Laboratory protocol for nanopore sequencing of whole mitochondrial genomes

Interreg North Sea Region GEANS
European Regional Development Fund

EUROPEAN UNION

Genetic tools for Ecosystem health Assessment in the North Sea region

Laboratory protocol for nanopore sequencing of whole mitochondrial genomes

This research was supported as part of GEANS, an Interreg project supported by the North Sea Program of the European Regional Development Fund of the European Union.

Authors:
Saskia Oosterbroek and Reindert Nijland

Project:
GEANS

Date:
14 June 2023

Lead Partner:
Marine Animal Ecology, Department of Animal Sciences,
Wageningen University (MAE-WUR)

Contact:
reindert.nijland@wur.nl

How to cite this protocol?
Saskia Oosterbroek & Reindert Nijland, 2023.
Laboratory protocol for nanopore sequencing of whole mitochondrial genomes.
https://www.geans.eu/protocols/

Introduction

This protocol is based on experiments at Marine Animal Ecology, Wageningen University, aimed at retrieving mitochondrial genomes of a variety of species. Single species samples of brittlestar *Ophiotrix fragilis*, sand goby *Pomatoschistus minutus* and flat oyster *Ostrea edulis* were multiplexed, sequenced and assembled using this protocol (Table 1). Assemblies were considered successful when most mitochondrial genes were retrieved in a single contig. A mock community consisting of 10 marine species was also tested and successful for 5 out of 10 species (Table 2). This mock community consisted of equal concentrations of DNA from sea anemone *Cylista troglodytes*, sea snail *Crepidula fornicata*, crab *Necora puber*, crab *Maja brachydactyla*, annelid worm *Spirobranchus triqueter*, annelid worm *Platynereis dumerilii*, barnacle *Austrominius modestus*, sea urchin *Psammechinus miliaris*, sponge *Suberites ficus* and European plaice *Pleuronectes platessa*. This gives some insight on the possibilities and hurdles to be taken for the future sequencing of entire natural communities. When multiple single species samples need to be sequenced multiplexing them is a more appropriate approach. The genome size of the species has a big impact on the number of mitochondrial reads in the run, where a large genome has a relatively lower number of mitochondrial reads. The high number of split/duplicated genes in these draft genomes has two causes. The single species experiment in table 1 was run on a r9.4.1 flowcell, which has a higher error rate than the newer r10.4.1 flowcell type that was used in the mock community experiment. The mock community experiment however has a very low coverage for most species, also resulting in split genes.

Multiple DNA extraction kits were tested for the retrieval of whole mitochondrial genomes. Kits for the isolation of high molecular weight DNA (Monarch Plasmid Miniprep Kit and Gentra Puregene) did not prove to outperform DNeasy blood and tissue kit, based on the amount and length of mitochondrial reads in sequencing data. The quality of the sample is key in retrieving long mitochondrial reads. Tissues that have been stored for a longer period generally show a lower DNA quality and shorter read lengths after sequencing.

All sequencing was done with Nanopore technology on Mk1C using R9 and later R10 flow cells. When multiplexing multiple samples Ligation sequencing gDNA – native barcoding kit (SQK-LSK109) with native barcoding expansion kits (EXP-NBD104 and EXP-NBD114) is recommended over the rapid barcoding kit SQK-RBK004. The rapid barcoding kit cleaves DNA when attaching the barcoded adapters, which could result in shorter DNA fragments. Also, the overall sequencing yield is generally lower. This is offset by the much more rapid library prep, so what you choose can vary on the amount and quality of the samples and available time.

Precautions

Native sequencing with Nanopore technologies will sequence all DNA molecules present in a sample. Take care preventing contamination between samples and between samples and the lab environment. Consider that also off target DNA will be sequenced, such as part of the microbiome of the sample and possible parasites or symbionts present in the sample.

Phylum	*Species*	*reference(s) used at level*		*Assembly length*	*Assembly program*	*Coverage*	*Circular*	*Genes not found*	*Split/duplicated genes*
Chordata	*Pomatoschistus minutus*	Species	*Pomatoschistus minutus*	16377	Flye	17	yes		cox1, cob, atp6, nad2, nad4, nad5
Echinodermata	*Ophiothrix fragilis*	Superfamily	Ophiactoidea	15631	Canu	7	yes		cox1, cox3, atp6, atp8, nad4
Mollusca	*Ostrea edulis*	Species	*Ostrea edulis*	16336	Flye	14	yes		nad2, nad5, trnM

Table 1. Single species sequenced separately, run on a r9.4.1 flowcell. Run N50 was 1.2 kb. Reads were filtered for Q score of 8 and lengths between 250 - 30000 bases resulting in: *Pomatoschistus minutus* 191027 raw, 138062 filtered; *Ophiothrix fragilis* 148323 raw, 98013 filtered; *Ostrea edulis* 267894 raw, 155059 filtered reads.

Phylum	*Species*	*reference(s) used at level*		*Assembly length*	*Assembly program*	*Coverage*	*Circular*	*Genes not found*	*Split/duplicated genes*
Annelida	*Platynereis dumerilii*	Species	*Platynereis dumerilii*	15,710	Flye	3	yes		cox1, cob, atp6, nad1, nad6
Annelida	*Spirobranchus triqueter*	Genus	Spirobranchus			0			
Arthropoda	*Necora puber*	Superfamily	Portunoidea			0			
Arthropoda	*Austrominius modestus*	Order	Balanomorpha	15,344	Flye	3	yes		cox1, cox2, cox3, atp6, nad1, nad2, nad3, nad4, nad5, trnC, trnY
Arthropoda	*Maja brachydactyla*	Genus	Maya			0			
Chordata	*Pleuronectes platessa*	Species	*Pleuronectes platessa*	15,305	Flye	3	no	nad5, nad6	cox1, cox2, cob, nad1
Cnidaria	*Cylista troglodytes*	Genus	Cylista	4,837	Flye	3	no	cox1, cox2, cox3, cob, nad2, nad4, nad6, trnA, trnC, trnD, trnE, trnF, trnG, trnH, trnK, trnL1, trnM, trnP, trnQ, trnR, trnS1, trnS2, trnT, trnW, trnY, trnV, rrnS, rrnL	
Echinodermata	*Psammechinus miliaris*	Family	Parechinidae	15,719	Flye	24	yes	rrnL	
Mollusca	*Crepidula fornicata*	Order	Littorinimorpha			0			
Porifera	*Suberites ficus*	Genus	Suberites	22,169	Flye	4	no	trnA, trnV, rrnS	cox2, cox3, cob, nad2, nad4, nad5, trnM, trnS1

Table2. Mock community sample, run on a r10.4.1 flowcell. Run N50 was 7.4 kb. 165115 reads were generated of which 162124 were used after filtering for Q score 10 and lengths between 250 - 30000 bases.

Materials

Sequencing libraries are prepared following Nanopore's protocol "ligation sequencing DNA V14" (SQK-LSK114). When sequencing multiple samples at once (up to 24) you can instead use the Native Barcoding Kit 24 V14 (SQK-NBD114.24).

All materials, consumables and equipment are as following Nanopore protocols SQK-LSK114 or SQK-NBD114.24 if multiplexing is desired.

Equipment

MinION Mk1C sequencer

Consumables

- DNeasy Blood & Tissue Kit, Cat. No. / ID: 69506 Qiagen
- Ligation sequencing DNA kit (SQK-LSK114)
- For multiplexing: Native Barcoding Kit 24 V14 (SQK-NBD114.24)
- AMPure XP Reagent, magnetic beads, A63881 Beckman Coulter – AMPure XP beads are supplied with LSK 114 but if samples need to be multiplexed more beads may be needed.

Laboratory protocol

To prevent contamination, ensure all working materials and surfaces are clean.

Extract DNA from samples according to DNeasy Blood & Tissue Kit manufacturer protocol.
If sample quality and DNA yield is expected to be low consider performing the final DNA elution step in a lower volume of elution buffer (EB).

Perform quality checks on the isolated DNA. Measure DNA concentration on a Qubit fluorometer and run a few microliters on a 1% agarose gel to estimate fragment length.

DNA repair and end-prep (SQK-LSK114)/(SQK-NBD114.24)

Follow the DNA repair and end-prep protocol per individual sample following "Ligation sequencing DNA V14 (SQK-LSK114)". Using the DNA Control Sample (DCS) is not necessary. Smaller volumes of repair reaction can be used when the concentration of the extracted DNA is high, or to save costs. The following reaction volumes work well:

16 µl	DNA extraction at ~20 ng/ul
1.17 µl	NEBNext FFPE DNA Repair Buffer
0.67 µl	NEBNext FFPE DNA Repair Mix
1.17 µl	Ultra II End-prep Reaction Buffer
1 µl	Ultra II End-prep Enzyme Mix

After incubation in the thermocycler according to the protocol add the of resuspended AMPure XP beads (AXP) to the end-prep reaction tube. If this scheme is followed use 20ul of beads, if your reaction has a different end volume also adjust the volume of the beads clean-up after the incubation).

Optional *for multiplexing samples:*

Native barcoding ligation Native Barcoding Kit 24 V14 (SQK-NBD114.24)
Follow the native barcoding ligation protocol. The first steps are identical to the protocol without barcoding. However, when combining many samples, lower input amount is needed for each individual sample. An example of using less reagents is given below.

11.25 µl	~250 ng end-prepped DNA
1.25 µl	Native Barcode
12.5 µl	Blunt/TA Ligase Master Mix

If this scheme is followed also adjust the volume of the beads clean-up **after the incubation** and add 25 µl of resuspended AMPure XP beads (AXP) to the end-prep reaction tube.

At the end of the beads clean-up adjust the final elution volume according to the number of samples that are being multiplexed. Total input for the Adapter Ligation and clean-up is 60 µl.

Adapter ligation and clean-up (SQK-LSK114)/(SQK-NBD114.24)

A mixture of 1:2 Long Fragment Buffer and Short Fragment Buffer can be used to enrich fragments of >1 kb. Only apply this if the DNA quality was checked on agarose gel and quality and amount of DNA is sufficient.

Priming and loading the SpotON flow cell and **Ending the experiment** (SQK-LSK114)/(SQK-NBD114.24)

The rest of the protocol should also follow (SQK-LSK114) "Ligation sequencing gDNA or Native Barcoding Kit 24 V14 (SQK-NBD114.24)

Bioinformatics

The mitochondrial assembly script is available at https://github.com/Saskia-Oosterbroek/NanoMito

Required dependencies:

NanoFilt https://github.com/wdecoster/nanofilt (De Coster et al., 2018)
Minimap2 https://github.com/lh3/minimap2 (Li, 2018)
Samtools https://github.com/samtools/samtools (Li et al., 2009)
Flye assembler https://github.com/fenderglass/Flye (Kolmogorov et al., 2019)

Optional dependencies:

Guppy_basecaller https://community.nanoporetech.com/downloads
Canu assembler https://github.com/marbl/canu (Koren et al., 2017)

Adjusting Flye

Flye was developed for long read data and uses a minimum read overlap of 1000 bases. When assembling mitochondrial genomes, especially from degraded samples, shorter reads might still be valuable. The 1000 bases minimum can be circumvented by adjusting the "main.py" file in flye's installation folder.

Find the following piece of code:

```
parser.add_argument("-m", "--min-overlap", dest="min_overlap", metavar="int",
                    type=lambda v: check_int_range(v, 1000, 10000),
                    default=None, help="minimum overlap between reads [auto]")
```

Change (v, 1000, 10000) to (v, 100, 10000) and your minimum read length overlap is set to 100 bases.

Run mitochondrial assembly script:

Be sure required dependencies are installed. The script is not meant to be run as is, adjust file locations to your specific data locations. Between consecutive rounds of assembly, draft assemblies need to be curated manually and set as input for the next round.

Quality check

After successful mitochondrial assembly the quality can be checked by annotating the genome on Mitos webserver, http://mitos.bioinf.uni-leipzig.de (Bernt et al., 2013). The annotation will reveal any missing, split, or duplicate elements of the mitochondrial genome. These errors may be caused by mistakes due to Nanopore sequencing or low sequencing coverage. Using an r10 flowcell combined with enough coverage will resolve most of these issues, as seen in the mitochondrial genome of *Psammechinus miliaris* in table 2.

References

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., & Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, *69*(2), 313–319. https://doi.org/10.1016/j.ympev.2012.08.023

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics*, *34*(15), 2666–2669. https://doi.org/10.1093/bioinformatics/bty149

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, *37*(5), Article 5. https://doi.org/10.1038/s41587-019-0072-8

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, *27*(5), 722–736. https://doi.org/10.1101/gr.215087.116

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Interreg
North Sea Region
GEANS
European Regional Development Fund

EUROPEAN UNION

Genetic tools for Ecosystem health Assessment in the North Sea region

Interreg North Sea Region GEANS

European Regional Development Fund

EUROPEAN UNION

Genetic tools for Ecosystem health Assessment in the North Sea region