. A standard variation file format for human genome sequences. *Genome Biol*. 2010;11(8):R88. Available from: <https://doi.org/10.1186/gb-2010-11-8-r88>.
  65. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*. 2016;32(5):767–769. Available from: <https://doi.org/10.1093/bioinformatics/btv661>.

66. Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*. 2021;3(1):lqaa108. Available from: <https://doi.org/10.1093/nargab/lqaa108>.
67. Lomsadze A, Burns PD, Borodovsky M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res*. 2014;42(15):e119–e119. Available from: <https://doi.org/10.1093/nar/gku557>.
68. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics*. 2008;24(5):637–644. Available from: <https://doi.org/10.1093/bioinformatics/btn013>.
69. Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics*. 2006;7(1):1–11. Available from: <https://doi.org/10.1186/1471-2105-7-62>.
70. Ensembl Metazoa Home. *Strongylocentrotus purpuratus* (Spur 01) (Spur.5.0); 2020. Accessed 12 Aug 2020. Available from: [https://metazoa.ensembl.org/Strongylocentrotus\\_purpuratus](https://metazoa.ensembl.org/Strongylocentrotus_purpuratus).
71. Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*. 2005;6(1):1–11. Available from: <http://www.biomedcentral.com/1471-2105/6/31>.
72. Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*. 2005;21(9):1859–1875. Available from: <https://doi.org/10.1093/bioinformatics/bt1310>.
73. Kent WJ. BLAT—the BLAST-like alignment tool. *Genome Res*. 2002;12(4):656–664. Available from: <https://doi.org/10.1101/gr.229202>.
74. Jacob Machado D, Janies D, Brouwer C, Grant T. A new strategy to infer circularity applied to four new complete frog mitogenomes. *Ecol Evol*. 2018;8(8):4011–4018. Available from: <https://doi.org/10.1002/ece3.3918>.
75. Donath A, Jühling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF, et al. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic Acids Res*. 2019;47(20):10543–10552. Available from: <https://doi.org/10.1093/nar/gkz833>.
76. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997;25(5):955–964. Available from: <https://doi.org/10.1093/nar/25.5.955>.
77. Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. In: *Gene prediction*. Springer; 2019. p. 1–14. Available from: [https://doi.org/10.1007/978-1-4939-9173-0\\_1](https://doi.org/10.1007/978-1-4939-9173-0_1).
78. Hennebert E, Leroy B, Wattiez R, Ladurner P. An integrated transcriptomic and proteomic analysis of sea star epidermal secretions identifies proteins involved in defense and adhesion. *J Proteomics*. 2015;128:83–91. Available from: <https://doi.org/10.1016/j.jprot.2015.07.002>.
79. 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(6):1087–1102. Available from: <https://doi.org/10.1111/mec.15386>.
80. Baughman KW, McDougall C, Cummins SF, Hall M, Degnan BM, Satoh N, et al. Genomic organization of Hox and Parahox clusters in the echinoderm, *Acanthaster planci*. *Genesis*. 2014;52(12):952–958. Available from: <https://doi.org/10.1002/dvg.22840>.
81. Yasuda N, Hamaguchi M, Sasaki M, Nagai S, Saba M, Nadaoka K. Complete mitochondrial genome sequences for Crown-of-thorns starfish *Acanthaster planci* and *Acanthaster brevispinus*. *BMC Genomics*. 2006;7(1):1–10. Available from: <https://doi.org/10.1186/1471-2164-7-17>.
82. Jung G, Lee YH. Complete mitochondrial genome of Chilean sea urchin: *Loxechinus albus* (Camarodonta, Parechinidae). *Mitochondrial DNA*. 2015;26(6):883–884. Available from: <https://doi.org/10.3109/19401736.2013.809449>.
83. Humble E, Dobrynin P, Senn H, Chuvén J, Scott AF, Mohr DW, et al. Chromosomal-level genome assembly of the scimitar-horned oryx: Insights into diversity and demography of a species extinct in the wild. *Mol Ecol Resour*. 2020;20(6):1668–1681. Available from: <https://doi.org/10.1093/gbe/evaa101>.
84. Bronstein O, Kroh A. The first mitochondrial genome of the model echinoid *Lytechinus variegatus* and insights into Odontophoran phylogenetics. *Genomics*. 2019;111(4):710–718. Available from: <https://doi.org/10.1016/j.ygeno.2018.04.008>.
85. Morrison AMS, Goldstone JV, Lamb DC, Kubota A, Lemaire B, Stegeman JJ. Identification, modeling and ligand affinity of early deuterostome CYP51s, and functional characterization of recombinant zebrafish sterol 14 $\alpha$ -demethylase. *BBA-Gen Subjects*. 2014;1840(6):1825–1836. Available from: <https://doi.org/10.1016/j.bbagen.2013.12.009>.
86. Tu Q, Cameron RA, Worley KC, Gibbs RA, Davidson EH. Gene structure in the sea urchin *Strongylocentrotus purpuratus* based on transcriptome analysis. *Genome Res*. 2012;22(10):2079–2087. Available from: <https://doi.org/10.1101/gr.139170.112>.
87. Stevens ME, Dhillon J, Miller CA, Messier-Solek C, Majeske AJ, Zuelke D, et al. SpTie1/2 is expressed in coelomocytes, axial organ and embryos of the sea urchin *Strongylocentrotus purpuratus*, and is an orthologue of vertebrate Tie1 and Tie2. *Dev Comp Immunol*. 2010;34(8):884–895. Available from: <https://doi.org/10.1016/j.dci.2010.03.010>.
88. Tartari M, Gissi C, Lo Sardo V, Zuccato C, Picardi E, Pesole G, et al. Phylogenetic comparison of huntingtin homologues reveals the appearance of a primitive polyQ in sea urchin. *Mol Biol Evol*. 2008;25(2):330–338. Available from: <https://doi.org/10.1093/molbev/msm258>.
89. Tu Q, Brown CT, Davidson EH, Oliveri P. Sea urchin Forkhead gene family: phylogeny and embryonic expression. *Dev Biol*. 2006;300(1):49–62. Available from: <https://doi.org/10.1016/j.ydbio.2006.09.031>.
90. Neill AT, Moy GW, Vacquier VD. Polycystin-2 associates with the polycystin-1 homolog, suREJ3, and localizes to the acrosomal region of sea urchin spermatozoa. *Mol Reprod Dev*. 2004;67(4):472–477. Available from: <https://doi.org/10.1002/mrd.20033>.
91. Multerer KA, Smith LC. Two cDNAs from the purple sea urchin, *Strongylocentrotus purpuratus*, encoding

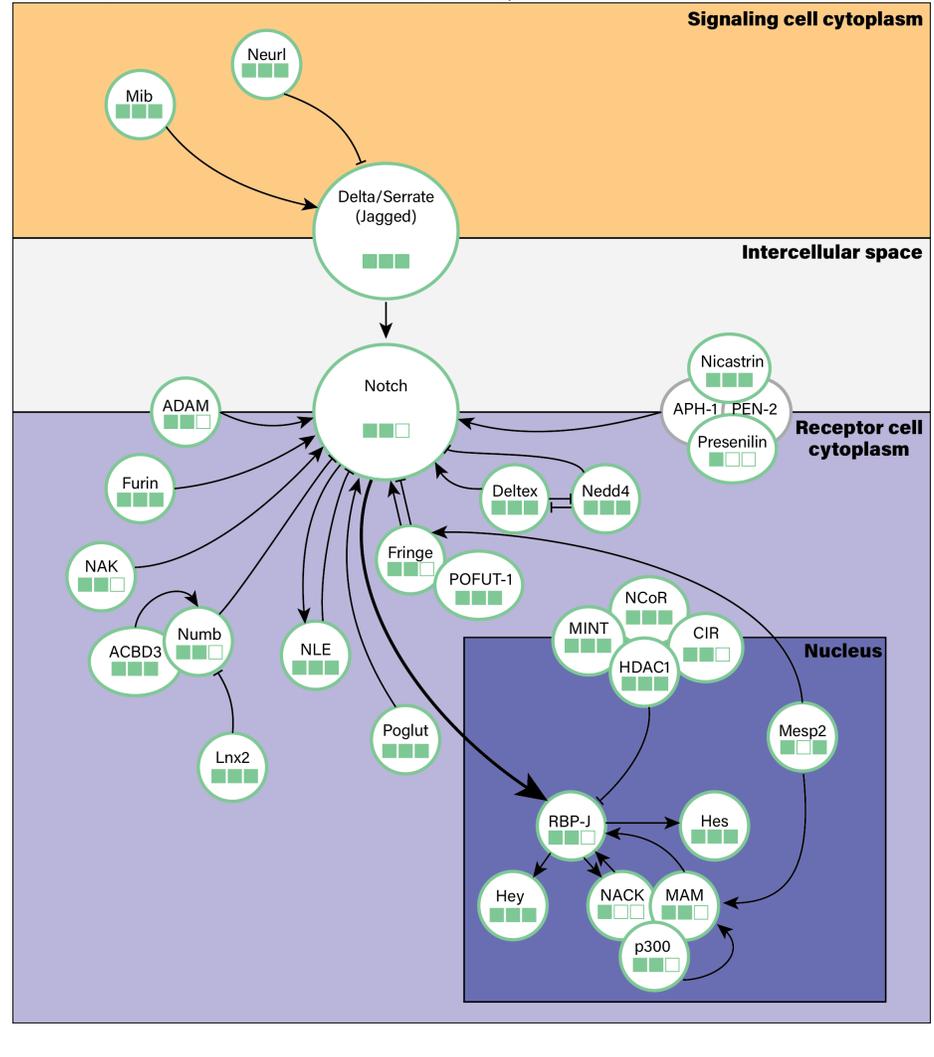
- mosaic proteins with domains found in factor H, factor I, and complement components C6 and C7. *Immunogenetics*. 2004;56(2):89–106. Available from: <https://doi.org/10.1007/s00251-004-0665-2>.
92. Kamei N, Glabe CG. The species-specific egg receptor for sea urchin sperm adhesion is EBR1, a novel ADAMTS protein. *Gene Dev*. 2003;17(20):2502–2507. Available from: <https://doi.org/10.1101/gad.1133003>.
  93. Tombes RM, Faison MO, Turbeville J. Organization and evolution of multifunctional Ca<sup>2+</sup>/CaM-dependent protein kinase genes. *Gene*. 2003;322:17–31. Available from: <https://doi.org/10.1016/j.gene.2003.08.023>.
  94. Sirotkin V, Seipel S, Krendel M, Bonder EM. Characterization of sea urchin unconventional myosins and analysis of their patterns of expression during early embryogenesis. *Mol Reprod Dev*. 2000;57(2):111–126. Available from: [https://doi.org/10.1002/1098-2795\(200010\)57:2<3C111::aid-mrd2%3E3.0.co;2-t](https://doi.org/10.1002/1098-2795(200010)57:2<3C111::aid-mrd2%3E3.0.co;2-t).
  95. Pancer Z. Dynamic expression of multiple scavenger receptor cysteine-rich genes in coelomocytes of the purple sea urchin. *P Natl Acad Sci*. 2000;97(24):13156–13161. Available from: <https://doi.org/10.1073/pnas.230096397>.
  96. Pancer Z, Rast JP, Davidson EH. Origins of immunity: transcription factors and homologues of effector genes of the vertebrate immune system expressed in sea urchin coelomocytes. *Immunogenetics*. 1999;49(9):773–786. Available from: <https://doi.org/10.1007/s002510050551>.
  97. LaFleur Jr GJ, Horiuchi Y, Wessel GM. Sea urchin ovoperoxidase: oocyte-specific member of a heme-dependent peroxidase superfamily that functions in the block to polyspermy. *Mechanisms Devel*. 1998;70(1-2):77–89. Available from: [https://doi.org/10.1016/s0925-4773\(97\)00178-0](https://doi.org/10.1016/s0925-4773(97)00178-0).
  98. Hartman JJ, Mahr J, McNally K, Okawa K, Iwamatsu A, Thomas S, et al. Katanin, a microtubule-severing protein, is a novel AAA ATPase that targets to the centrosome using a WD40-containing subunit. *Cell*. 1998;93(2):277–287. Available from: [https://doi.org/10.1016/s0092-8674\(00\)81578-0](https://doi.org/10.1016/s0092-8674(00)81578-0).
  99. Marsden M, Burke RD. The  $\beta$ L integrin subunit is necessary for gastrulation in sea urchin embryos. *Dev Biol*. 1998;203(1):134–148. Available from: <https://doi.org/10.1006/dbio.1998.9033>.
  100. Marsden M, Burke R. Cloning and characterization of novel  $\beta$  integrin subunits from a sea urchin. *Dev Biol*. 1997;181(2):234–245. Available from: <https://doi.org/10.1006/dbio.1996.8451>.
  101. Valverde JR, Marco R, Garesse R. A conserved heptamer motif for ribosomal RNA transcription termination in animal mitochondria. *P Natl Acad Sci*. 1994;91(12):5368–5371. Available from: <https://doi.org/10.1073/pnas.91.12.5368>.
  102. Qureshi SA, Jacobs HT. Two distinct, sequence-specific DNA-binding proteins interact independently with the major replication pause region of sea urchin mtDNA. *Nucleic Acids Res*. 1993;21(12):2801–2808. Available from: <https://doi.org/10.1093/nar/21.12.2801>.
  103. Kitagawa M. Notch signalling in the nucleus: roles of Mastermind-like (MAML) transcriptional coactivators. *J Biochem*. 2016;159(3):287–294. Available from: <https://doi.org/10.1093/jb/mvv123>.
  104. Jin K, Zhou W, Han X, Wang Z, Li B, Jeffries S, et al. Acetylation of mastermind-like 1 by p300 drives the recruitment of NACK to initiate notch-dependent transcription. *Cancer Res*. 2017;77(16):4228–4237. Available from: <https://doi.org/10.1158/0008-5472.CAN-16-3156>.
  105. Wallberg AE, Pedersen K, Lendahl U, Roeder RG. p300 and PCAF act cooperatively to mediate transcriptional activation from chromatin templates by notch intracellular domains in vitro. *Method Mol Cell Biol*. 2002;22(22):7812–7819. Available from: <https://doi.org/10.1128/MCB.22.22.7812-7819.2002>.
  106. Weaver KL, Alves-Guerra MC, Jin K, Wang Z, Han X, Ranganathan P, et al. NACK is an integral component of the Notch transcriptional activation complex and is critical for development and tumorigenesis. *Cancer Res*. 2014;74(17):4741–4751. Available from: <https://doi.org/10.1158/0008-5472.CAN-14-1547>.
  107. Kim GS, Park HS, Lee YC. OPTHIS identifies the molecular basis of the direct interaction between CSL and SMRT corepressor. *Mol Cells*. 2018;41(9):842. Available from: <https://doi.org/10.14348/molcells.2018.0196>.
  108. Gazave E, Lapébie P, Richards GS, Brunet F, Ereskovsky AV, Degnan BM, et al. Origin and evolution of the Notch signalling pathway: an overview from eukaryotic genomes. *BMC Evol Biol*. 2009;9(1):1–27. Available from: <https://doi.org/10.1186/1471-2148-9-249>.
  109. Shi S, Stanley P. Protein O-fucosyltransferase 1 is an essential component of Notch signaling pathways. *P Natl Acad Sci*. 2003;100(9):5234–5239.
  110. Taylor P, Takeuchi H, Sheppard D, Chillakuri C, Lea SM, Haltiwanger RS, et al. Fringe-mediated extension of O-linked fucose in the ligand-binding region of Notch1 increases binding to mammalian Notch ligands. *P Natl Acad Sci*. 2014;111(20):7290–7295.
  111. Van Tetering G, Vooijs M. Proteolytic cleavage of Notch: “HIT and RUN”. *Curr Mol Med*. 2011;11(4):255–269. Available from: <https://doi.org/10.2174/156652411795677972>.
  112. Koutelou E, Sato S, Tomomori-Sato C, Florens L, Swanson SK, Washburn MP, et al. Neuralized-like 1 (Neur1) targeted to the plasma membrane by N-myristoylation regulates the Notch ligand Jagged1. *J Biol Chem*. 2008;283(7):3846–3853. Available from: <https://doi.org/10.1074/jbc.M706974200>.
  113. Koo BK, Yoon KJ, Yoo KW, Lim HS, Song R, So JH, et al. Mind bomb-2 is an E3 ligase for Notch ligand. *J Biol Chem*. 2005;280(23):22335–22342.
  114. Kopan R, Ilagan MXG. The canonical Notch signaling pathway: unfolding the activation mechanism. *Cell*. 2009;137(2):216–233. Available from: <https://doi.org/10.1016/j.cell.2009.03.045>.
  115. Teider N, Scott DK, Neiss A, Weeraratne SD, Amani VM, Wang Y, et al. Neuralized1 causes apoptosis and downregulates Notch target genes in medulloblastoma. *Neuro Oncol*. 2010;12(12):1244–1256. Available from: <https://doi.org/10.1093/neuonc/noq091>.
  116. Iso T, Kedes L, Hamamori Y. HES and HERP families: multiple effectors of the Notch signaling pathway. *J Cell Physiol*. 2003;194(3):237–255. Available from: <https://doi.org/10.1002/jcp.10208>.
  117. Young PW. LNX1/LNX2 proteins: Functions in neuronal signalling and beyond. *Neuronal Signal*. 2018;2(2):NS20170191.
  118. Zhou Y, Atkins JB, Rompani SB, Bancescu DL, Petersen PH, Tang H, et al. The mammalian Golgi regulates

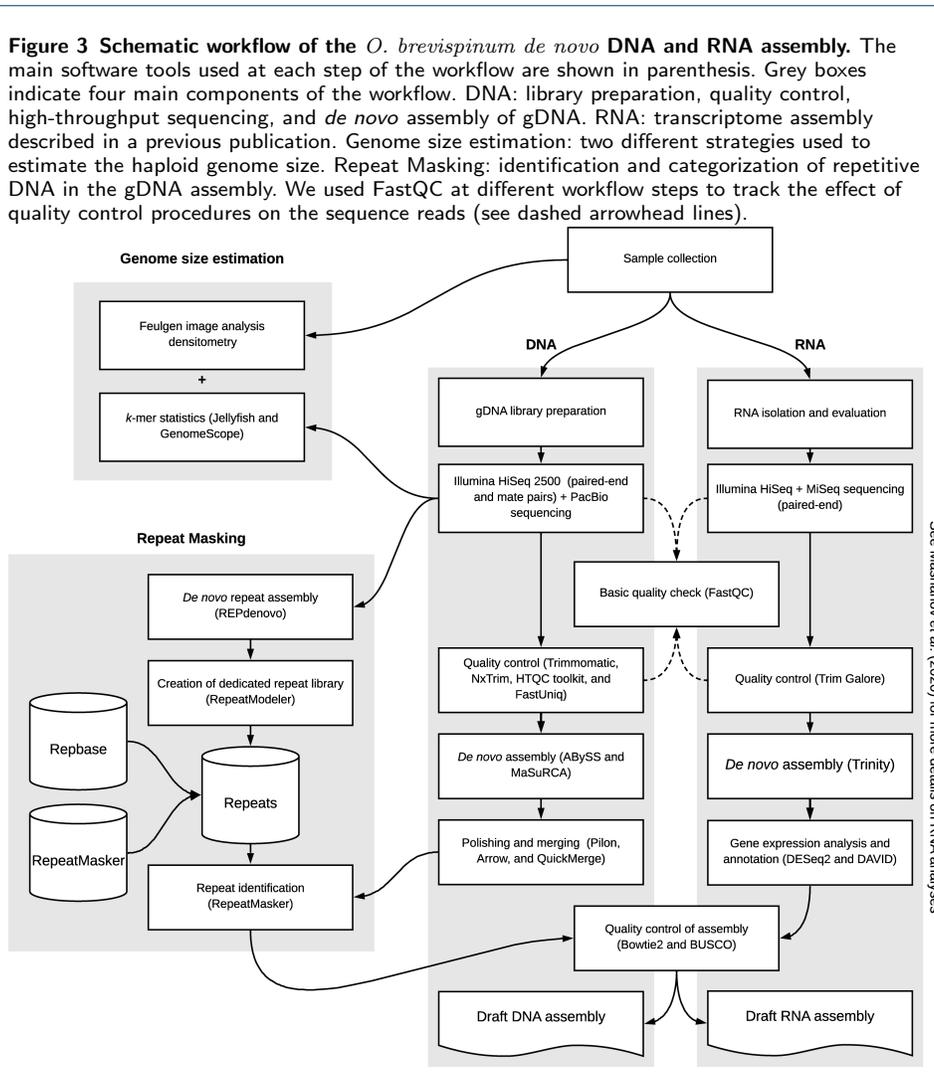
- numb signaling in asymmetric cell division by releasing ACBD3 during mitosis. *Cell*. 2007;129(1):163–178.
119. Sakata T, Sakaguchi H, Tsuda L, Higashitani A, Aigaki T, Matsuno K, et al. *Drosophila* Nedd4 regulates endocytosis of notch and suppresses its ligand-independent activation. *Curr Biol*. 2004;14(24):2228–2236. Available from: <https://doi.org/10.1016/j.cub.2004.12.028>.

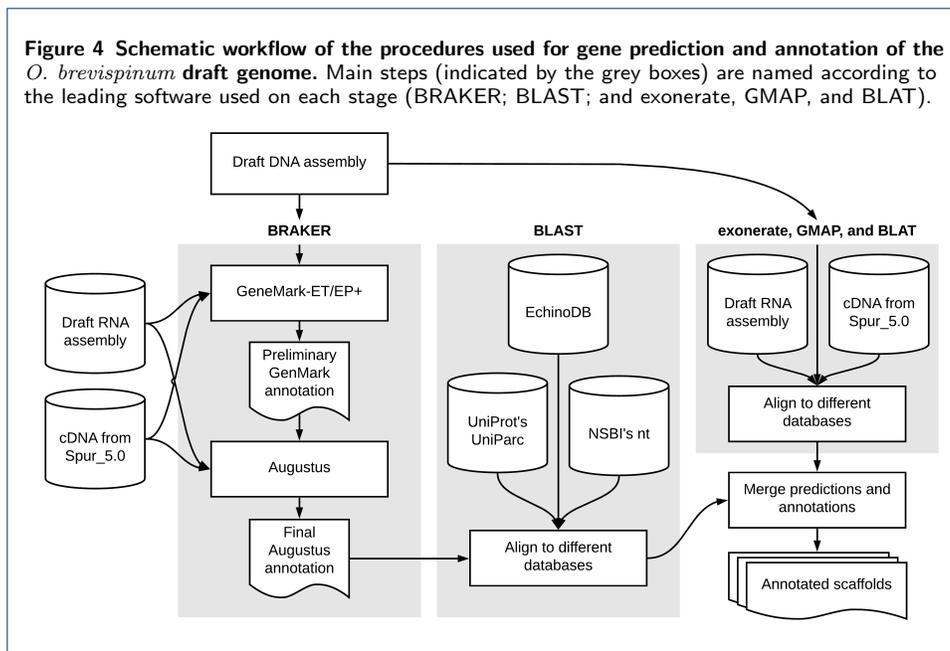
## Figures

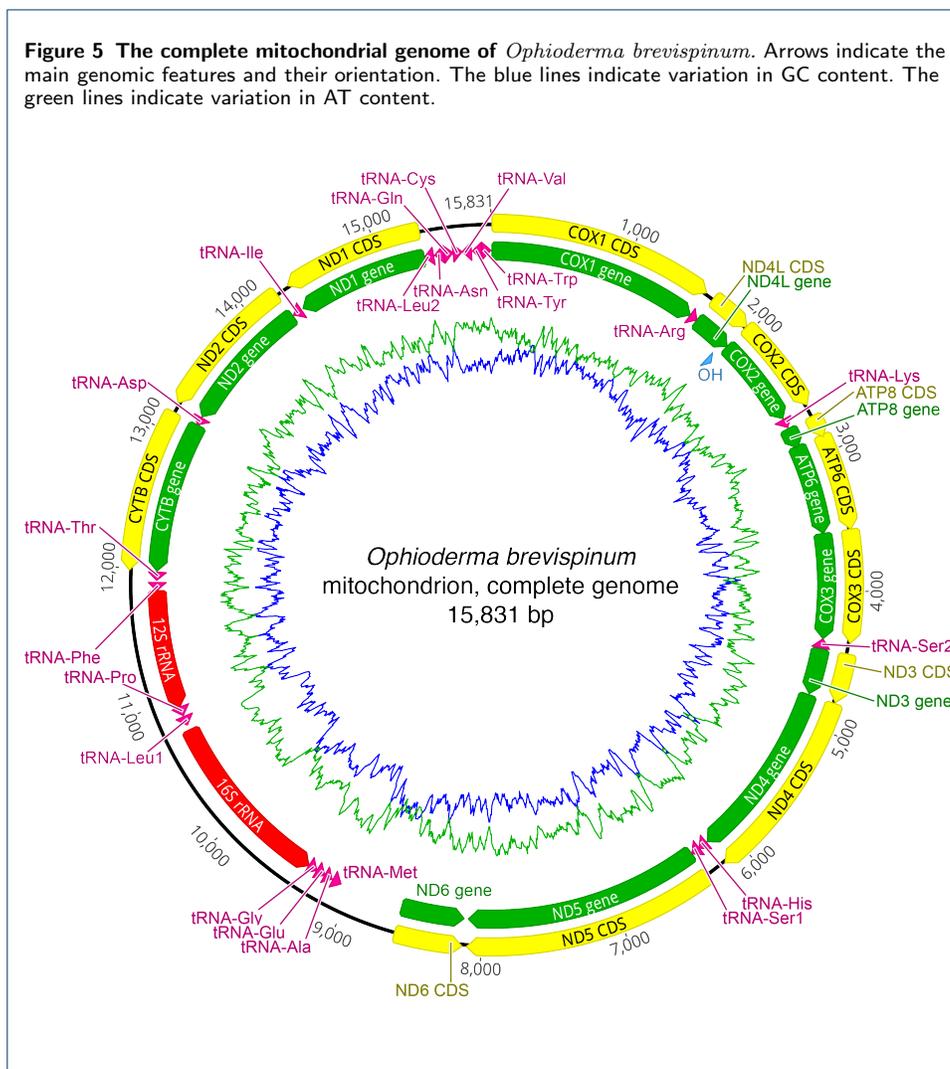


**Figure 2 Simplified diagram of the Notch signaling pathway.** The pathway is mediated by juxtacrine signaling that requires direct physical contact between the signaling and receptor cells. The *Delta/Serrate (Jagged)* ligands and *Notch* receptors are transmembrane proteins embedded into the plasma membrane of the signaling and receptor cells, respectively. Ligand-receptor interaction triggers conformational changes in the Notch protein that allows for proteolytic cleavage of the receptor by the ADAM metalloprotease and the multiprotein  $\gamma$ -secretase complex. The latter includes the catalytic component *presenilin*, as well as regulatory/stabilizing subunits *nicastrin*, *Aph-1*, and *Pen (presenilin enhancer)-2*. This proteolytic cleavage releases the Notch intercellular domain that translocates into the nucleus and activates the transcription factor *RBP-J* by inducing the release of co-repressors (e.g., *NCOR*, *CIR*, *MINT*, and *HDAC*) and recruitment of co-activators, such as *Mastermind (MAM)*, *p300*, and *NACK*. The activated transcription factor complex initiates transcription of the direct targets of the pathway, including *Hes* and *Hey*. Even though the pathway itself is conceptually simple, it is subjected to a multitude of regulatory inputs at multiple levels, including receptor post-translational maturation and stability/availability of the key pathway components in both the signaling and receptor cells. One of the properties of the Notch pathway is the ability to sustain itself through a series of feed-forward loops, thus resulting in an all-or-nothing response. For example, *NACK*, which is a transcriptional co-activator in the pathway is itself positively regulated by Notch. The genes shown in the diagram were searched for and identified in the *O. brevispinum* draft genome (see Table 5). Three different searches for Notch related genes in the draft genome were performed, designated by green boxes (BLAST, Exonerate, and conserved domain search respectively), filled boxes indicate positive identification.









## Tables

**Table 1** Genome assembly statistics for echinoderm genomes available at the NCBI's Assembly database (<https://www.ncbi.nlm.nih.gov/assembly>) on Jun. 17, 2021. Ditto marks ("") indicate values identical to the cell above. Asterisks (\*) indicate GenBank assembly accessions that have corresponding RefSeq assemblies.

Organism name	Class: Order	GenBank assembly	Length (Mbp)	Scaffold No.	Scaffold N50 (bp)	Citations
<i>Asterias rubens</i> (European starfish)	Asteroidea: Forcipulatida	GCA_902459465.3*	418	150	20,558,067	[78]
<i>Pisaster ochraceus</i> (purple sea star)	"	GCA_010994315.1	402	1,844	20,188,303	[79]
<i>Acanthaster planci</i> (crown-of-thorns starfish)	Asteroidea: Valvatida	GCA_001949145.1*	384	1,766	1,521,119	[14, 80, 81]
<i>Patiria miniata</i> (bat star)	"	GCA_000285935.1	811	60,183	52,614	NA
<i>Patiriella regularis</i> (starfish)	"	GCA_900067625.1	949	3,006,458	557	[15]
<i>Anneissia japonica</i> (crinoids)	Crinoidea: Comatulida	GCA_011630105.1*	590	76,727	623,489	NA
<i>Hemicentrotus pulcherrimus</i> (sea urchins)	Echinoidea: Camarodonta	GCA_003118195.1	559	16,251	142,559	[10, 82]
<i>Lytechinus pictus</i> (painted urchin)	"	GCA_900067635.1	811	60,183	52,614	[9]
<i>Lytechinus variegatus</i> (green sea urchin)	"	GCA_011630205.1	974	113,362	628,067	[83, 84]
<i>Strongylocentrotus purpuratus</i> (purple sea urchin)	"	GCA_000002235.4*	922	871	37,282,239	[8, 76, 85–102]
<i>Eucidaris tribuloides</i> (sea urchins)	Echinoidea: Cidaroida	GCA_001188425.1	2,187	637,071	39,192	NA
<i>Actinopyga echinites</i> (sea cucumbers)	Holothuroidea: Holothuriida	GCA_010015985.1	899	895,374	1,907	NA
<i>Holothuria glaberrima</i> (sea cucumbers)	"	GCA_009936505.1	1,128	346,783	1,221,172	NA
<i>Apostichopus japonicus</i> (Japanese sea cucumber)	Holothuroidea: Synallactida	GCA_002754855.1	805	3,278	487,241	[13]
<i>Apostichopus leukothele</i> (sea cucumbers)	"	GCA_010014835.1	481	74,445	1,493,354	NA
<i>Apostichopus parvimensis</i> (sea cucumbers)	"	GCA_000934455.1	873	21,559	89,133	NA
<i>Australostichopus mollis</i> (sea cucumbers)	"	GCA_900067635.1	1,252	3,712,641	626	[15]
<i>Paelopatides confundens</i> (sea cucumbers)	"	GCA_011317855.1	1,379	764,445	504,687	NA
<i>Stichopus horrens</i> (warted sea cucumber)	"	GCA_009801055.1	689	423,833	3,896	NA
<i>Ophioneis fasciata</i> (brittle stars)	Ophiuroidea: Amphilepidida	GCA_900067615.1	1,185	3,968,282	484	[15]
<i>Ophiothrix spiculata</i> (brittle stars)	"	GCA_000969725.1	2,764	75,696	72,780	NA

**Table 2** Summary metrics of the *O. brevispinum* genome assembly

Metrics	Quantification
Total assembly length	2,684,874,461 bp
Number of scaffolds	88,538
Shortest scaffold	2,035 bp
Longest scaffold	61,2917 bp
Mean scaffold length	30,325 bp
N50 scaffold length	48,505 bp
L50 (scaffolds)	15,677
Assembly GC content	38.4%
Repetitive DNA	1,410,344,530 bp (52.53%)

**Table 3** Summary of results from RepeatMasker v4.0.8, run with rmblastn v2.6.0+. This table corresponds to the classification of 1,410,344,530 bp (GC content of 38.40%) of repetitive DNA in the draft genome assembly of *Ophioderma brevispinum*. (\*) Most repeats fragmented by insertions or deletions have been counted as one element.

	Number of elements (*)	Length occupied (bp)	Percentage of sequence
<b>SINEs</b>	20,483	2,893,927	0.11%
ALUs	0	0	0.00%
MIRs	20483	2893927	0.11%
<b>LINEs</b>	356,695	144,072,365	5.37%
LINE1	25	1,598	0.00%
LINE2	142,719	53,518,810	1.99%
L3/CR1	21,833	7,360,396	0.27%
<b>LTR elements</b>	44174	30,059,426	1.12%
ERVL	15	795	0.00%
ERVL-MaLRs	0	0	0.00%
ERV_classI	1	47	0.00%
ERV_classII	15	963	0.00%
<b>DNA elements</b>	68,647	19,627,664	0.73%
hAT-Charlie	295	13,023	0.00%
TcMar-Tigger	0	0	0.00%
<b>Unclassified</b>	4,673,979	1,143,432,629	42.59%
<b>Total interspersed repeats</b>		1,340,086,011	49.91%
<b>Small RNA</b>	0	0	0.00%
<b>Satellites</b>	219	83,876	0.00%
<b>Simple repeats</b>	644,098	68,988,714	2.57%
<b>Low complexity</b>	72,298	4,790,218	0.18%

**Table 4** Summary of BUSCO v4.0.6 results. The database column names each odb10 BUSCO database used. The species column intricate the Augustus training parameter. Ditto marks (") indicate values identical to the cell above. The names "Human", "Fly" and "Spur" correspond to *Homo sapiens*, *Drosophila melanogaster*, and *Strongylocentrotus purpuratus*, respectively.

Database	Species	Complete					N
		All	Single-copy	Duplicated	Fragmented	Missing	
Metazoa	Human	30.20%	29.60%	0.60%	16.00%	53.80%	954
"	Fly	30.10%	29.40%	0.70%	18.10%	51.80%	954
"	Spur	25.60%	25.30%	0.30%	19.70%	54.70%	954
Eukaryota	Human	21.6%	21.6%	0.00%	17.60%	60.80%	255
"	Fly	22.40%	22.00%	0.40%	15.30%	62.30%	255
"	Spur	20.00%	19.60%	0.40%	16.50%	63.50%	255

**Table 5** Select components of the Notch signaling pathway identified in the draft genome of *O. brevispinum* using reference sequences from UniProt and Echinobase. For each gene, we list its name, the known function in the pathway, and whether or not the gene was recovered from the draft genome with independent BLAST and exonerate alignments. In addition, we also indicate if we could identify conserved protein domains in the predicted protein sequences.

Name	Role in the Pathway	BLAST	Exonerate	Conserved domains	References
Delta/Serrate (Jagged)	Ligand of the Notch receptor	Yes	Yes	Yes	[18, 32]
Notch	Receptor	Yes	Yes	No	[18, 103]
RBP-J	Transcription factor activated by Notch	Yes	Yes	No	[18, 32]
Mastermind	Co-activator of RBP-J	Yes	Yes	No	[18, 32, 103]
p300	Co-activator of RBP-J	Yes	Yes	No	[104, 105]
NACK	Co-activator of RBP-J	Yes	No	No	[104, 106]
CIR1	Co-repressor of RBP-J	Yes	Yes	No	[18, 32]
NCoR	Co-repressor of RBP-J	Yes	Yes	Yes	[107]
HDAC1	Co-repressor of RBP-J	Yes	Yes	Yes	[108]
MINT	Co-repressor of RBP-J	Yes	Yes	Yes	[107]
Fringe	Post-translational maturation of Notch	Yes	Yes	No	[18, 32]
POFUT1	Post-translational maturation of Notch	Yes	Yes	Yes	[109]
Poglut	Post-translational maturation of Notch	Yes	Yes	Yes	[110]
Furin	Receptor proteolysis	Yes	Yes	Yes	[111]
Neuralized	Ubiquitination of Jagged	Yes	Yes	Yes	[3, 112]
Mindbomb	Ubiquitination of Delta	Yes	Yes	Yes	[113]
Nicastrin	Receptor proteolysis	Yes	Yes	Yes	[18, 114]
Presenilin 1	Receptor proteolysis	Yes	No	No	[18, 114]
ADAM 10/17	Metalloprotease	Yes	Yes	No	[18]
HES	Canonical target gene	Yes	Yes	Yes	[114–116]
HEY1	Canonical target gene	Yes	Yes	Yes	[114–116]
Numb	Negative regulator of the Notch pathway	Yes	Yes	No	[114]
LNK2	Negative regulator of Numb	Yes	Yes	Yes	[117]
ACBD3	Activator of Numb	Yes	Yes	Yes	[114, 118]
NAK	Positive regulator of the Notch pathway	Yes	Yes	No	[114]
Mesp2	Activates Fringe, induces degradation of Mastermind	Yes	No	Yes	[103]
Nedd4	Targets Notch and Deltex for degradation	Yes	Yes	Yes	[119]
Notchless	Context-dependent positive or negative regulator	Yes	Yes	Yes	[33]
Deltex	Context-dependent positive or negative regulator. Antagonizes Nedd4	Yes	Yes	Yes	[108, 119]

**Supplementary information**

Supplementary information accompanies this paper at Zenodo, DOI: [10.5281/zenodo.4247706](https://doi.org/10.5281/zenodo.4247706).

Additional file 1: File S1

Bioinformatics protocols for quality control of raw sequence reads and subsequent genome assembly.

Additional file 2: File S2

Protocols for Feulgen image analysis densitometry and sequence-based genome size estimation.

Additional file 3: File S3

Expected variation of haploid genome sizes in echinoderms.

Additional file 4: File S4

Protocol for *de novo* DNA repeat assembly from shotgun sequence reads.

Additional file 5: File S5

Protocol for DNA repeat identification.

Additional file 6: File S6

Template scripts for gene prediction and annotation.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FileS1.pdf](#)
- [FileS2.pdf](#)
- [FileS3.pdf](#)
- [FileS4.pdf](#)
- [FileS5.pdf](#)
- [FileS6.pdf](#)