Genepi Molmed Hackathon

Welcome to the Genepi MinION Hackathon! Your task today is to analyse mtDNA data generated from Oxford Nanopore. The data has been generated using a transposase library prep kit on a mixture of two long-rage PCR amplicons, which each amplify about half of the mtDNA genome. The amplicons overlap by about 200 at each end. Mind that a transposase library prep on so short amplicons (about 8.5 kb each) produces rather short reads. These are rather untypical for nanopore sequencing at genome scale or when using a ligation library prep kit, which would specifically ligate the adaptors to the end of the amplicons and thus produce full-length amplicon reads.

Setup

- Create a Word Document with your last name (named: `SS24_NGS_Molmed_Lastname.docx`) and document what you do (e.g. commands, screenshots, graphics). This file must be sent at the end of the class to Sebastian.
- Login to `genepi-lehre.i-med.ac.at` and create a folder `project-day` within `genepi-teaching/students/<your-q-number>`. All experiments must me executed from here.
- Copy your barcode folder (`cp -r <your-barcode> .`) to your `project-day` folder. The files are located here: `~/genepi-teaching/ngs/2024-ngs-molmed`. The structure should look like below (e.g. for barcode04).
- After the `cp` has been finished, verify that fastq files are available within your folder.

Structure

```
└── project-day
    └── barcode04
        ├── file1.fastq.gz
        └── file2.fastq.gz
        └── ...
```

Quality Control

Task 1

Count the reads in your sequencing experiment by combining the commands `zcat *fastq.gz`, `|`, and `wc -l` and explain the command and output in your Word file. If you see almost no reads included (< 400), copy a different barcode to your `project-day` folder and redo Task 1.

Task 2

The second task is to run a QC analysis on your data. For Nanopore data, we use a tool called Nanoplot. The usage looks complicated but at the end, you only need the following options: `--fastq_rich <fastq-files>`, `--outdir <output-folder>`, `--maxlength 15000`).

```
NanoPlot --fastq-rich <your-fastq-files> ...
```

HINT: Use `*fastq.gz` if you have several files!

This tool outputs a HTML report, look for the report, download it to your local filesystem and see what information you can identify. Add some plots and information to your Word file!

Mapping

Now it's time to align the data using a tool called minimap2. Replace with your input-file (included in your barcode folder) and with the name of your actual barcode. The reference can be found here: `~/genepi-teaching/students/reference-data/chrM.fasta`

```
minimap2 -ax map-ont <reference> <your-fastq-files> | samtools sort >
<yourbarcode>.bam
```

HINT: Use `*fastq.gz` for `<your-fastq-files>` if you have several files!

Copy the command into your doc-file and describe what the command is doing (e.g. `minimap2`, `|`, `samtools sort`).

Visualize Mapping

Below we execute a new command that calculcates the coverage for each position. Run `samtools depth` on your BAM file. The output is written on the command-line by default. Adapt the command and write the result into a file (using `>`), transfer the file to Windows and visualize the result.txt file with Excel or R.

```
samtools depth -a <yourbamfile>
```

Variant Calling

Now we run the actual variant calling with a tool created specific for mitochondrial genomes.

```
/opt/tools/mutserve/mutserve call <your-bam.bam> --reference <reference> -
-level 0.1 --output <output-name>.vcf.gz
```

Questions:

- What is `--level 0.1` doing? Try different parameters.
- Open the vcf file with `zcat` and see what information you can identify. Discuss it in your Word file. The VCF file format is explained in the slides or also here.

Haplogroups

mtDNA profiles can also be grouped into so called haplogroups. Download the file to your local PC and upload the VCF File to https://haplogrep.i-med.ac.at.

Questions:

- Interpret the results (use the Haplogrep docs) and add information to your Word file.

PCR Products