Oxford NANOPORE Technologies

WORKFLOW: ANIMAL GENOME ASSEMBLY

Assembling animal genomes using long nanopore sequencing reads

The release of the first animal genome sequence, for the nematode *Caenorhabditis elegans*[1], instigated a new era of animal genome sequencing and the acquisition of a wealth of knowledge about animal biology, evolution, and biodiversity[2]. However, despite such progress, only around 0.2% of animal species have had their genomes sequenced, and because many assemblies have been derived using short-read sequencing technology, many remain incomplete[2].

Compared to short-read data, long and ultra-long nanopore sequencing reads enable the resolution of repeat-rich sequences and large-scale structural variants, and demonstrate a lack of bias in GC-rich regions, supporting the assembly of high-quality, highly contiguous animal genomes. Epigenetic modifications can also be explored through direct sequencing of native DNA. These strengths, in combination with the high yield and throughput of PromethION™ sequencing devices, allow nanopore technology to further our understanding of genetic variation across the animal kingdom, and therefore may ultimately advance breeding and conservation efforts.

Here we present a simple workflow for animal genome assembly from a mammalian blood sample using PromethION Flow Cells.

EXTRACTION: obtaining high molecular-weight DNA

Find more extraction protocol recommendations for your sample type, from avian blood to insect and reptilian tissue: **community.nanoporetech.com/docs/prepare**

Selecting a suitable extraction method for obtaining high molecular-weight DNA greatly depends on sample type. For the extraction of ultra-high molecular-weight DNA (>50 kb) from whole animal blood, we recommend using the **NEB Monarch HMW DNA Extraction Kit**.

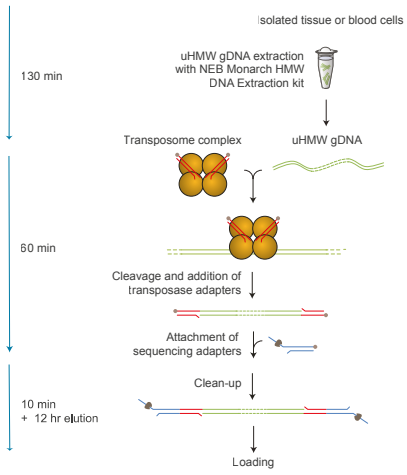
However, if it is not possible to use this option — for example if the starting sample is too fragmented or insufficient sample is available — we recommend using the **QIAGEN Puregene Blood Kit**, which we have found to generate high sequencing yields and maximise read lengths between 25–35 kb.

LIBRARY PREPARATION: selecting a kit

Find out more about library prep solutions: **nanoporetech.com/products/kits**

Obtaining long sequencing reads is important for genome assembly, as it maximises the overlap between reads at the downstream analysis stage. There is no upper read length limit in nanopore sequencing and generating ultra-long reads will increase both assembly continuity and contiguity. We recommend preparing ultra-high molecular-weight gDNA for sequencing using the **Ultra-Long DNA Sequencing Kit**, generating read length N50s >50 kb.

If it is not possible to extract ultra-high molecular-weight DNA, we recommend using the **Ligation Sequencing Kit**, which provides the greatest yield and control over read lengths. When using this kit, we recommend the Oxford Nanopore **Short Fragment Eliminator Kit** to size select for fragments >25 kb and light shearing of the extracted gDNA using the **Diagenode Megaruptor 3**.



Learn more about whole-genome sequencing and assembly, including how to incorporate ultra-long reads, in our Getting started guide: **nanoporetech.com/resource-centre/whole-genome-sequencing-large-genomes**

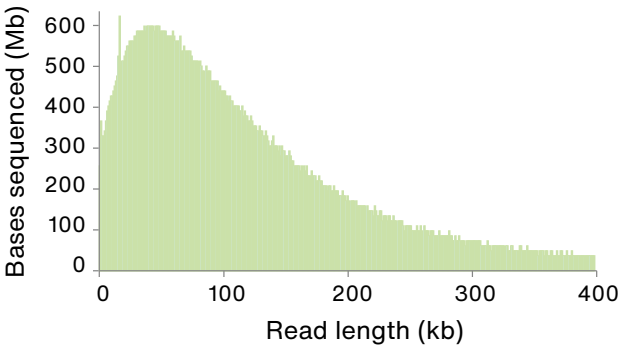
SEQUENCING: generating high yields of long reads with PromethION

Find out more about the PromethION range:
nanoporetech.com/products/promethion

Assuming an animal genome of ~2–4 Gb in length, we recommend sequencing to a minimum depth of 20x when using the **Ultra-Long DNA Sequencing Kit**. This can be achieved using one PromethION Flow Cell. For best assembly metrics, sequencing to a depth of 30x will further improve completeness and contiguity. Throughput can be maximised by a nuclease flush using the **Flow Cell Wash Kit** and loading fresh library every 24 hours. We recommend basecalling using high accuracy mode or super accuracy mode.

If using the **Ligation Sequencing Kit**, we recommend sequencing 25–35 kb reads to a minimum depth of 30x, which can also be achieved on one PromethION Flow Cell.

Read length distribution obtained from a library (mammalian blood) using the Ultra-Long DNA Sequencing Kit



The high-throughput PromethION 24 and 48 sequencing devices are configured for sequencing up to 24 or 48 high-yield PromethION Flow Cells, providing ultimate flexibility and adaptability to your sequencing needs. For lower throughput requirements, the compact PromethION 2 Solo — which can be plugged into a GridION™ or existing computer infrastructure — and the standalone PromethION 2 enable sequencing on up to two flow cells, for PromethION-scale sequencing in any lab.

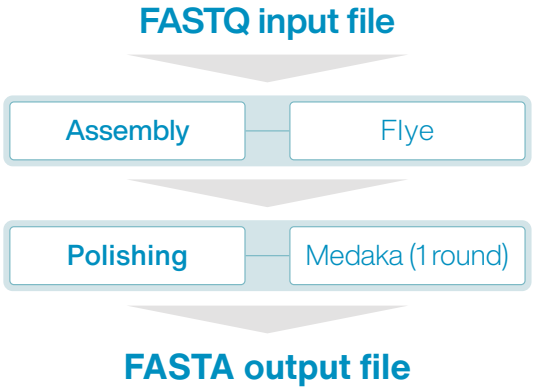
Find out more about the Flow Cell Wash Kit: store.nanoporetech.com/flow-cell-wash

ANALYSIS: selecting an assembly tool

Find out more about data analysis solutions:
nanoporetech.com/analyse

To assemble animal genomes, we suggest using the third-party *de novo* assembly tool **Flye**[3]. This analysis package represents a complete pipeline, taking raw nanopore reads as input, and producing polished contigs as output. We also advise one round of additional polishing of the assembly with **Medaka**[4]. These tools can both be found on GitHub.

Regarding analysis runtime, assuming an animal genome of ~2–4 Gb sequenced at 30x, assembly with **Flye** would require approximately one day (based on an AWS instance, with 1 TB RAM and 128 CPU threads). Polishing with **Medaka** would require an additional day.



Find out more at: **nanoporetech.com/applications/animal-genomics**

Oxford NANOPORE Technologies

Twitter: @nanopore
www.nanoporetech.com

References:
1. The *C. elegans* Sequencing Consortium. Science. 282:2012-2018 (1998).
2. Hotaling, S. et al. PNAS. 118(52):e2109019118 (2021).
3. Kolmogorov, M. et al. Nat. Biotech. 37:540-546 (2019).
4. Oxford Nanopore Technologies. Medaka. Software available at: https://github.com/nanoporetech/medaka

WF_1158(EN)_V2_11Jan2023