ABSTRACT

Whilst much sequencing effort has focused on key mammalian model organisms such as mouse and human, little is known about the correlation between genome sequencing techniques for non-model mammals and genome assembly quality. This is especially relevant to non-model mammals, where the samples to be sequenced are often degraded and low quality. A key aspect when planning a genome project is the choice of sequencing data to generate. This decision is driven by several factors, including the biological questions being asked, the quality of DNA available, and the availability of funds. Cutting-edge sequencing technologies now make it possible to achieve highly contiguous, chromosome-level genome assemblies, but relies on good quality high-molecular-weight DNA. Here we use a range of different genomic technologies generated from a roadkill European Polecat (Mustela putorius) to assess various assembly techniques on this low-quality sample. We evaluated different approaches for de novo assemblies and discuss their value in relation to biological analyses. The high degree of variability between each de novo assembly method (assessed from the seven key metrics) highlights the importance of carefully devising the sequencing strategy to be able to carry out the desired analysis. Adding more data to genome assemblies does not always result in better assemblies so it is important to understand the nuances of genomic data integration explained here, in order to obtain cost-effective value-for-money when sequencing genomes.

DOI

dx.doi.org/10.17504/protocols.io.bd3ri8m6

COLLECTION CITOATION


https://dx.doi.org/10.17504/protocols.io.bd3ri8m6

COLLECTIONS

2020 Featured Protocols

KEYWORDS

polecat, vertebrate, non-model organism, Illumina, chromium, Bionano, assembly, sequencing

LICENSE

This is an open access collection distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

CREATED

Mar 23, 2020


This is an open access protocol distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
ABSTRACT
Whilst much sequencing effort has focused on key mammalian model organisms such as mouse and human, little is known about the correlation between genome sequencing techniques for non-model mammals and genome assembly quality. This is especially relevant to non-model mammals, where the samples to be sequenced are often degraded and low quality. A key aspect when planning a genome project is the choice of sequencing data to generate. This decision is driven by several factors, including the biological questions being asked, the quality of DNA available, and the availability of funds. Cutting-edge sequencing technologies now make it possible to achieve highly contiguous, chromosome-level genome assemblies, but relies on good quality high-molecular-weight DNA. Here we use a range of different genomic technologies generated from a roadkill European Polecat (Mustela putorius) to assess various assembly techniques on this low-quality sample. We evaluated different approaches for de novo assemblies and discuss their value in relation to biological analyses. The high degree of variability between each de novo assembly method (assessed from the seven key metrics) highlights the importance of carefully devising the sequencing strategy to be able to carry out the desired analysis. Adding more data to genome assemblies does not always result in better assemblies so it is important to understand the nuances of genomic data integration explained here, in order to obtain cost-effective value-for-money when sequencing genomes.