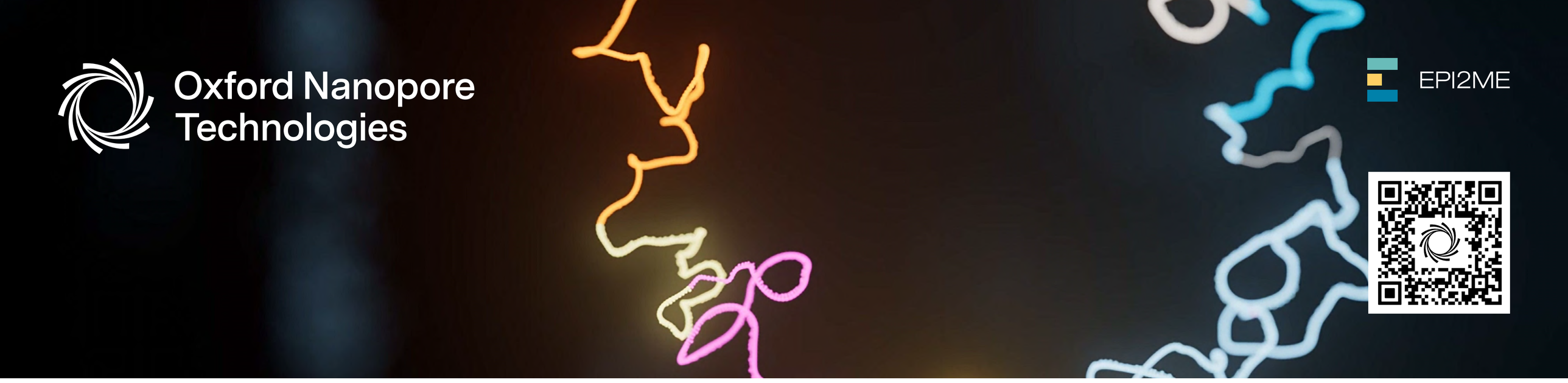


Validation and quality control of a molecular cloning experiment using *de novo* assembly of Oxford Nanopore reads

wf-clone-validation — for rapid verification of artificial constructs such as plasmids, bacterial artificial constructs (BACs), and fosmids

Contact: **support@nanoporetech.com**. More information can be found at: **labs.epi2me.io/workflows/wf-clone-validation/**.
Data used in this analysis is available to download from: **epi2me.nanoporetech.com/plasmid_2025.04/**

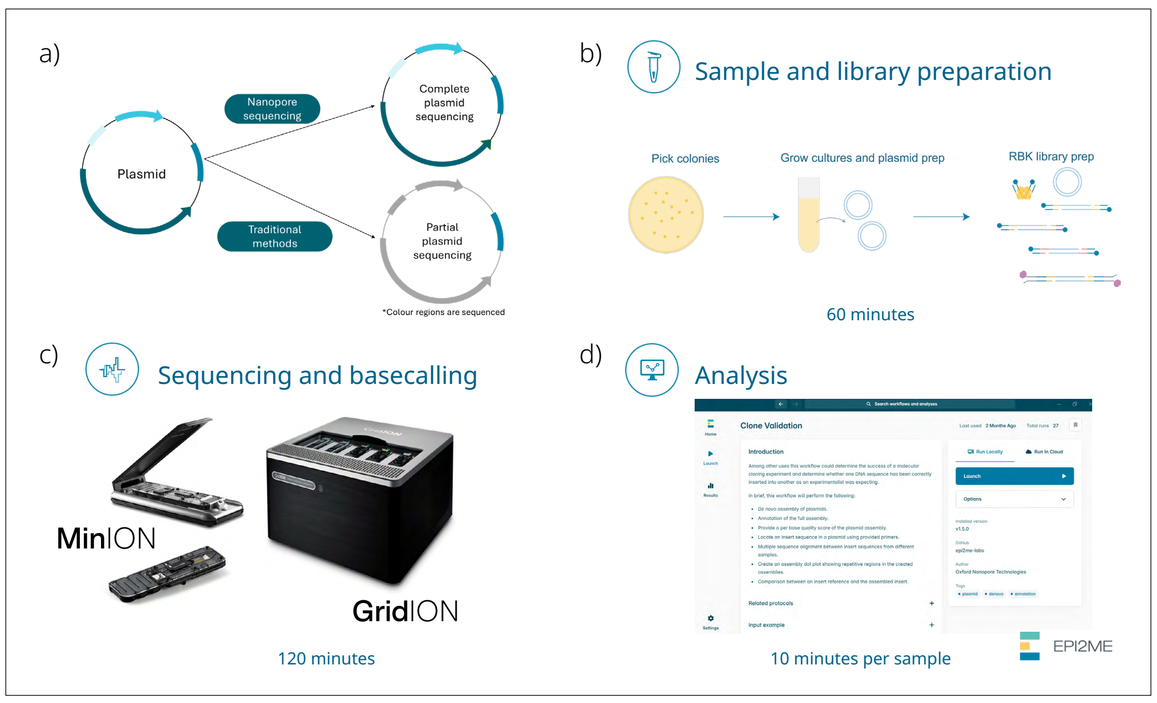
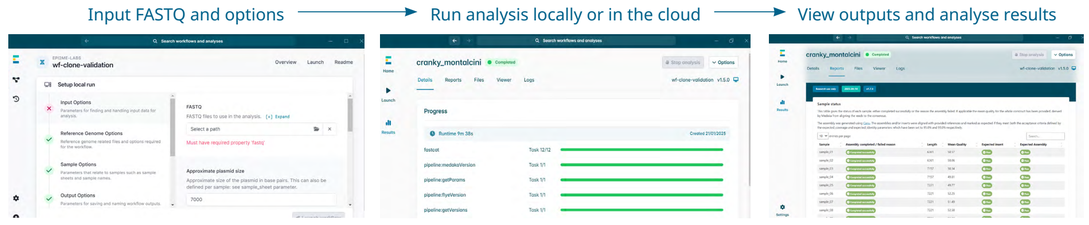


Fig. 1 a) Legacy methods vs clone validation end-to-end workflow b) sample and library preparation diagram c) sequencing and basecalling devices d) EPI2ME™ application for analysis.

End-to-end solution

Legacy methods for validation of cloned vectors usually only consider a small section of the construct. Therefore, they can miss changes in the vector backbone, such as recombination and mutations, or insertion failure (Fig. 1a). The end-to-end clone-validation workflow overcomes these issues by assembling whole constructs, ranging in size from 1,500 to 300,000 bp (Fig. 1b, c, d). To demonstrate its utility, we have used a 96 sample plasmid dataset, which contains two sample prep replicates and two technical replicates, across four MinION Flow Cells. The samples were prepared following the the Rapid sequencing V14 — Plasmid sequencing protocol (SQK-RBK114.96) and sequenced on a MinION™. Basecalling was performed using the Dorado dna_r10.4.1_e8.2_400bps_hac@v5.0.0 HAC model and the FASTQ files were analysed with the EPI2ME wf-clone-validation workflow (v1.8.0) on a standard laptop (16 GB RAM; 8 CPUs).

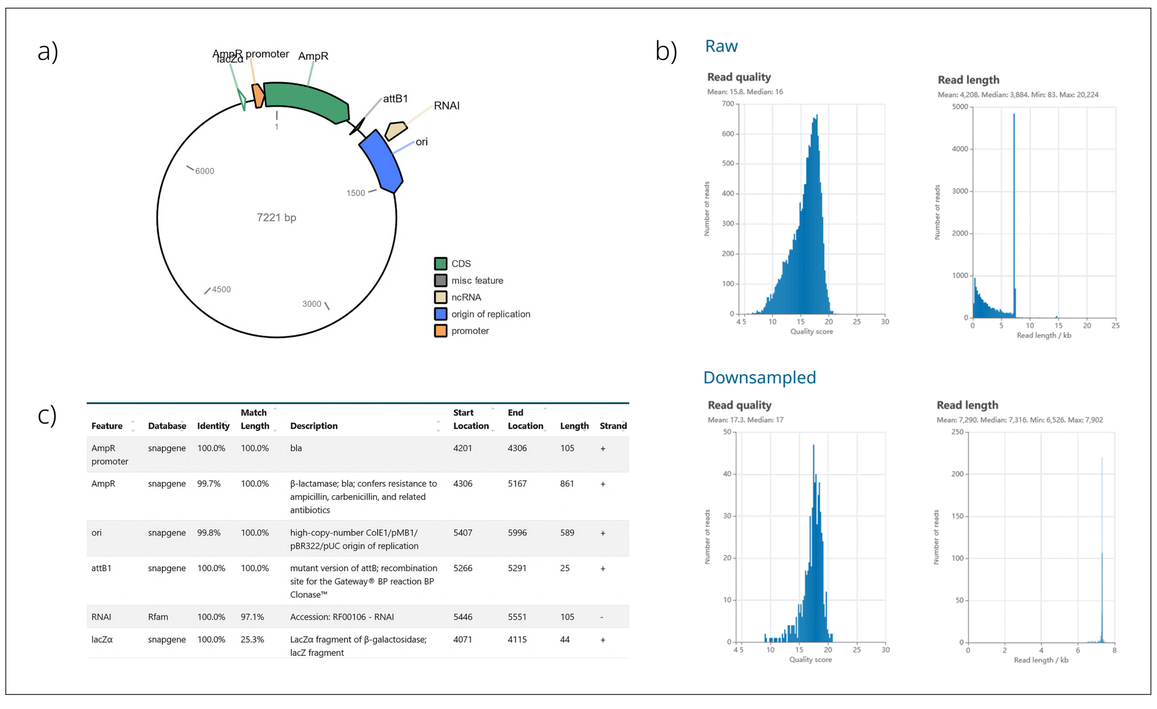


Sample	Assembly completed / failed reason	Length	Mean Quality	Expected insert	Expected Assembly
sample_01	Completed successfully	6361	50.57	Pass	Pass
sample_02	Completed successfully	6361	50.06	Pass	Pass
sample_03	Completed successfully	7157	50.34	Pass	Pass
sample_04	Completed successfully	7157	49.81	Pass	Pass
sample_05	Completed successfully	7221	49.77	Pass	Pass
sample_06	Completed successfully	7221	52.25	Pass	Pass
sample_07	Completed successfully	7221	51.49	Pass	Pass
sample_08	Completed successfully	7221	52.58	Pass	Pass
sample_09	Completed successfully	7221	53.22	Pass	Pass
sample_10	Completed successfully	7221	53.86	Pass	Pass
sample_11	Completed successfully	7222	46.79	Fail	Fail
sample_12	Completed successfully	7221	51.96	Pass	Pass

Fig. 2 a) Running analysis in the EPI2ME application b) sample status table indicating if assembly was successful, assembly length, mean quality, and if the insert and construct recovered were expected compared to the reference.

Comprehensive assemblies

The sequencing output files were directly input into the wf-clone-validation workflow in the EPI2ME application. A sample sheet with per-sample approximate construct sizes, references, and insert references was provided for additional quality control after the *de novo* assembly. The workflow completed in less than an hour and progress was monitored in the workflow overview tab. The workflow produced a per-sample construct FASTA, insert FASTA, plasmid annotations in BED and GBK formats, and a report. The sample status table in the report (Fig. 2b) showed that all assemblies completed successfully, and we can see the read lengths and mean qualities. However, for sample_11, although a plasmid was assembled, when compared to the provided reference it did not meet the acceptance criteria (default 95% coverage, 99% identity) so it showed as failed.

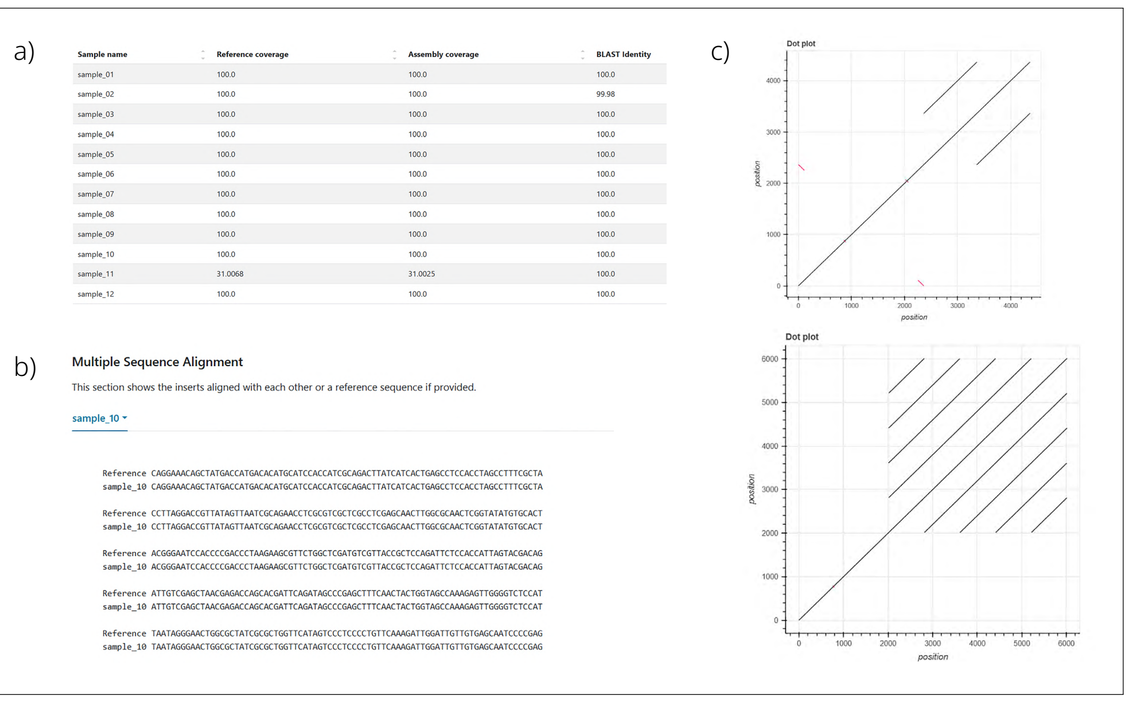


Feature	Database	Identity	Match Length	Description	Start Location	End Location	Length	Strand
AmpR promoter	snapgene	100.0%	100.0%	bla	4201	4306	105	+
AmpR	snapgene	99.7%	100.0%	β-lactamase; bla; confers resistance to ampicillin, carbenicillin, and related antibiotics	4306	5167	861	+
ori	snapgene	99.8%	100.0%	high-copy-number ColE1/pMB1/ pBR322/pUC origin of replication	5407	5996	589	+
attB1	snapgene	100.0%	100.0%	mutant version of attB; recombination site for the Gateway® BP reaction BP Clonase™	5266	5291	25	+
RNAI	Rfam	100.0%	97.1%	Accession: RF00106 - RNAI	5446	5551	105	-
lacZα	snapgene	100.0%	25.3%	LacZα fragment of β-galactosidase; lacZ fragment	4071	4115	44	+

Fig. 3 a) pLannotate annotation plot b) pLannotate feature table c) raw and downsampled read statistics.

Detailed annotations

As can be seen for the example of sample_11, the pLannotate annotation plot in the report provides a visual representation of each construct, confirming whether the expected genes or features are present and in the correct orientation relative to the plasmid (Fig. 3a). If a reference was provided, the orientation of the plasmid will match that of the reference. A full feature table (Fig. 3b) offers additional detail, including percent identity, match length, and descriptive annotations. The read summary section (Fig. 3c) provides insight into read quality and length distributions; for instance, in sample_11, a sharp peak at approximately 7 kb indicates that most reads span the full plasmid. The downsampled data used for assembly can be adjusted by modifying the coverage and minimum quality score settings. Sample_11 has good read lengths and qualities indicating no issues with the sequencing experiment.



Sample name	Reference coverage	Assembly coverage	BLAST Identity
sample_01	100.0	100.0	100.0
sample_02	100.0	100.0	99.98
sample_03	100.0	100.0	100.0
sample_04	100.0	100.0	100.0
sample_05	100.0	100.0	100.0
sample_06	100.0	100.0	100.0
sample_07	100.0	100.0	100.0
sample_08	100.0	100.0	100.0
sample_09	100.0	100.0	100.0
sample_10	100.0	100.0	100.0
sample_11	31.0068	31.0025	100.0
sample_12	100.0	100.0	100.0

Fig. 4 a) Assembly QC from b) multiple sequence alignment c) self-alignment dot plot.

Quality control

The QC table in the report (Fig. 4a), which compares the assembled constructs with the provided reference, shows that sample_11 only has 30% of coverage, which is less than the expected ≥95%. A multiple sequence alignment between provided insert references and the sample confirms the sample_11 insert is not as expected (Fig. 4b). The matching coverage is likely because the vector matches but the insert is not as expected. There are various other QC metrics available to help inspect constructs, including a self-alignment dot plot: repetitive regions appear as diagonal lines off the main diagonal, indicating internal duplications or repeats. These can be used to check repetitive regions have assembled as expected (Fig. 4c). To learn more and try out the workflow, visit: **epi2me.nanoporetech.com/epi2me-docs/workflows/wf-clone-validation/**.