The Report of Nanopore for Full-Length Sequencing

1. Clean data QC

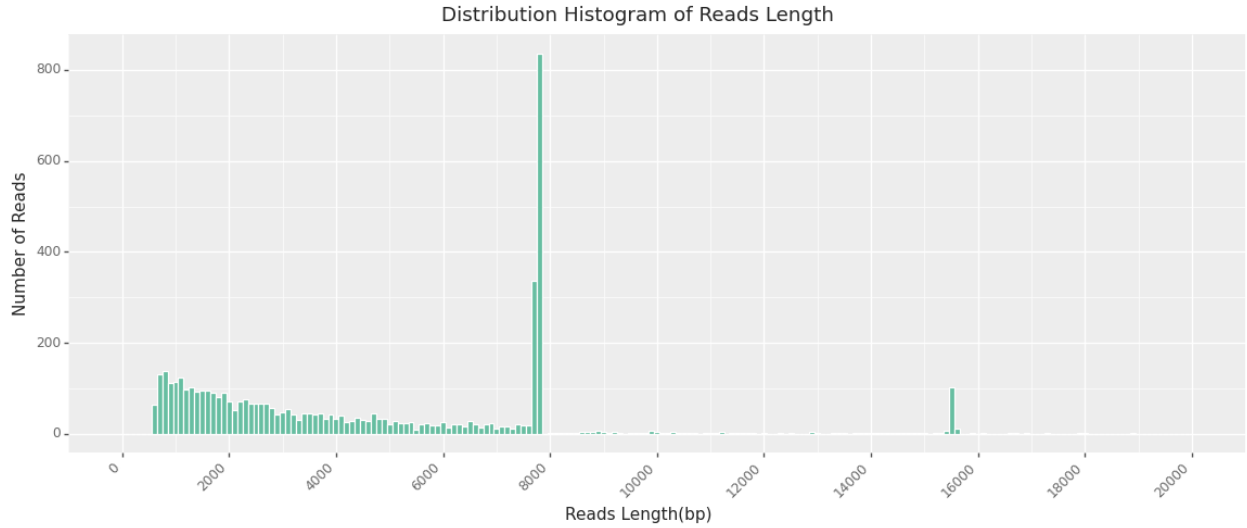
The raw data obtained from Nanopore sequencing undergoes quality control processes, such as trimming residual adapter sequences, length filtering, and sequencing quality filtering to generate clean data. This clean data is used for further data analysis.

- **Total Reads:** Total number of reads;
- **Total Bases:** Total number of bases;
- **Read Length N50:** The reads in the clean data are sorted in descending order of length, and the base counts are summed up. When the cumulative base count reaches half of the total bases, the length of the current read is determined as the N50 value. This value is primarily used to evaluate the read lengths in the result;
- **Mean read length:** Average length of reads(bp);
- **Mean read quality(Q score):** Average quality value of bases in reads;

SampleID	CloneID	Total Reads	Total Bases	Read Length N50	Mean read length	Mean read quality
C872ALLTG0-1	PF62489	4833	23128200	7747	4785	17.7

2. Distribution of read length

A distribution plot of read lengths in sequencing results. The x-axis represents the read lengths, while the y-axis represents the number of reads. The bins represent the length intervals of the histogram.



3. Clean data mapping result

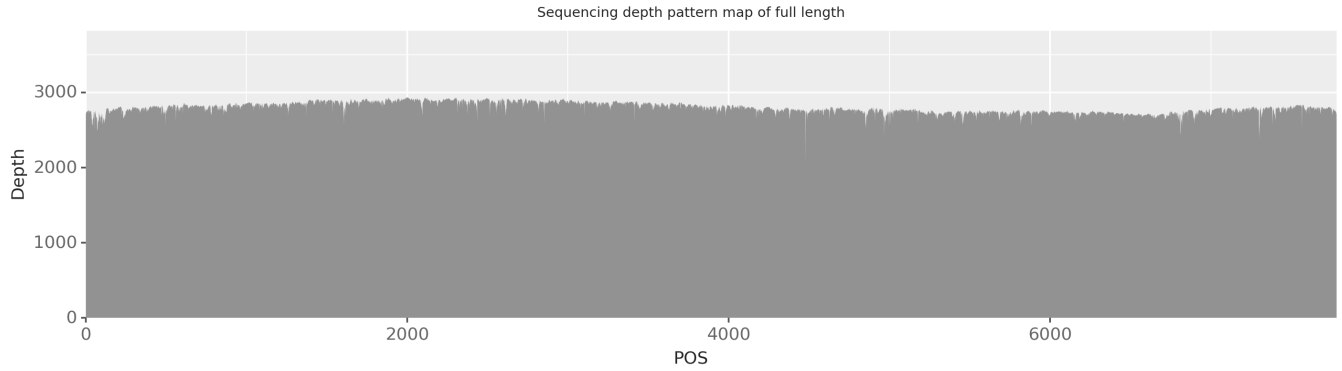
The clean data is aligned to a reference sequence using minimap2, and the results are analyzed to obtain information such as coverage and depth.

- **RefLen(bp):** Length of the reference sequence;
- **Map Ratio:** The proportion of bases in the clean data that align to the reference sequence;
- **Avg depth:** The average number of times a given nucleotide in a DNA sequence has been sequenced;
- **Median depth:** The median number of times a given nucleotide in a DNA sequence has been sequenced;
- **Coverage:**Proportion of bases with depth greater than 0x in reference;
- **Cov 30x:** Proportion of bases with depth greater than 30x in reference;
- **Cov 100x:** Proportion of bases with depth greater than 100x in reference;
- **MutationCounts:**Total number of SNP(Single Nucleotide Polymorphism)，Indel(Insertion, deletion)，SV(Structural Variation);
- **Result:** Based on the sequencing data, the final conclusion is derived by evaluating the sample coverage depth, and the consistency between the observed sample sequence and the theoretical sequence;

SampleID	RefLen	Map Ratio	Avg depth	Median depth	Coverage	Cov 30x	Cov 100x	MutationCounts	Result
C872ALLTG0-1	7783	97.93%	2805	2801	100.0%	100.0%	100.0%	0	Pass

4. Mapping Result

The pattern diagram of the sequencing results aligned to the reference sequence is shown below. The X-axis represents the position of bases in the reference sequence, and the Y-axis represents the sequencing depth at each position.



5. Mutations

This section displays the differences between the actual sequencing results and the reference sequence, primarily including SNP and Indel (insertions or deletions with a length of 35 bases or fewer). If the table contains only a "-", it indicates that no credible variations were detected in sample.

- **POS:** Locations of mutations occurring in the reference sequence;
- **REF:** The bases in the reference sequence at the positions where the mutations occur;
- **ALT:** The bases in the reference sequence after the occurrence of mutations at specific positions;
- **AF(Allele frequency, %):** Allele frequency refers to the proportion of reads in which a mutation occurs at a specific position, out of all the reads sequenced at that position;

5.1 High-confidence sites

These are confident mutations as there are no specific structures near these types of mutations, and the allele frequency is higher.

SampleID	POS	REF	ALT	AF
-				

5.2 Low-confidence sites

These types of mutations occur in low complexity regions of nucleic acids, such as oligonucleotide regions, and have a low allele frequency. Consequently, these sites have a low level of confidence.

SampleID	POS	REF	ALT	AF
-				

6. Structural Variation(SV)

This section displays structural variations, which are large-scale Indels (insertions or deletions with a length of more than 35 bases). The sequences of inserted and deleted bases are annotated in the attached .dna file. If the table contains only "-", it indicates that no credible structural variations were detected in the tested plasmid.

- **POS:** Locations of structural variations (SVs) occurring in the reference sequence;
- **SVTYPE:** Types of structural variants, insertions(INS) or deletions(DEL);
- **SVLEN:** It indicates that the corresponding number of bases is insert/deleted at that position;
- **AF(Allele frequency, %):** The frequency of SV occurrence refers to the proportion of reads with SV at a specific position out of all the reads obtained;

SampleID	POS	SVTYPE	SVLEN	AF
-				