Yes, we can use it: A formal test on the accuracy of low-pass nanopore long-read sequencing for mitophylogenomics + barcoding research using the Caribbean spiny lobster *Panulirus argus*

J. Antonio Baeza

Department of Biological Sciences, Clemson University, Clemson, SC 29634, USA. Email: jbaezam@clemson.edu.

ABSTRACT

Whole mitogenomes or short fragments (e.g., 300-700 bp of the *cox1* gene) are markers of choice for revealing within- and among-species geneticities. Protocols for sequencing and assembling mitogenomes include primer walking or 'long PCR' followed by Sanger sequencing or low-coverage whole genome (LC-WGS) sequencing with or without prior mitochondrial enrichment and Illumina sequencing. The aforementioned strategies assemble complete and accurate mitochondrial genomes but are time consuming and/or expensive. In this study, I tested first if mitogenomes can be sequenced from long-read nanopore sequencing data exclusively. Second, I explored the accuracy of the long-read assembled genomes by comparing them to a 'gold' standard reference mitogenome retrieved from the same individual using Illumina sequencing. LC-WGS using a MiniON ONT device and various de-novo and reference-based assembly pipelines did retrieve a complete and highly accurate mitogenome for the Caribbean spiny lobster *Panulirus argus*. Discordance between each of the long-read assemblies and the reference mitogenome was mostly due to indels at the flanks of homopolymer regions. Although not 'perfect', phylogenetic analyses using entire mitogenomes or a fragment of the *cox1* gene demonstrated that mitogenomes assembled using long reads can reliably identify the sequenced specimen as belonging to *P. argus* and distinguish it from other closely and distantly related species in the same genus, family, and superorder. This study serves as a proof-of-concept for the future implementation of in-situ surveillance protocols using the MiniON as to detect mislabeling in *P. argus* across its supply chain. Mislabeling detection will improve fishery management in this overexploited lobster. This study will additionally aid in decreasing costs for exploring meta-population connectivity in the Caribbean spiny lobster and will aid with the transfer of genomics technology to low-income and developing countries.

INTRODUCTION

The aims of this study were threefold. First, I tested if a mitochondrial genome can be sequenced and assembled from long read nanopore sequencing data exclusively using both a de novo and a reference-based strategy. Second, I explored the quality (i.e., accuracy) of the long read-based assembled genomes by comparing them to a 'gold' standard mitochondrial genome retrieved from the same individual but generated using short read Illumina sequencing data. Sequence accuracy was explored for different long read assembly pipelines with multiple metrics including completeness, identity, and coverage. Furthermore, a detailed quantitative analysis of error type for long read assemblies was conducted. Third and lastly, I tested if the de novo and reference-based long read assemblies are useful for mitophylogenomic and barcoding research. I specifically assessed if long read assemblies do have phylogenomic information that permit to reliably identify the sequenced specimen and distinguish it from other closely and distantly related species in the same genus, family, and superorder.

To accomplish the goals above, I used as a model system the Caribbean spiny lobster *Panulirus argus*, an ecologically relevant species in shallow water coral reefs and target of the most lucrative fishery in the greater Caribbean region.

MATERIAL & METHODS

Field collections were approved by FWCC (SAL-11-1319-SR). A single adult specimen of *P. argus* was collected from Alligator Reef, FL, USA. A small piece of muscle was dissected from a dismembered pereopod, and immediately preserved in 95% ethanol. Next, total genomic DNA from ~15-25 mg muscle tissue was extracted using standard protocols. Approximately one microgram of gDNA extracted from each sample was used for Illumina and MiniON ONT library construction following standard protocols.

The pipeline Novoplasty was used for mitogenome assembly with short reads. For long reads, the mitochondrial genome of *P. argus* was assembled using the following workflow:

RESULTS

Mitogenome assembly using short & long reads

Novoplasty assembled and circularized the mitochondrial chromosome of *P. argus* (15 729 bp) with an average coverage of 710x (MH068821). The mtDNA of *P. argus* comprises 13 protein-coding genes (PCGs), 2 ribosomal RNA genes (rrnS and rrnL), and 22 transfer RNA (tRNA) genes.

All long read pipelines, with and without 'extra' polishing with Medaka, also assembled and circularized the mitochondrial genome of *P. argus*.

All long read assemblies were very similar (a close match) to the reference genome (patristic distance range: 6.36821×10^{-5} - 6.3755×10^{-4}) (Fig. 1).