Molecular Microbiology 2024

Protocol 6

Assembling a prokaryotic genome from Oxford Nanopore Sequencing data

This protocol covers the assembly of prokaryotic genomes using trimmed, filtered, long-read sequencing data. It uses the following software tools:

- UniCycler - tested with version 0.5.0
- Flye - tested with version

1. Log into your server using **ssh** or **MobaXterm** as described in protocol 1, for example in ssh:

   ```
   ssh molmicroX@dnaseq1a.bio.au.dk
   ```

2. Change into the genome directory where you have saved your trimmed and filtered reads:

   ```
   cd genome
   ```

3. Run the assembly tool `Flye` to assemble your genome:

   ```
   screen -L \
   flye \
   --nano-hq inputfile_trimmed_filtered.fastq.gz \
   --genome-size Xm \
   --threads 3 \
   --out-dir flye_assembly
   ```

   - `flye` - This is the name of the program to be run.
   - `--nano-hq inputfile_trimmed_filtered.fastq.gz` - This parameter specifies the name of the file containing the trimmed and filtered reads for assembly. Note that this parameter is specifically for "Q20" high quality Nanopore reads made with recent chemistry and flow cells - different options exist for different kinds of reads.
   - `--genome-size Xm` - This parameter specifies the expected approximate size of the genome. `X` should be replaced with expected size in megabases (e.g. 5.3m).
   - `--threads 3` - This specifies the number of CPU threads to be used (in this case 3).
   - `--out-dir flye_assembly` - This parameter specifies the output directory and will need to be changed if it's run more than once.

4. Run the assembly tool `unicycler` to assemble your genome - perhaps a different tool will result in a better assembly?:

   ```
   screen -L \
   unicycler \
   --long inputfile_trimmed_filtered.fastq.gz \
   --out unicycler_assembly \
   --threads 3
   ```

- `unicycler` - This is the name of the program to be run.
- `--long inputfile_trimmed_filtered.fastq.gz` - This specifies the long reads to be used in the assembly. Other read types (for example short reads) can be included with other input flags.
- `--out unicycler_assembly` - This parameter specifies the output directory and will need to be changed if it's run more than once.
- `--threads 3` - This specifies the number of CPU threads to be used (in this case 3).