Manuscript Title

Authors

- **John Doe**
  XXXX-XXXX-XXXX-XXXX · johndoe · johndoe
  Department of Something, University of Whatever · Funded by Grant XXXXXXXX

- **Jane Roe**
  XXXX-XXXX-XXXX-XXXX · janeroe
  Department of Something, University of Whatever; Department of Whatever, University of Something

Abstract

The blue whale, *Balaenoptera musculus*, is the world's largest animal and, to our knowledge, the largest to have ever existed. We report a high-quality assembly of the genome of this fascinating species. This is one of the first genomes of a cetacean sequenced using a long-read technology. We sequenced genomic DNA using PacBio long read, Illumina short read, and 10X Genomics synthetic long read technologies. We also obtained long-range mapping information using Bionano optical mapping and Dovetail Hi-C. Additionally, we sequenced the transcriptome of blue whale fibroblasts using Illumina RNA-seq and PacBio Iso-seq technologies. We also measured genome size using a flow cytometry technique. We report on 1) comparison of alternative long read and short read based assemblies, 2) segmental duplications within the blue whale genome, and 3) historical demography of Pacific and Atlantic blue whale populations. A high quality, annotated blue whale genome shall serve as an important resource for developmental and stem cell biology, evolution, and conservation research.

Test

Introduction

[Erich, Arang, Adam: please review] We report a high-quality genome assembly of the world's largest animal, the blue whale Balaenoptera musculus. The blue whale is a mammal, and mammalian genomes are challenging to assemble because they are relatively large, in the 2-4 billion base pairs (GB) range, and rich in repetitive elements. Ours is one of the first cetacean genome assemblies generated using PacBio long-read technology, which enables better resolution of repetitive regions and improved contiguity compared to older, short read technologies. In order to assess how the use of long reads affects assembly contiguity, we compare our assembly to previously reported assemblies of other cetacean genomes.

Genome assembly is a rapidly developing field. Several sequencing and long-range mapping technologies are available. Both experimental methods and assembly algorithms are continually updated, while new contenders emerge. Several alternative paths to a genome assembly are possible, with various strengths and weaknesses, making it a challenge to choose an optimal strategy. In particular, there is a continuing debate on whether using long read technologies is worth the extra cost. Our blue whale assembly is based on a combination of PacBio long read sequencing, linked long reads generated by Illumina sequencing of a library prepared using 10X Genomics Chromium technology, Bionano optical maps, and Dovetail Hi-C. This assembly was generated using a computational workflow developed by the Vertebrate Genomes Project (VGP) consortium (Rhie, McCarthy et al. 2020). We compare it to an alternative assembly based on 10X linked long reads technology without the use of PacBio, generated with the help of assay vendors using readily available software.

The blue whale appears to be the largest animal to have ever existed (Würsig, Thewissen et al. 2017). An adult can reach up to 110 feet and weigh 330,000 pounds . Genomic studies of giant animals are of interest to several sub-fields of biomedical science. Understanding developmental mechanisms that control body size may have applications in regenerative medicine and animal husbandry. Furthermore, large mammals tend to have long lives and have developed mechanisms that make them resistant to cancer. They rarely succumb to cancer in spite of having orders of magnitude more cells (and thus more cell divisions) and substantially longer life spans then their smaller cousins. This puzzling phenomenon is known as Peto's Paradox (Caulin and Maley 2011).

Previously sequenced genomes of large animals, including other large whales, yielded some clues with regard to their possible mechanisms of cancer resistance. These include, for example, mutations and duplications of known tumor suppressors and other potentially relevant genes involved in processes such as DNA repair and apoptosis (Keane, Semeiks et al. 2015, Sulak, Fong et al. 2016, Tollis, Robbins et al. 2019), and control over the abundance of microsatellite repeats (Park, An et al. 2016). Mammalian genomes contain relatively large regions of duplicated sequence, known as segmental duplications (SDs), many of which contain genes. We report on detecting SDs in the blue whale genome and on genes that are found in those regions.

[Phil and Aimee] The blue whale is classified as Endangered by the IUCN Red List and is on Appendix I of both the Conservation on International Trade in Wild Species of Fauna and Flora (CITES) and the Convention on the Conservation of Migratory Species of Wild Animals (CMS) (Cooke 2018). Genomes of endangered species facilitate studies of their population structure and diversity, thereby assisting conservation efforts (e.g. (Foote, Martin et al. 2019)). We report an investigation of genome heterozygosity and demographic histories of North Pacific and North Atlantic blue whales.

References