Data in Brief 55 (2024) 110588

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Nuclear and mitochondrial genome datasets for spiny lobsters genus *Panulirus* (Decapoda: Achelata: Palinuridae)

J. Antonio Baeza[a,b,c,*], Alyssa Baker[a], Michael Childress[c], Stacy Pirro[d]

[a] *Department of Biological Sciences, Clemson University, Clemson, SC, USA*
[b] *Smithsonian Marine Station at Fort Pierce, Smithsonian Institution, Fort Pierce, FL, USA*
[c] *Departamento de Biología Marina, Universidad Catolica del Norte, Coquimbo, Chile*
[d] *Iridian Genomes, Bethesda, MD, USA*

ARTICLE INFO

Article history:
Received 15 January 2024
Revised 24 April 2024
Accepted 29 May 2024
Available online 3 June 2024

Dataset link: Nuclear and mitochondrial genome datasets for spiny lobsters genus Panulirus (Decapoda: Achelata: Palinuridae) (Original data)

Keywords:
Mitogenome
Assembly
Crustacea
Arthropoda

ABSTRACT

Spiny lobsters (Decapoda: Palinuridae) in the genus *Panulirus* are targets of lucrative fisheries globally and have relevant ecological functions in tropical and subtropical environments. Only a few, but increasing, number of genetic and genomic resources exist for them. Nuclear and mitochondrial genome assemblies can provide insights into their phylogenetic relationships and support fishery management strategies in species that are heavily exploited. Herein, using Illumina short reads whole genome sequencing, we assembled the nuclear and mitochondrial genomes of a total of 14 species. Genomic DNA was extracted from specimens deposited at Clemson University Crustacean Collection and sequenced in a HiSeq X Ten system. The number of paired-end (PE) reads generated for the different studied species varied between 219,917,346 in *P. argus* and 70,215,423 in *P. cygnus*. Nuclear and mitochondrial genomes were '*de novo*' assembled. Nuclear genomes ranged between 1,624,400,357 bp in *P. guttatus* and 935,571,898 bp in *P. cygnus* with scaffold numbers varying between 466,583 in *P. versicolor* and 852,228 in *P. longipes*. Mitochondrial genomes varied between 15,613 bp and 15,768 bp in *P. pascuensis* and *P. versicolor*, respectively.

* Corresponding author at: Department of Biological Sciences, Clemson University, Clemson, SC, USA.
E-mail address: jbaezam@clemson.edu (J.A. Baeza).

https://doi.org/10.1016/j.dib.2024.110588

The totality of the short reads, nuclear, and mitochondrial genome assemblies are available at NCBI's GenBank.

Specifications Table

Subject	Crustacean Science
Specific subject area	Genomics
Data format	Raw and Analyzed
Type of data	Table
Data collection	Genomic DNA was extracted from 14 spiny and slipper lobster muscle tissue samples using the DNeasy Blood and Tissue Kit. Libraries were prepared using the Illumina TruSeq kit (Illumina, San Diego, CA, USA). Next generation sequencing was performed using Illumina HiSeq X Ten. Illumina reads were assembled using SPAdes v2.5. Sequence contamination was removed using FSCR v3.0 and GX. Contigs were integrated using Zanfona v1.0. Mitochondrial genomes were de novo assembled using GetOrgnanelle v1.6.4 using as a 'seed' the mitochondrial genome of *P. argus* (MH068821).
Data source location	Institution: Clemson University Crustacean Collection City: Clemson State: South Carolina Country: United States
Data accessibility	Repository name: NCBI GenBank Data identification number: Bioproject IDs: PRJNA1014903, PRJNA996201–PRJNA996212, PRJNA996222 BioSample IDs: SAMN36530433–SAMN36530444, SAMN36531993, SAMN37343104 SRA Accession Numbers: SRR13036344, SRR25338398, SRR25338439, SRR25338495, SRR25338518, SRR25338519, SRR25338784, SRR25338798, SRR25340716, SRR25340736, SRR25340737, SRR25341181, SRR25341344, SRR25341383 Mitogenome Accessions: NC_039671, OR612305–OR612317 Nuclear Assemblies: ASM3808873, ASM3236138, ASM3236140, ASM3236144, ASM3236148, ASM3236176, ASM3236186, ASM3236188, ASM3236170, ASM3227352, ASM3227360, ASM3227372. ASM3227384, ASM3478053

1. Value of the Data

- These nuclear and mitochondrial genome sequences will aid in evaluating phylogenetic relationships among species in the genus, understanding diversification processes, and exploring the historical biogeography of spiny lobsters.
- The dataset will be useful for bioprospecting and marine biodiversity monitoring using non-intrusive environmental DNA (eDNA) approaches.
- The generated dataset can be used as a base to detect misidentification/mislabeling of spiny lobsters in the marketplace and in monitoring and enforcing fisheries and trade management restrictions.

2. Background

The infraorder Achelata consists of the exclusively marine spiny (fam. Palinuridae) and slipper (fam. Scyllaridae) lobsters, which are characterized by enlarged antennae, a long-lived phyllosoma larval stage, and the absence of chelae [1,2]. *Panulirus* is the most specious genus of spiny lobster and contains 20 species and 5 subspecies that are targeted by lucrative tropical and

subtropical fisheries worldwide [2]. The aim of this study is to assemble nuclear and mitochondrial genomes in spiny lobsters belonging to the genus *Panulirus* using a Illumina short read whole genome sequencing (WGS) strategy. The genomic datasets produced by this study represent new resources to aid with conservation and management via application in eDNA sampling and in species identification via barcoding and have further use in evaluating the adaptive evolution of protein-coding genes.

3. Data Description

Illumina short read WGS was conducted for a total of 13 species belonging to the spiny lobster genus *Panulirus* and the slipper lobster *Scyllarides nodifer* to assemble nuclear and mitochondrial genomes. Nuclear genomes were assembled using SPAdes v2.5, FSCR v3.0, and Zanfona v1.0 and are available in the National Center for Biotechnology Information (NCBI) Genbank (Table 1). Draft nuclear genomes ranged from 935.6 Mb in *P. cygnus* to 1.6 Gb in *P. guttatus* and consisted of 618,989 to 989,875 contigs in *P. laevicauda* and *P. pascuensis*, respectively. Scaffold numbers varied between 466,583 in *P. versicolor* and 852,228 in *P. longipes.* Mitogenome sequences were assembled via the GetOrganelle v1.6.4 pipeline and are available at NCBI GenBank accession numbers NC_039671 and OR612306–OR612317. Mitogenome lengths varied between 15,613 bp and 15,768 bp in *P. pascuensis* and *P. versicolor*, respectively. Corresponding raw reads are available as SRA datasets in NCBI GenBank under BioProjects PRJNA996201-PRJNA996212 and PRJNA1014903. Nuclear, mitochondrial, and raw read assembly details are summarized in Table 1.

4. Experimental Design, Materials and Methods

4.1. Specimens, DNA extraction, and sequencing

All specimens used for sequencing belonging to the different species of *Panulirus* were available at the Clemson University Crustacean Collection (Table 1). Small tissue samples (approx. 5 mm^3) were dissected from pereiopods and immediately stored separately in sterile centrifuge tubes containing ethyl alcohol (95 %) that were shipped to Iridian Genomes, Inc. (Bethesda, MD) for genomic DNA (gDNA) extraction and next generation sequencing (NGS). gDNA was extracted from each tissue sample with the DNeasy Blood and Tissue Kit (Qiagen, Germany) using the manufacturer's protocol. Library preparation for each sample was performed using the Illumina TruSeq kit following the manufacturer's instructions. NGS was performed in a Illumina HiSeq X Ten system (Illumina, San Diego, CA, USA) using a 2 $\times$ 150 cycle. Between 219,917,346 and 70,215,423 pairs (PE) of reads were produced, respectively, for *P. argus* in *P. cygnus* by Iridian Genomes and are available in the short read archive (SRA) repository (Bioprojects: PRJNA1014903, PRJNA996201–PRJNA996212, PRJNA996222; BioSamples: SAMN36530433–SAMN36530444, SAMN36531993, SAMN37343104; SRA accession number: see Table 1) at NCBI's GenBank.

4.2. Nuclear genome assembly

The totality of the reads available for each specimen were used for nuclear genome assembly. First, the raw reads were trimmed of adapter sequences and low-quality regions with Trimmomatic v0.33 [3]. Next, trimmed sequences were assembled using SPAdes v2.5 [4]. Then, we used NCBI's sequence contamination screening pipeline FSCR v3.0 and GX (https://github.com/ncbi/fcs/wiki/FCS-GX) [5] for identifying and removing technical and biological contaminants from the newly assembled genomes. Lastly, we applied a finishing step to the assembly with the

Table 1
Raw reads, nuclear, and mitochondrial assemblies for spiny lobsters *Panulirus* spp. and the slipper lobster *Scyllarides nodifer*.

	Raw Reads				Mitochondrial Genome			Nuclear Genome				
Species Name	SRA	Spots (bp)	Bases (G)	GC (%)	Genbank Accession Number	Length (bp)	Mitogenome Coverage	Genbank Accession Code	Assembly Length (bp)	Assembly Coverage	Contigs (n)	Scaffolds (n)
Panulirus argus	SRR13036344	219,917,346	66	44.9	NC_039671	15,739	70x	ASM3808873	1812,672,735	90x	434,038	424,852
Panulirus cygnus	SRR25340737	70,215,423	21.1	44.4	OR612313	15,731	30x	ASM3236148	935,571,898	50x	926,678	844,337
Panulirus gracilis	SRR25338518	76,845,122	23.1	43.4	OR612307	15,745	30x	ASM3236144	1279,827,122	50x	751,506	679,949
Panulirus guttatus	SRR25338519	91,313,030	27.4	45.2	OR612312	15,702	30x	ASM3236138	1624,400,357	50x	746,306	657,369
Panulirus homarus	SRR25340716	82,043,440	24.6	43.3	OR612305	15,665	30x	ASM3236140	1313,005,488	50x	704,650	618,289
Panulirus inflatus	SRR25341181	78,968,575	23.7	43.0	OR612314	15,670	30x	ASM3236176	1321,811,166	50x	705,867	632,377
Panulirus interruptus	SRR25338398	83,625,231	25.1	44.8	OR612311	15,657	30x	ASM3227372	1407,497,722	50x	859,704	748,487
Panulirus laevicauda	SRR25338798	82,408,003	24.7	43.2	OR612306	15,675	30x	ASM3227360	1371,664,687	50x	618,989	552,489
Panulirus longipes	SRR25338439	80,710,679	24.2	43.8	OR612309	15,706	30x	ASM3227384	1243,798,248	50x	972,289	852,228
Panulirus marginatus	SRR25341383	86,771,275	26.0	43.9	OR612310	15,725	30x	ASM3236188	1276,116,527	50x	959,315	842,519
Panulirus ornatus	SRR25341344	82,272,566	24.7	43.9	OR612315	15,677	30x	ASM3227352	1364,558,317	50x	739,006	666,649
Panulirus pascuensis	SRR25340736	74,534,229	22.4	44.3	OR612316	15,613	30x	ASM3236186	1147,801,627	50x	989,875	848,215
Panulirus versicolor	SRR25338784	89,867,779	27.0	43.4	OR612308	15,768	30x	ASM3236170	1465,571,352	50x	518,312	466,583
Scyllarides nodifer	SRR25338495	81,609,917	24.5	44.0	OR612317	15,648	30x	ASM3478053	1014,165,050	50x	888,648	801,177

pipeline Zanfona v1.0 [6] to make additional contig joins based on conserved regions in related species.

4.3. Mitochondrial genome assembly

The mitochondrial genomes of each of the studied species were '*de novo*' assembled with the program GetOrganelle v1.6.4 [7]. The complete mitochondrial genome of the congeneric Caribbean spiny lobster *Panulirus argus* (GenBank's accession number MH068821- [8]) was used as a 'seed' for each of the assemblies. All assemblies were run using k-mer sizes of 21, 55, 85, and 115.

Limitations

Non Applicable

Ethics Statement

The authors declare that this work follows the ethical requirements for publication in Data in Brief and does not involve human subjects or animal experiments that require ethical approval.

CRediT Author Statement

J. A. Baeza: Conceptualization, Resources, Methodology, Formal analysis, Data curation, Visualization, Writing – original draft. **A. Baker:** Methodology, Formal analysis, Data curation, Visualization, Writing – review & editing. **S. Pirro:** Resources, Methodology, Formal analysis, Data curation, Visualization, Writing – review & editing. **M. Childress:** Resources, Writing – review & editing.

Data Availability

Nuclear and mitochondrial genome datasets for spiny lobsters genus Panulirus (Decapoda: Achelata: Palinuridae) (Original data) (NCBi's GenBank).

Acknowledgments

J.A. Baeza thanks Dr. Vincent. P. Richards for bioinformatics support. The great majority of the specimens available in the collection were donated by Dr. R. W. George to Dr. M. Childress.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] L.B. Holthuis, XVI. – On a small collection of isopod crustacea from the greenhouses of the Royal Botanic Gardens, Kew, Ann. Mag. Nat. Hist. 98 (1946) 122–137, doi:10.1080/00222934608654533.
[2] L.B. Holthuis, FAO species catalogue. Vol. 13. Marine lobsters of the world. An annotated and illustrated catalogue of species of interest to fisheries known to date, FAO Fish. Synop. 125 (1991) 13.
[3] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120, doi:10.1093/bioinformatics/btu170.
[4] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (2012) 455–477, doi:10.1089/cmb.2012.0021.
[5] A. Astashyn, E.S. Tvedte, D. Sweeney, V. Sapojnikov, N. Bouk, V. Joukov, E. Mozes, P.K. Strope, P.M. Sylla, L. Wagner, S.L. Bidwell, Rapid and sensitive detection of genome contamination at scale with FCS-GX, bioRxiv (2023), doi:10.1101/2023.06.02.543519.
[6] M. Kieras, Zanfona, a genome finishing process for use with paired-end short reads (2021) https://github.com/zanfona734/zanfona.
[7] J.J. Jin, W.B. Yu, J.B. Yang, Y. Song, C.W. DePamphilis, T.S. Yi, D.Z. Li, GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes, Genome Biol. 21 (2020) 241, doi:10.1186/s13059-020-02154-5.
[8] J.A. Baeza, The complete mitochondrial genome of the Caribbean spiny lobster *Panulirus argus*, Nat., Sci. Rep. 8 (2018) 17690, doi:10.1038/s41598-018-36132-6.