QUICK START GUIDE

PCR-EZ
Long-Read PCR Sequencing

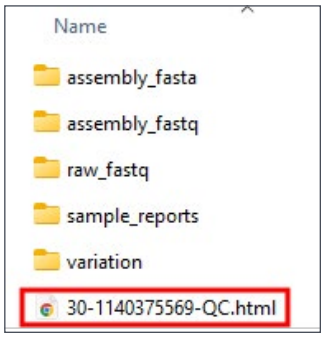
GENEWIZ®
From Azenta Life Sciences

genewiz.com

Table of Contents

1. Getting Started

To begin, click on the 30-xxxxxxxxx.QC.html file. This file provides an overall QC report.



VIEWING PROJECT REPORT

The result summary provides a table with a simple pass/fail assembly report and sequence length. The example below shows three samples successfully assembled for contigs, with sequence lengths 1181, 1554 and 1611.

1 ANALYSIS RESULTS

1.1 Results summary

Table 1.1 shows the assembled consensus sequence summary for all samples in this order.

Raw reads were cleaned, downsampled, assembled, polished and trimmed (remove low coverage regions), resulted in the final consensus sequence.

Note that the consensus sequence will miss up to 25 bp at the beginning and/or end of the linear/PCR product.

Table 1.1: Summary of the analysis results

Project	Sample	Assembly Status	Consensus sequence length
30-1140375569	BI1	Passed	1181
30-1140375569	BI10	Passed	1554
30-1140375569	BI11	Passed	1611

Following the analysis report, the raw data summary outlines key statistics for the raw reads. The table lists the total number of reads, total bases, and the minimum and maximum read lengths. In most cases, the total number of raw reads correlates with the sample's concentration and size.

1.2 Raw data summary

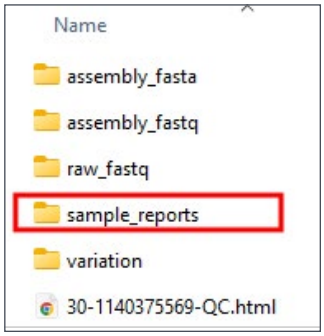
Table 1.2 shows the summary of the raw data.

Table 1.2: Summary of the raw reads statistics

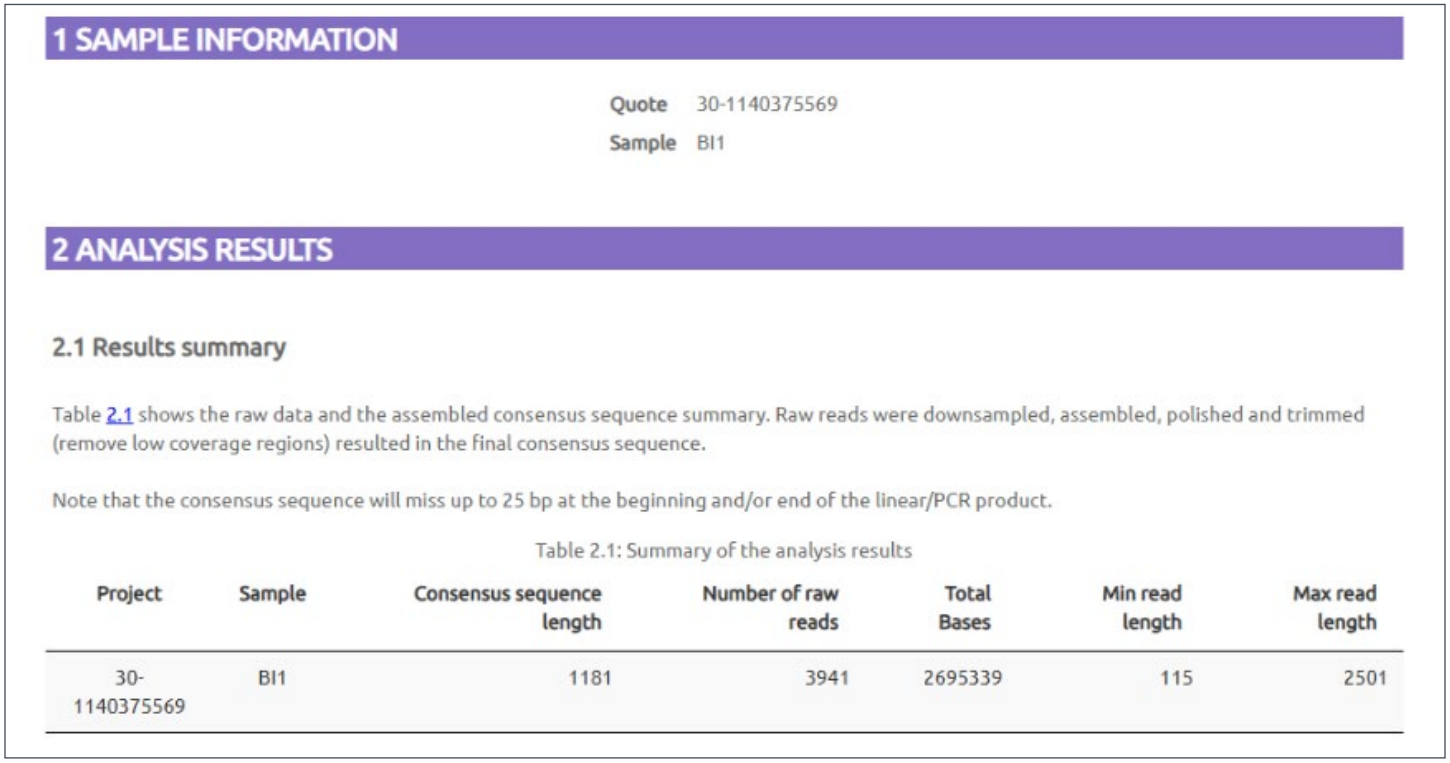
Project	Sample	Number of raw reads	Total Bases	Min read length	Max read length
30-1140375569	BI1	3941	2695339	115	2501
30-1140375569	BI10	1488	1114701	106	4042
30-1140375569	BI11	1394	1021351	101	3792

VIEWING SAMPLE REPORT

Open the sample reports folder to assess the individual report files by clicking on the sample name.



The sample report includes sample information, analysis results, and data quality control metrics. The analysis results section summarizes the assembled sequences and raw reads, while the data quality control section shows the distribution of raw reads by length and quality. The assembly QC dot plot visually represents the reads used for assembly. Additional details on these tables are provided in Section 3, *Assessing the Quality of Your Sample*.



1 SAMPLE INFORMATION

Quote 30-1140375569
Sample BI1

2 ANALYSIS RESULTS

2.1 Results summary

Table 2.1 shows the raw data and the assembled consensus sequence summary. Raw reads were downsampled, assembled, polished and trimmed (remove low coverage regions) resulted in the final consensus sequence.

Note that the consensus sequence will miss up to 25 bp at the beginning and/or end of the linear/PCR product.

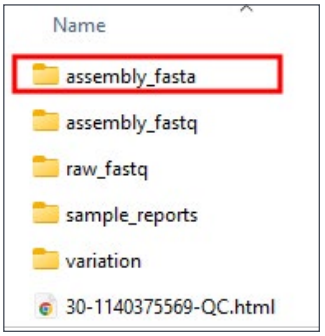
Table 2.1: Summary of the analysis results

Project	Sample	Consensus sequence length	Number of raw reads	Total Bases	Min read length	Max read length
30-1140375569	BI1	1181	3941	2695339	115	2501

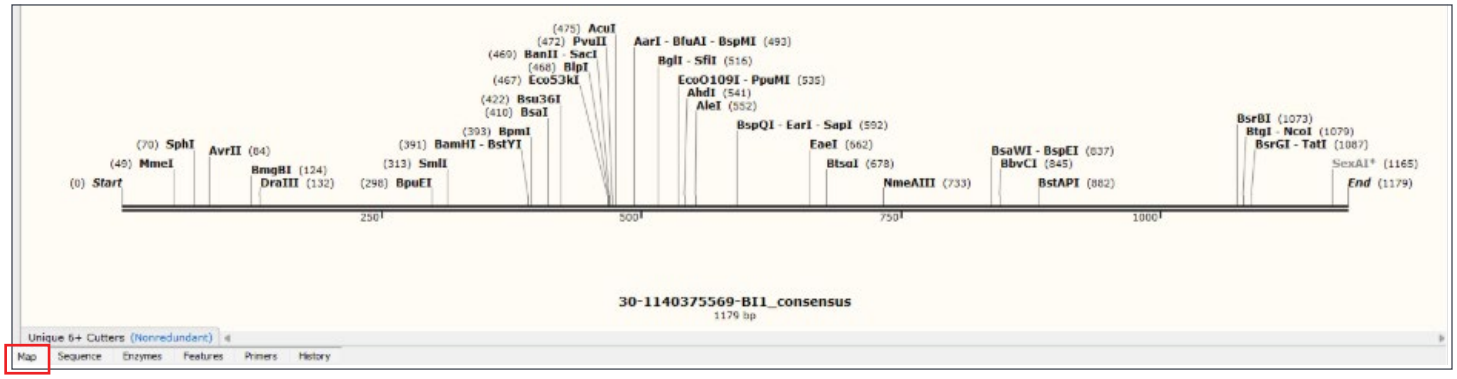
2. Viewing Your PCR Sequence

VIEWING FASTA FILES

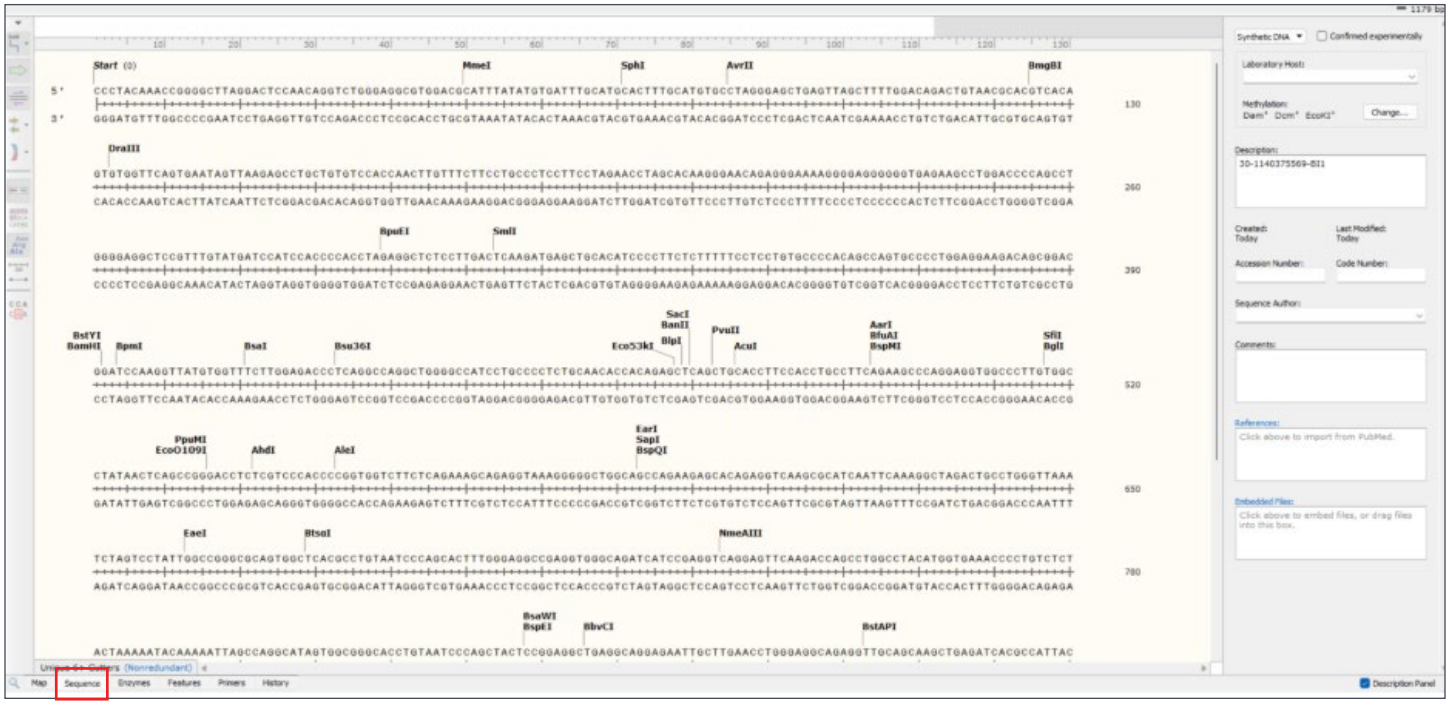
To view the assembled PCR sequence, open the assembly_fasta folders shown to the right.



SnapGene Viewer or your preferred fasta viewer can be used to open and review the sequence. The assembly can be viewed in the *Map* tab (highlighted in the red box) to review the total assembled length and restriction sites, as seen below.



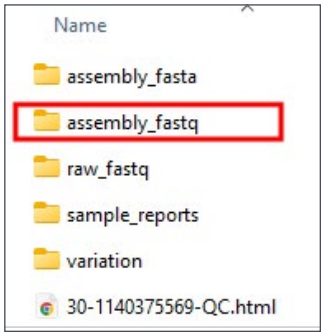
Clicking the sequence tab (highlighted in a red box below) will open the nucleotide sequence.

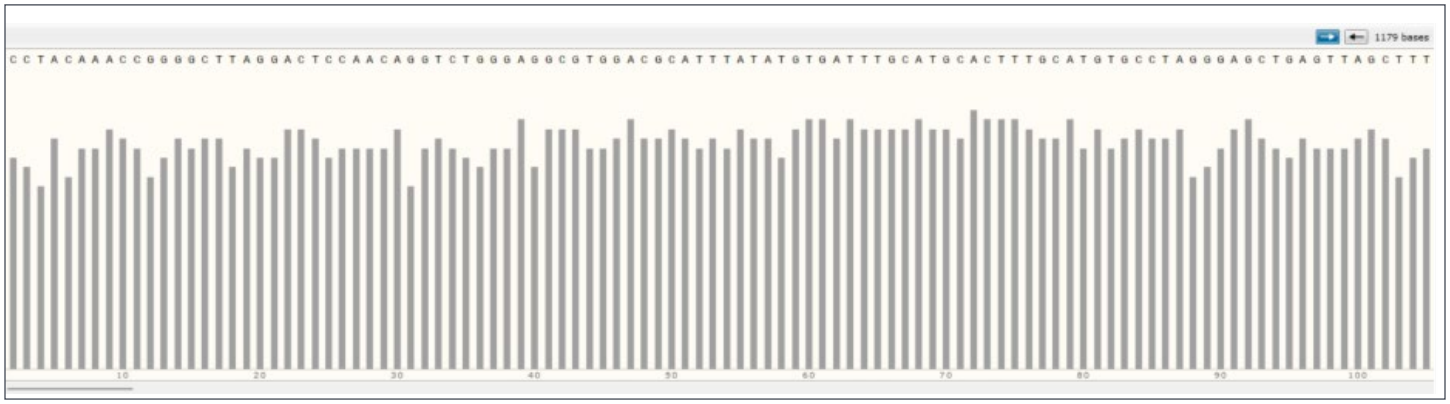


VIEWING FASTQ FILES

We also provide a FASTQ file with a confidence Q score per base that can be viewed in SnapGene Viewer or a similar program.

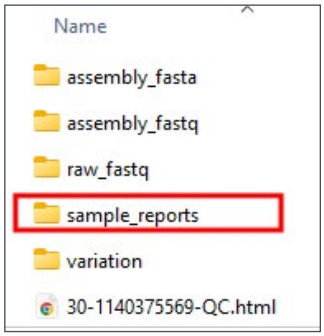
The higher the bar, the higher the confidence for the base call at this position. If a bar is lower, it could indicate either low-quality sequence data or the presence of a polymorphism at the site.





3. Assessing the Quality of Your Sample

The sample report provides detailed visuals to assess the quality of each sample.



VIEWING THE READ-LENGTH DISTRIBUTION

The data quality control section displays the distribution of raw reads by length and quality. The graph below illustrates the length distribution of the raw reads. Ideally, this distribution should show a comparable read count across lengths up to the expected total length. In this example, the graph depicts the raw read distribution for a 1,300 bp PCR fragment.

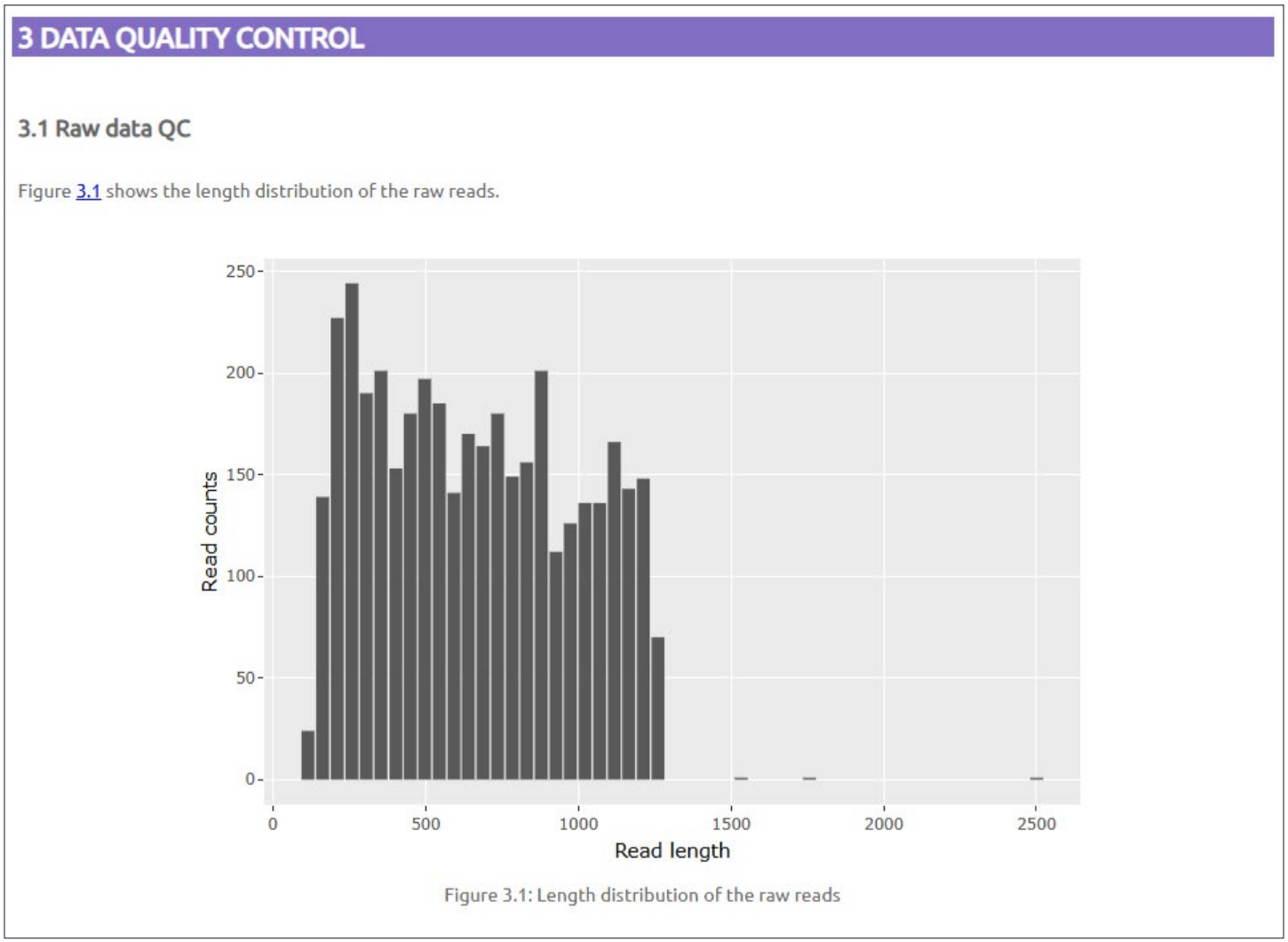


Figure 3.1: Length distribution of the raw reads

VIEWING THE QUALITY DISTRIBUTION

An example of the quality distribution of raw reads is shown below. An ideal graph should show peak read counts toward the right, indicating majority of the reads are high quality in the sample.

Figure 3.2 shows the quality distribution of the raw reads.

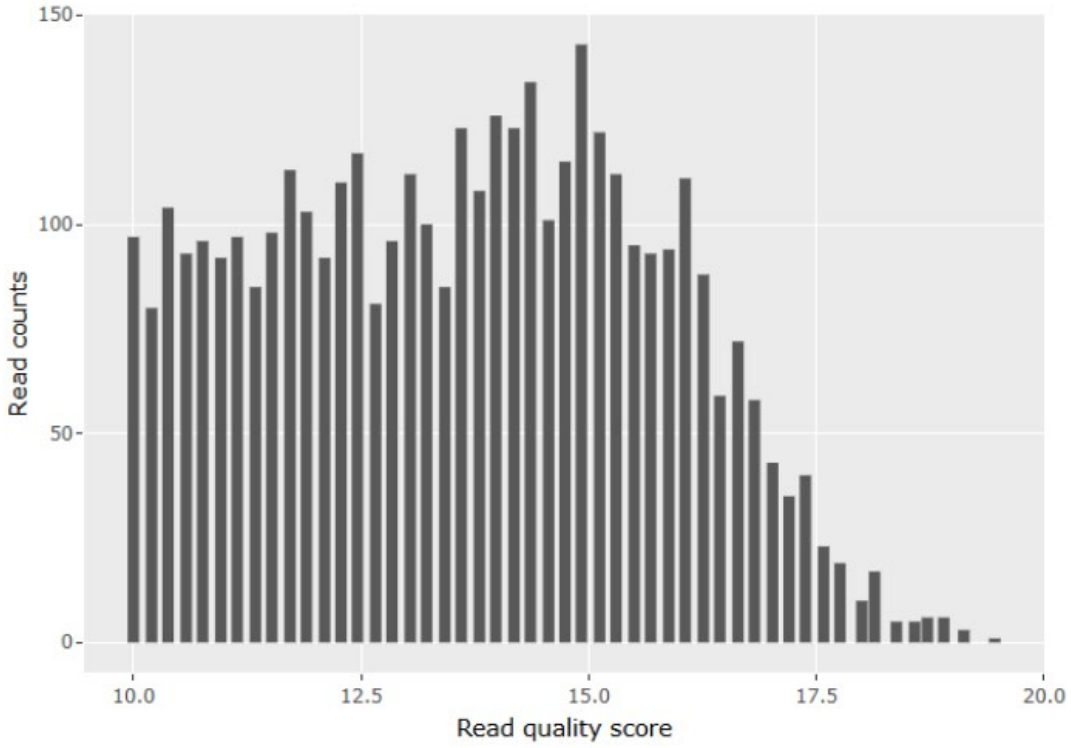


Figure 3.2: Quality distribution of the raw reads

ASSESSING THE COVERAGE BY BASE POSITION

The Assembly QC graph displays the coverage across the assembled consensus sequence. Typically, coverage is lower at the fragment ends and higher through the central region. In the example below, positions ~50–1,100 have coverage scores above 400, while both ends show slightly reduced coverage. Any dips within the central region would indicate decreased coverage at those positions.

3.2 Assembly QC

Reads used for assembly and polishing were aligned to the final consensus sequence. Figure 3.3 shows the coverage along the polished consensus sequence.

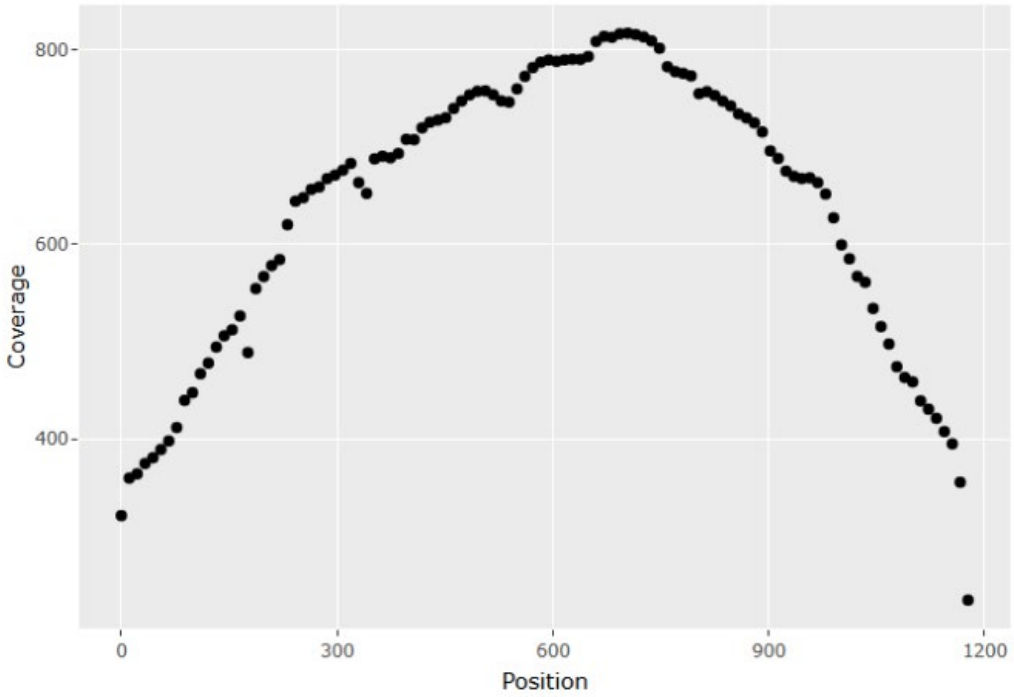
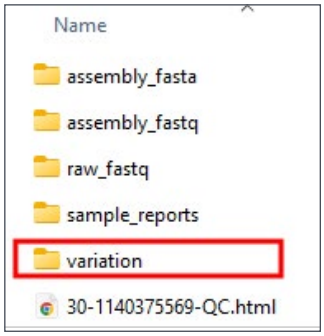


Figure 3.3: Coverage of the polished consensus sequence

4. Variant Calling

Any bases with a second nucleotide represented in >10% of reads are noted in the vc.tsv files. The example shown below shows 4 of the positions in the assembly have alternative base calling. The column labeled AD shows the ratio of the read counts between the called and alternate base at the position. The AF column provides the percentage of mixed base calling at the position.



CHROM	POS	REF	ALT	GT	DP	AD	AF
30-1140375569-BI1	218	A	G	0/1	580	439,136	0.2345
30-1140375569-BI1	545	T	C	0/1	750	349,380	0.5067
30-1140375569-BI1	551	C	T	0/1	758	596,160	0.2111
30-1140375569-BI1	1009	C	T	0/1	594	299,282	0.4747

The base_count.tsv file lists the number of reads for each base at every position. Any position with insertions or deletions is also shown with the number of read counts for said insertion or deletion.

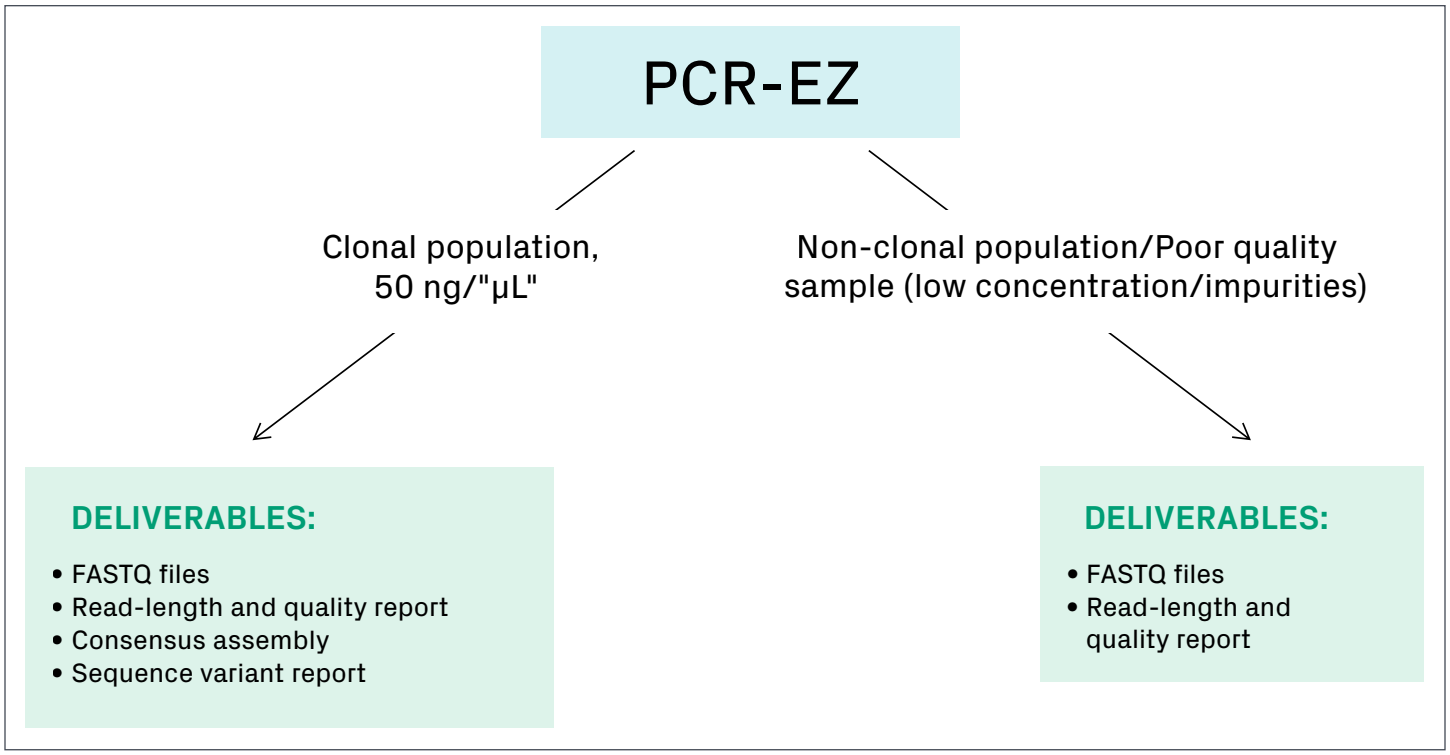
Position	Reference	Coverage	A	T	G	C	N	Insertions	Top Insert	Deletions
1	C	216	0	0	0	216	0	0	-	0
2	C	275	0	0	0	275	0	0	-	0
3	C	311	0	0	0	311	0	0	-	0
4	T	317	1	311	2	3	0	3	G (3)	0
5	A	337	337	0	0	0	0	0	-	0
6	C	341	0	0	0	341	0	1	A (1)	0
7	A	348	348	0	0	0	0	0	-	0
8	A	348	348	0	0	0	0	0	-	0
9	A	349	348	0	0	0	0	0	-	1
10	C	349	1	0	0	345	0	0	-	3

5. Keys for Successful Assembly

CONCENTRATION AND CLONAL DNA

To provide a fast turnaround time at a low cost, we do not perform sample QC to determine why samples failed assembly. However, the most common reason for failure is the sample not meeting the required 50 ng/ul concentration. Low concentrations may lead to increased fragmentation during library preparation and/or a low number of reads generated for the sample. We strongly recommend verifying sample concentration and using clonal DNA to minimize the risk of failure.

PCR-EZ also requires clonal DNA, as a mixed population may lead to failure. For failed samples, the raw fastq files are provided so you can perform additional analysis.



FAILED ASSEMBLY

In the event the sample fails to produce an assembly, the order folder will only contain the raw FASTQ reads and order QC report that includes the summary of the raw reads statistics.