Plasmid Sequencing Analysis Report

Research Use Only (RUO)

22 January, 2025

Eurofins Order ID: **11109208324**
Sample Name: **Sample1**
Plasmid Length: **2.5-25KB**
Sample Received on: **22 January, 2025**
Sample Analyzed on: **22 January, 2025**
Technology Used: **Oxford Nanopore Technology (ONT) Sequencing**
Pipeline: **Plasmid Analysis Pipeline v3.1.0**
Report Version: **v3.1**

Eurofins 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
Sample1	3197	50.10948	3167	2847

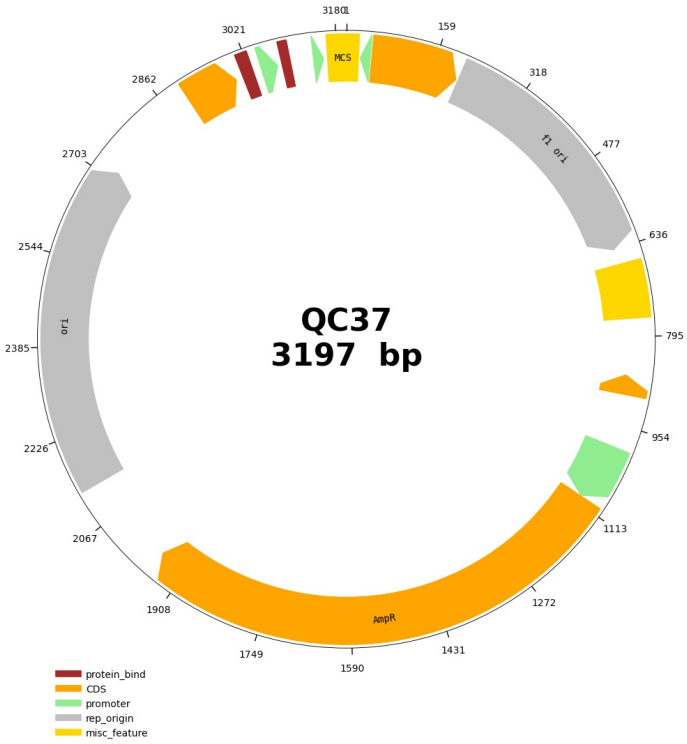
- SAMPLE NAME column represents the name of sample processed.
- PLASMID LENGTH column represents the size of the assembled plasmid in basepairs (bp).
- GC % column represents percentage GC content of the assembled plasmid.
- TOTAL READS column represents the number of reads in plasmid assembly.
- COVERAGE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

The process of plasmid assembly from Nanopore sequencing data is a complex task that requires careful optimization to ensure accuracy. Eurofins proprietary Nanopore data analysis pipeline is designed to enhance the assembly of plasmids, especially when dealing with challenging sequences that include repetitive elements or a mix of full-length and partial reads. Despite the advanced technology, there are instances where manual intervention is necessary. When automated assemblies do not align with expectations, a meticulous manual review of the data quality, coupled with parameter adjustments, can often lead to successful reconstruction of the plasmid sequence. For situations where significant discrepancies are observed, Eurofins offers an “Assembly Review” service (https://eurofinsgenomics.eu/en/next-generation-sequencing/ngs-additional-services). This service provides a thorough manual examination and reassembly of the plasmid using the original sequencing data, ensuring that the final results meet the anticipated outcomes. Providing details such as the expected plasmid size enhances the review process, allowing for a more targeted

approach to resolving assembly issues.

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.

Copy CSV Excel PDF Print

Show 10 entries Search:

Feature	Type	Start Location	End Location	Percent Identity
f1 ori	rep_origin	207	636	100
AmpR promoter	promoter	994	1099	100
AmpR	CDS	1099	1960	99.768
ori	rep_origin	2130	2719	99.83
MCS	misc_feature	3163	23	100
lac promoter	promoter	3042	3073	100
CAP binding site	protein_bind	3006	3028	100
T7 promoter	promoter	25	44	100
SP6 promoter	promoter	3138	3157	100
lac operator	protein_bind	3080	3097	100

Showing 1 to 10 of 14 entries Previous 1 2 Next

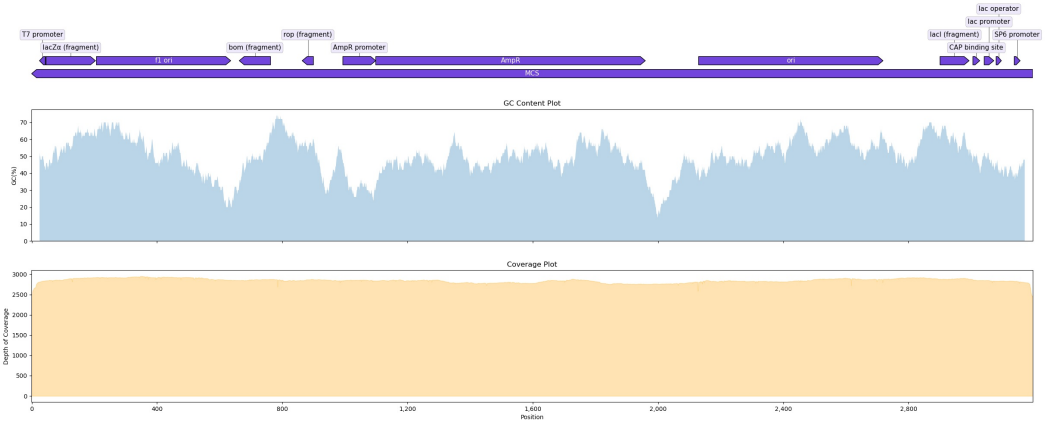
Plasmid Clonal Purity Check

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 0.3 minor allele frequency(MAF) and >20x read support are shown below

No variants detected.

Plasmid Coverage Plot

The coverage and GC content of the assembled plasmid is shown below. If variants are observed, they are also plotted below along with the annotated features.



Variants not found

Deliverables

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

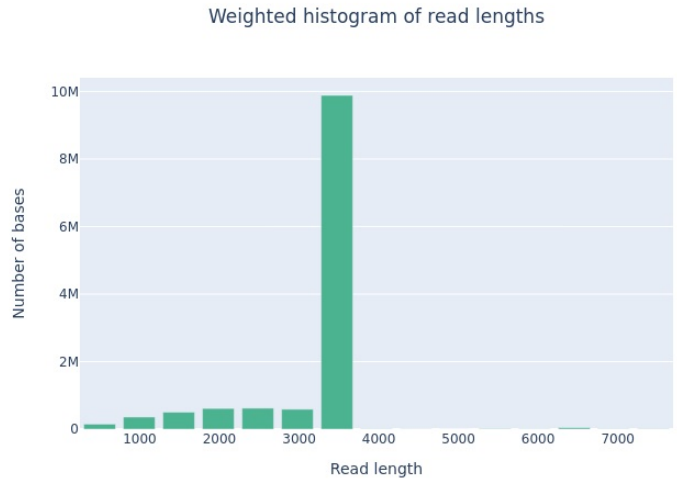
1. **SAMPLE.plasmid_analysis_report.html**: This is the analysis report.
2. **SAMPLE.plasmid_assembly.fasta**: Assembled plasmid FASTA sequence.
3. **SAMPLE.plasmid_assembly.fastq**: Assembled plasmid sequence with quality scores for each nucleotide position.
4. **SAMPLE.plasmid_assembly_per_base_qualities.csv**: Per base Quality with confidence score.
5. **SAMPLE.plasmid_annotations.gbk**: Annotated plasmid sequence in GENBANK* format.
6. **SAMPLE.plasmid_annotations.csv**: Annotated feature table, including the nucleotide sequence of each feature.
7. **SAMPLE.rawdata.fastq.gz**: Raw Sequencing data^

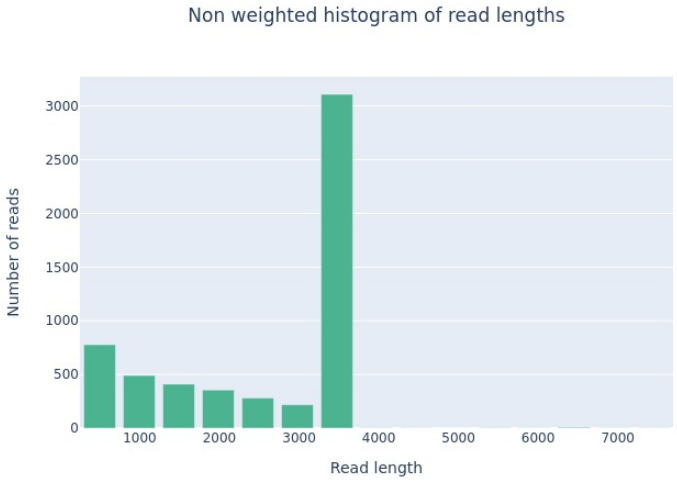
*Use this file to import the plasmid sequence and features in software like SnapGene, Geneious Prime, Benchling, SeqBuilder Pro etc.

^This file provides sequencing raw data of reads that align to the assembled plasmid consensus sequence. During this process, any non-target sequences, such as host genomic DNA, low-abundance molecular species, and other irrelevant sequences, are filtered out. Consequently, the remaining reads represent the target sequence of interest and can be used for further downstream analyses.

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). Please see the interpretation guide https://eurofinsgenomics.eu/en/custom-dna-sequencing/eurofins-services/whole-plasmid-sequencing/wps-data-interpretation/ and FAQs (https://eurofinsgenomics.eu/en/eurofins-genomics/product-faqs/sanger-sequencing/whole-plasmid-sequencing/)