Plasmid Sequencing Analysis Report

Research Use Only (RUO)

10 May, 2023

Eurofins Order ID: **1000001**
Sample Name: **Sample01**
Sample Received on: **9 May, 2023**
Sample Processed on: **9 May, 2023**
Sample Analyzed on: **10 May, 2023**
Technology Used: **Oxford Nanopore Technology (ONT) Sequencing**
Pipeline: **Plasmid Analysis Pipeline v1.0**
Report Version: **v1.0**

Eurofin's proprietary Nanopore data analysis pipeline is used to prepare and sequence samples with Oxford Nanopore Technologies sequencers, which utilize a third-generation sequencing technology capable of real-time long-read sequencing of DNA. The technology involves feeding a single-stranded DNA molecule through a protein nanopore and measuring changes in electrical current as the DNA passes through. The resulting reads are then subjected to quality filtering, assembly, and annotation using the Nanopore data analysis pipeline developed by Eurofins.

Results

Plasmid Assembly

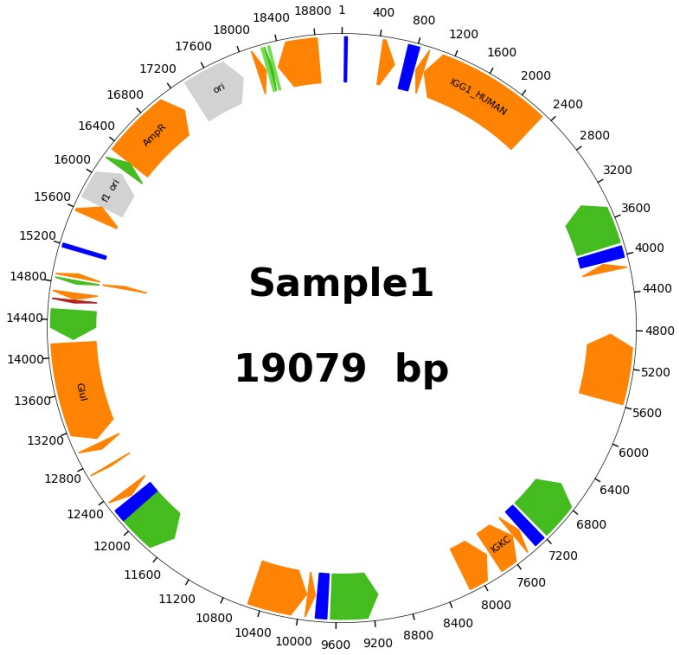
The assembly statistics of the assembled plasmid is shown in the following table -

SAMPLE NAME	PLASMID LENGTH	GC %	TOTAL READS	COVERAGE DEPTH
Sample01	19079	50.96	1811	1622

- SAMPLE NAME column represents the name of sample processed.
- PLASMID LENGTH column represents the size of the assembled plasmid in basepairs (bp).
- GC % column represents percentage GC content of the assembled plasmid.
- TOTAL READS column represents the number of reads in plasmid assembly.
- COVERAGE DEPTH column represents the average coverage depth of the assembled plasmid.

Plasmid Map

The assembled plasmid is represented graphically showing the locations and types of genetic features such as genes, promoters, restriction enzyme sites, and other functional elements as a plasmid map. Typically, a circular plasmid map is shown with the DNA sequence depicted as a circle, with the start and end points of the sequence joined together. Plasmid maps may also include labels indicating the names or functions of the genetic features and their relative positions on the plasmid.



Plasmid Features Table

The features annotated and displayed in the plasmid map are shown in the table below. The full annotations, including nucleotide sequence and descriptions of each feature are found in the deliverables.

Plasmid Clonal Purity Check

To check the assembled plasmid purity, the sequenced reads are mapped against the assembled plasmid and the variants (SNPs, insertions & deletions) are determined. Variants detected with at least 10% minor allele frequency and >30x read support are shown below

```
No variants detected.
```

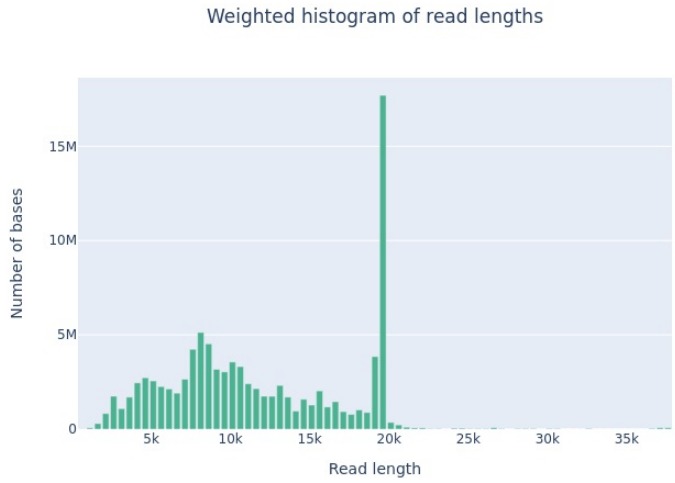
Deliverables

The **ORDERID.SAMPLE.plasmid_analysis.zip** archive contains the following files:

1. **SAMPLE.plasmid_analysis_report.html**: This is the analysis report.
2. **SAMPLE.plasmid_assembly.fasta**: Assembled plasmid FASTA sequence.
3. **SAMPLE.plasmid_annotations.gbk**: Annotated plasmid sequence in GENBANK format.
4. **SAMPLE.plasmid_annotations.csv**: Annotated feature table, including the nucleotide sequence of each feature.

Sequencing Reads QC

Distribution of read lengths from sequenced data is shown in the following histogram. The histogram displays the number of sequenced bases (bp) on the y-axis and the read length on the x-axis. Each bar in the histogram represents a range of read lengths, and the height of the bar indicates the total number of bases (bp) falling within that range. This results in a weighted plot by the number of nucleotides per bin, as longer reads carry more weight in the histogram. Read length histograms can be used to assess the quality of sequencing data, as the distribution of read lengths can indicate the presence of contaminants or biases in the sequencing process. They can also be used to determine the size of the plasmid being sequenced.



In case of plasmid mixtures with several read length peaks, the pipeline will only report the plasmid sequence in the major peak.

REMARKS

Plasmid sequencing using nanopore technology has some limitations. One limitations is the relatively high error rate associated with long-read nanopore sequencing in comparison to short read sequencing technologies, which can lead to errors in the assembled sequence. Additionally, nanopore sequencing can be sensitive to sequencing errors, particularly in homopolymer regions, which can affect the accuracy of the sequencing data. The quality of input DNA is a very important factor that can influence the accuracy of generated sequence data. Any impurities in the DNA sample can significantly affect the accuracy of the sequencing data, which may result in failure of plasmid assembly reconstruction.

DISCLAIMER

The results presented and delivered are generated by following best practices available for nanopore sequencing of plasmids. Before interpretation of the results, customers are advised to inspect the results thoroughly and consider the technological and bioinformatical limitations carefully. This report and the provided deliverables are for research use only (RUO).
