High-Molecular-Weight DNA

Corrales, Carolina

Pensoft Publishers
2023

Corrales , C , Mackenzie-Dodds , J , Mulcahy , D , Poczai , P , Haring , E , Martín , M P & Astrin , J J 2023 , High-Molecular-Weight DNA . in C Corrales & J J Astrin (eds) , þÿ B i o d i v e r s i t y B i o b a n k i n g a H a n d b o o k o n P r o t o c o l s a n d P r a c t i c e s . P e r 137-139 . https://doi.org/10.3897/ab.e101876

http://hdl.handle.net/10138/357480
https://doi.org/10.3897/ab.e101876

CHAPTER 9
High-Molecular-Weight DNA

Carolina Corrales, Jackie Mackenzie-Dodds, Daniel Mulcahy, Péter Poczai, Elisabeth Haring, María Paz Martín, and Jonas J. Astrin

Introduction

International initiatives, such as the Earth BioGenome Project, are aiming to characterise complete genomes of all extant species using long-read sequencing technologies (e.g., PacBio, Oxford Nanopore) (Blaxter et al. 2022; Formenti et al. 2022). Hence, long DNA fragments are crucial, and can only be obtained by having access to high-molecular-weight DNA (HMW, 45–150 kb) and ultra-high-molecular-weight DNA (uHMW, over 150 kb) (Mulcahy et al. 2016; Ryder and Onuma 2018; Dahn et al. 2022).

Although second-generation technologies (e.g., Illumina) have transformed DNA analyses and become standard applications (Kchouk et al. 2017; Slatko et al. 2018), they are based on short-read sequencing and cannot deal with complex genomes that contain repetitive areas; thus, the resulting genome assemblies are often incomplete and fragmented (Kchouk et al. 2017; Dahn et al. 2022). Long-read sequencing or third-generation sequencing, on the other hand, can generate fragments long enough to overlap, despite of having a higher error rate (Rhoads and Au 2015; Amarasinghe et al. 2020; Delahaye and Nicolas 2021).

The performance of HMW DNA is dependent on the sample collection and DNA extraction procedures. From the start during field collection, adequate sample preservation methods have to be used to maintain high DNA integrity and purity. The best preservation method for non-live material is flash-freezing, but samples preserved in 95% ethanol or 20–25% DMSO-EDTA (for vertebrates) stored at 4 °C for up to one week also show little degradation (Dahn et al. 2022). If sampling occurs in hot climates, the use of insulated boxes, ice packs, wet ice, dry ice, or electronic coolers should be considered (Dahn et al. 2022), up to LN2-cooled dry shippers. Dahn et al. (2022) provide guidelines regarding sample preservation and choice of tissue for different vertebrates to ensure a high DNA quality. Further guidance on collection, preparation and storage of animal, plant, and fungal material to be used for WGS is provided in the PacBio technical note (2018).

RECOMMENDATION
Commercial reagents were mainly optimised for lower molecular weight DNA, and therefore, are not suitable either for uHMW preservation or for chromosomal 3D interactions (Hi-C) (Dahn et al. 2022).

Some recommendations should be followed if Hi-C methods are aimed for, as these require intact cell nuclei (Lajoie et al. 2015; ERGA 2021; **PacBio Hi-C requirements**, **National Genomics infrastructure**-Sweden):

- Tissue should be quantified: 20–200 mg soft, non-fatty animal tissue (internal organs, or muscle) is needed. Liver samples should not be used as input samples due to the amount of enzymes that can degrade DNA. Whole small animal specimens should occupy at least 50 µl. 1 ml non-nucleated blood in EDTA or heparin should be collected or 2–5 ml if flash-frozen. For nucleated blood, 200 µl flash frozen. 300 g young leaves with non-fibrous tissue.
- Cell cultures can also be used (10–50 million cells or more).
- Samples should remain frozen during transport, grinding and transfer procedures.

RECOMMENDATION
All samples that were preserved frozen with a preservation solution, in RNA later, or in ethanol, require pre-treatment before DNA extraction.

Specific DNA extraction methods are available to produce uHMW, such as bead-based methods (e.g., MagAttract HMW DNA kit), agarose plug methods (Bionano Prep Soft/Fibrous Tissue Protocol), or the the Circulomics thermoplastic magnetic disk (Nanobinds) method (Dahn et al. 2022). For HMW, the portal "**Extract DNA for PacBio**" provides a list of publications

describing extraction protocols for subsequent PacBio sequencing. Further DNA extraction protocols for third generation sequencing for several taxa can be found in the **protocols.io** repository, as well as in Green and Sambrook (2018) and Pereira (2022). Note that bead beating is required to break tough cell walls of plants, fungi, and some microorganisms, but it can also fragment DNA (Heavens et al. 2021). Li et al. (2020), Stark et al. (2020), Jones et al. (2021), and Russo et al. (2022) have optimised protocols that can overcome this disadvantage, as well as to remove contaminants (e.g., polysaccharides). For marine organisms, refer to Panova et al. (2016). For environmental samples, the **Earth Microbiome Project** (Marotz et al. 2017), Sakai (2021), and Trigodet et al. (2022) have established protocols for long-read sequencing.

If phenol-chloroform extraction protocols are used, the phenol has to be fresh and not oxidised (GTF 2020). If further purification is needed, the Qiagen Genomic-tip 500/G, MoBio PowerClean columns, or a high-salt clean-up protocol can be used for a wide range of samples (PacBio 2014; GTF 2020). Note that the latter may lead to a loss of 50% of the sample (UC Davis Genome Center 2022).

RECOMMENDATION
If possible, an RNAse digestion step should be included after DNA extraction, as samples must be RNA-free before proceeding with long-read sequencing. Further information can be found at the **UC Davis Genome Center website**.

Recording the technical value/quality of a DNA sample will allow researchers to estimate the probability of success of their planned downstream analyses. Sequencing technologies have technical requirements for successful sequencing results. If these are not met, sequencing results will not be reliable, and the sequence quality and quantity will be lower than expected.

The following table shows the ideal DNA quantities for long-read sequencing (ERGA 2021; UC Davis Genome Center 2022). Note that values can vary depending on the sample/library preparation and sequencing platform:

Table 6.

DNA	PacBio	Oxford Nanopore
Quantity	23 µg (min 15 µg) 7 µl small genomes 3.2 µl microorganisms	>5 µg
Concentration	50 ng/µl	100 ng/µl
Volume	50 µl – 400 µl	50 µl
DIN	>8	

RECOMMENDATION
The database for **long-read sequencing** provides access to existing analytical tools, and it can help in planning and performing best-practice analyses.