Advances in Plasmid Verification: The Role of NGS in Identity & Clonality Testing.

Keywords: sequence verification, Sanger sequencing, NGS, co-transformation, ExpressPlex™

Introduction

The rapid expansion of synthetic biology has required the generation of larger and more complex plasmid libraries. Advanced cloning techniques and pooled assembly strategies have made this feasible. However, with increased plasmid library complexity comes a bigger challenge: quality control (QC), including sequence and clonality verification, at an unprecedented scale. This application note compares 'gold standard' Sanger sequencing with next-generation sequencing (NGS) for plasmid QC focusing on their strengths, weaknesses, and performance metrics particularly for detecting contaminant constructs that can arise from co-transformation events. It highlights that adopting NGS-based QC incorporating ExpressPlex 2.0 library preparation addresses the modern challenges of cloning fidelity using simple, high throughput workflows that overcome the limitations of Sanger sequencing.

Gene Assembly and the Sequence Verification Bottleneck

Sequence verification of plasmids is essential to confirm that the DNA construct matches the intended design without mutations, insertions, or deletions. The consequences of even a single nucleotide error can be significant. Any unintended plasmid alterations can lead to failed experiments, inaccurate data, or compromise the safety and efficacy of therapeutic development. Thus, sequence verification of every construct within the larger libraries of today is an essential QC component and has shifted a major bottleneck in synthetic construct workflows from plasmid assembly to sequence verification.

Traditionally, sequence verification has relied on Sanger sequencing of construct inserts. While this approach works well for verifying a small number of clones, it becomes impractical for large libraries. Primer walking for each clone, poor read quality through certain secondary structures, and the need to manually interpret electropherograms make it slow and labor-intensive.

NGS offers a faster, more scalable alternative. It requires no primer design, can reconstruct entire plasmid sequences from modest coverage (<100×), and allows many samples to be multiplexed in a single run. With lower sequencing costs and faster turnaround, NGS is now a more cost-effective QC tool for large numbers of plasmids.

A central component of all NGS workflows is library preparation. While many options are available, most are not created to multiplex thousands of samples using a streamlined, easily scaled workflow which can lead to a sequencing bottleneck.

The simplicity of the ExpressPlex 2.0 workflow, its multiplexing of over 6000 samples, and its 96- or 384-well ready-to-use reagents make it uniquely well-suited for automated, ultra-high throughput library prep for sequence verification of plasmid and synthetic construct libraries.

Going Beyond Sequence Verification:

Older cloning methods relied on restriction enzymes, PCR, and ligation, which were reliable because of the low error rate of DNA from natural sources, but were limited by available cut sites and time-consuming subcloning steps. Seamless cloning technologies like Gibson Assembly[1] and NEBuilder® HiFi DNA Assembly (New England Biolabs, Ipswitch, MA) removed these constraints and opened the door to constructing completely novel, *in silico*-designed proteins and large variant libraries from chemically-synthesized DNA.

In addition to modern cloning methods, pooling strategies are also powerful means of generating highly diverse combinatorial libraries in one assembly reaction. However, pooled assembly significantly increases the risk of co-transformation[2], where more than one plasmid is taken up by the same cell. Furthermore, the carryover of errors from oligonucleotide mixtures or longer synthetic DNA fragments can co-transform unwanted contaminants into competent cells. Low levels of co-transforming plasmids can be maintained within single cells when cultures are scaled up and can compromise results in downstream protein expression or phenotypic screening (Figure 1). Erroneous results can lead to significant time and money losses particularly when being used within biotherapeutic development pipelines. Thus, using a high-throughput sequencing method with the sensitivity needed to detect low level co-transformants that express protein impurities is a crucial part of construct QC.

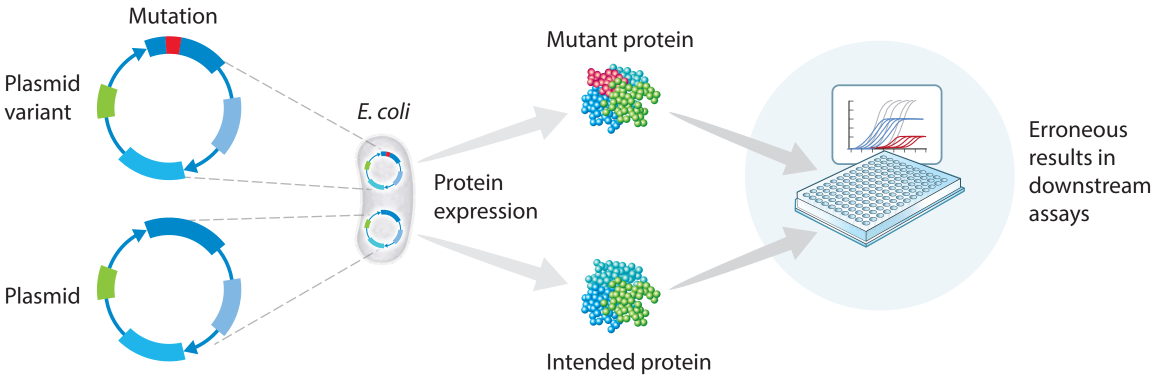


Figure 1. A single bacterium can accept multiple plasmids resulting in impurities that can impact the results of downstream applications.

Sequencing-Based QC: Sanger vs. NGS

Sanger sequencing has traditionally been seen as the 'gold standard' for plasmid verification. It is excellent for confirming the plasmid insert sequence in a single clone, and for identifying single-base errors in that dominant clone. However, since it reads one plasmid sequence per reaction, Sanger's throughput and scale are very limited. Moreover, if multiple plasmids or contaminants are present in a colony (co-transformation), Sanger cannot easily resolve them as mixed traces or minor peaks may be overlooked due to the difficulty of differentiating these from background noise.

NGS fundamentally changes this QC paradigm. This method of verification provides whole-plasmid coverage without primer design and offers orders-of-magnitude higher throughput. By sequencing millions of molecules in parallel, NGS resolves the entire plasmid population in each sample. When combined with an optimized, scalable library prep, NGS can accurately enumerate variant sequences at very low abundances.

Sanger sequencing

- Well-suited for confirming the dominant plasmid sequence in a single clone
- Accurately detects single-base changes in dominant sequence
- Requires design and management of a large number of primer sets
- Limited throughput—one sequence per reaction

Next Generation Sequencing

- Provides full-plasmid coverage without primers
- Processes thousands of samples in parallel
- Can quantify variant abundance directly from read counts

To experimentally examine the sensitivity of the two sequencing-based QC methods in assessing clonality concerns due to bacteria co-transformation, we evaluated each sequencing method's ability to detect several spike-in levels of plasmids containing single-base insertions, deletions, and substitutions.

Experimental Methods

Plasmids were mixed with varying concentration of 'contaminant' plasmids followed by sequencing using either Sanger sequencing or NGS. This assay verified the presence of the known plasmid sequence while examining the sequencing method's capability to detect low levels of contaminating variant sequences.

We designed a series of three plasmids with single-base mutations, addressing insertion, deletion, and substitutions to represent contaminant variants (Figure 2). The plasmids were synthesized by Integrated DNA Technologies (Newark, NJ) and sent to GenScript (Piscataway, NJ) for bulk plasmid preparation. The three (3) high-quality, high-quantity plasmid stocks (>500 µg) contained ≥ 80% of supercoiled DNA and endotoxin levels of ≤0.01 EU/µg.

Plasmid quantification was performed using the Infinite® 200 NanoQuant (Tecan, Männedorf Switzerland) and Quant-IT™ PicoGreen dsDNA Assay Kit (ThermoFisher Scientific, Waltham, MA). Plasmid concentrations were normalized and adjusted to 5.0 ng/µl based on PicoGreen results.

To mimic co-transformation plasmids carrying single-base mutations were spiked into control plasmid samples at 0%, 0.1%, 0.2%, 0.4%, 0.5%, 1%, 10% and 20% (each in triplicate). Aliquots from these mixed samples were sent to GENEWIZ from Azenta Life Sciences (Plainfield, NJ) for Sanger sequencing.

Control (135bp)	5'- ......CAATGTCG*TAA*CGTGATCAAGCGCTTGGATCGTACGAATCAGTCAATCCTGTCACTAGAACCTACC...... -3'
A->C substitution	5'- ......CAATGTCGTAACGTGATCAAGCGCTTGGCTCGTACGAATCAGTCAATCCTGTCACTAGAACCTACC...... -3'
Single Base Deletion	5'- ......CAATGTCG[/]AACGTGATCAAGCGCTTGGATCGTACGAATCAGTCAATCCTGTCACTAGAACCTACC...... -3'
Single Base Insertion	5'- ......CAATGTCGTA[T]ACGTGATCAAGCGCTTGGATCGTACGAATCAGTCAATCCTGTCACTAGAACCTACC...... -3'

Figure 2. Plasmids designed with single-base mutations, including control plasmid of 135 bp, single-base substitution from Adenine (A) to Cytosine (C), single-base deletion, and single-base insertion.

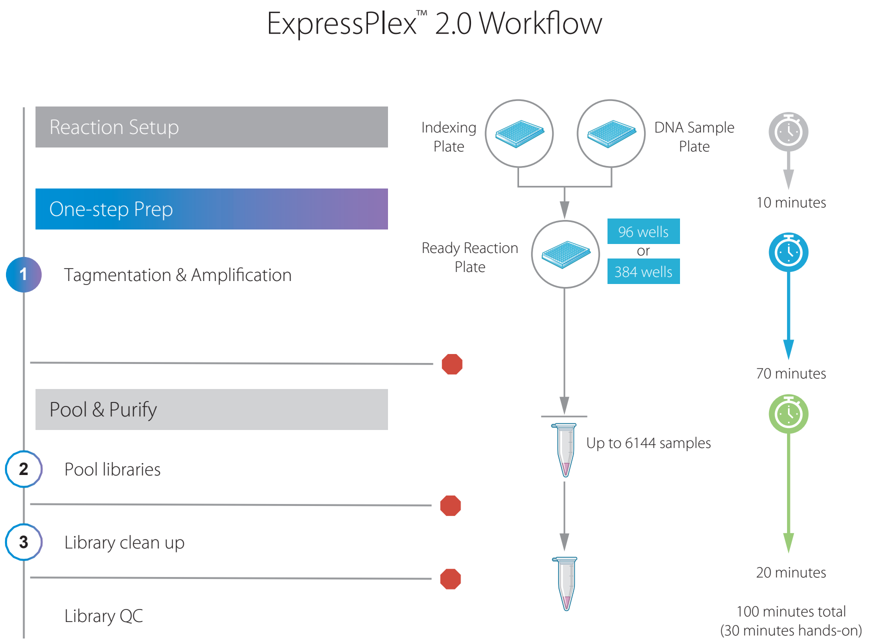


Figure 3. ExpressPlex 2.0 library preparation kit workflow.

Additionally, NGS libraries using these same samples were prepared following manufacturer's instructions using ExpressPlex 2.0 (seqWell, Beverly, MA), which tags DNA with indexed adapters and amplifies libraries in a single reaction (Figure 3). This one-step library prep workflow supports multiplexing of up to 6,144 samples on a single Illumina run, for streamlined sequence verification of very large plasmid libraries.

NGS libraries were sequenced (2×150 bp) on an Illumina NextSeq 2000. The sequencing run was demultiplexed using bcl2fastq, with one mismatch allowed in each index. Reads were aligned to the plasmid control reference using BWA. Variants at specific positions were called with SAMtools mpileup, and frequencies were calculated from coverage depth.

Results

Cloning techniques have vastly improved in both precision and efficiency. However, while improvements in transformation efficiency and pooled gene assembly strategies expand the potential scope of plasmid libraries, they introduce new risks of co-transformation. Understanding the implications of co-transformation and synthesis errors is crucial for reliable use in downstream applications. Robust and highly sensitive QC methods can mitigate the risks posed by contamination in molecular cloning workflows. Sanger sequencing, while reliable for identifying single-base errors or insertions in individual clones, is typically limited to reading one plasmid species at a time and provides no information about subdominant sequences.

To test the sensitivity of Sanger sequencing against NGS for detecting contaminants, control plasmids were mixed with known amounts of variant plasmids containing a single-base insertion, deletion, or substitution at defined ratios (0%–20%). Each mixture was subsequently sequenced using both Sanger and NGS technologies.

The results showed that NGS confidently detected the expected variant, even at spike-in levels below 1%. NGS using ExpressPlex 2.0 met the industry-standard sensitivity threshold of 0.5% co-transformation (i.e. ≥99.5% purity) and consistently detected single-base contaminants well under

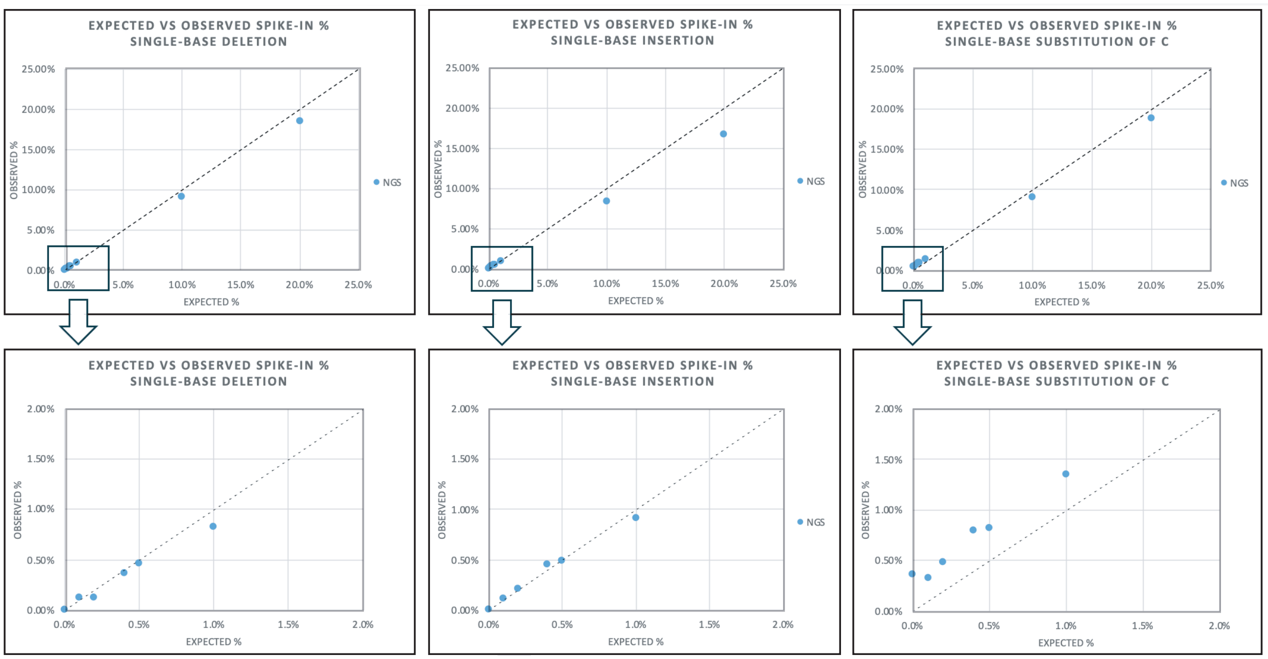
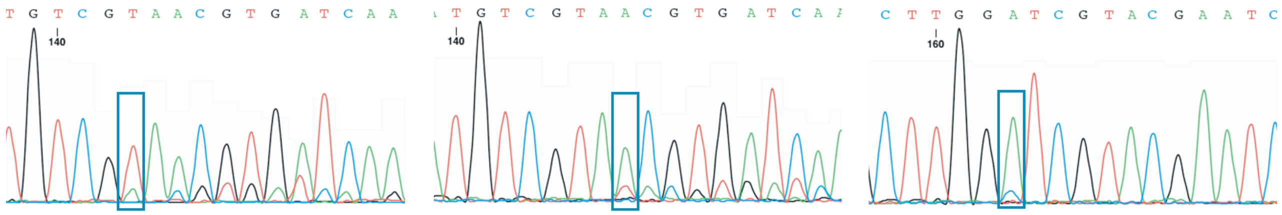


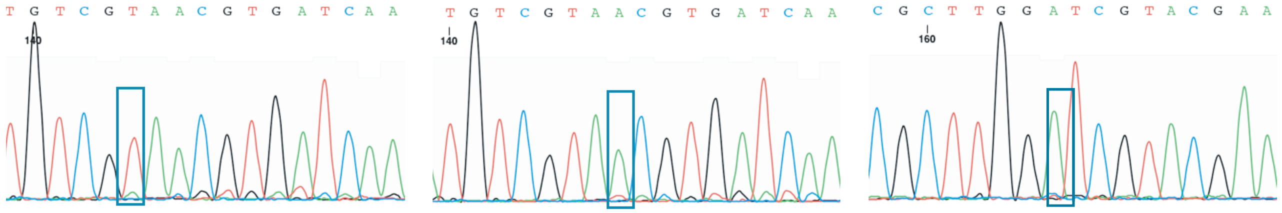
Figure 4. NGS quantification of low-frequency plasmid contaminants with single-base differences. Top row: Observed vs. expected spike-in percentages for plasmids containing a single-base deletion, insertion, or substitution (A→C) across a 0.1%–20% dynamic range. The dashed line represents perfect concordance. Bottom row: Zoomed-in views of the boxed 0%–2% regions from the top graphs, highlighting NGS sensitivity at detecting variants at ≤0.5% abundance. All three mutation types show high concordance between expected and observed frequencies, confirming the ability of NGS workflows to accurately quantify low-frequency sequence variants in plasmid populations.



Single- base Deletion 10% Spiked-in

Single- base Insertion 10% Spiked-in

Single- base Substitution 10% Spiked-in



Single- base Deletion 1% Spiked-in

Single- base Insertion 1% Spiked-in

Single- base Substitution 1% Spiked-in

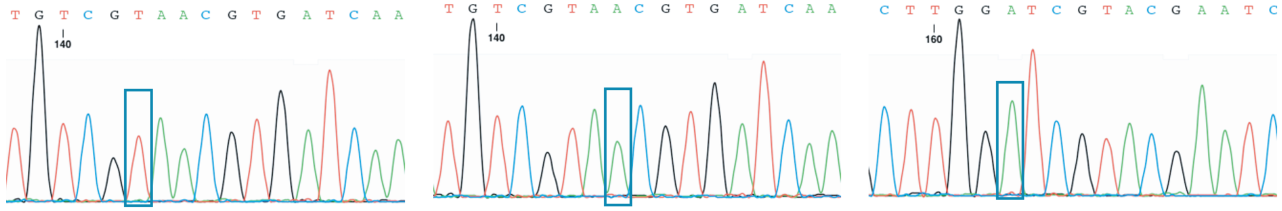


Figure 5. Sanger sequencing electropherograms of spike-in mixtures at decreasing variant frequencies. Each panel represents a chromatogram from mixtures of two plasmids at defined ratios, with the minor variant frequency decreasing from top to bottom and single base mutation changes from left to right. The boxed regions highlight the expected variant position. Mixed base calls (double peaks) are only clearly distinguishable when the minor variant is present at approximately ≥20% of the population. Below this threshold, the signal becomes difficult to resolve from background noise, demonstrating the limited sensitivity of Sanger sequencing in detecting low-frequency variants.

that level (Figure 4, bottom row demonstrates detection <0.5%). This suggests that a minority plasmid present at 0.5% of the population will be flagged with confidence.

In contrast, contaminating plasmids were very difficult to differentiate from background noise using Sanger sequencing, unless they were present at ≥20% of the mixed population (Figure 5). See table 1 for a full summary comparing the features of Sanger and NGS-based QC methods.

Discussion and Conclusion

Plasmids are no longer simple cloning tools — they are now critical starting materials for cell and gene therapies (e.g., CAR-T cells, AAV production), mRNA vaccines, and biotherapeutics (e.g. monoclonal antibodies, recombinant proteins), as well as playing a critical role in the burgeoning field of synthetic biology (e.g., biosensors, engineered microbes).

In these applications, even small errors (mutations, contaminations, or incorrect sequences) can have major downstream effects — from failed expression to artifacts and erroneous results. Therefore, rigorous plasmid QC is required to confirm their integrity and identity before use in high-value assays.

Innovations in cloning and synthetic gene assembly have driven large increases in plasmid library size and diversity, but also have introduced new risks of co-transformation and hidden mutations. Thus, access to robust, scalable sequence and clonality verification is required for modern laboratories.

Our results showed that NGS using ExpressPlex 2.0 library prep accurately measured the expected variant frequencies, even at spike-in levels below 1% for all mutation types. Sanger sequencing, however, could not detect contaminants unless present at >20%. Thus, while Sanger sequencing still has a role for final confirmation of a single clone, it cannot guarantee clonality nor meet the scalability requirement for high-throughput, full plasmid sequence verification.

Feature	NGS (ExpressPlex 2.0 + Illumina)	Sanger Sequencing
Co-transformation detection	Yes – detects low levels of contaminating plasmids in the sample	No – dominant plasmid only
Quantitative output	Yes – variant %	No – electropherogram only
Contamination detection limit	≥0.5%	≥20%
Multiplexing	Thousands of samples/run	1 reaction/sequencing capillary
Reactions/sample	1 reaction	Multiple reactions - highly dependent on plasmid size
Workflow time	~100 min (30 min hands-on)	3 - 5 hours
Best use case	High-throughput QC	Small-scale validation

Table 1. Features of NGS using ExpressPlex library prep from seqWell versus Sanger sequencing for plasmid QC.

NGS Plasmid QC Using ExpressPlex 2.0:

- Detects rare contaminants (≥0.5%)
- Processes thousands of samples in parallel
- Fits into automated pipelines
- Reduces time and cost for high-throughput QC

References

1. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A. 3rd and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. ***Nat. Methods***, 6, 343–345.

2. Tomoiaga D, Bubnell J, Herndon L, Feinstein P. High rates of plasmid cotransformation in E. coli overturn the clonality myth and reveal colony development. ***Sci Rep.*** 2022 Jul 7;12:11515. doi: 10.1038/s41598-022-14598-9. PMID: 35798773; PMCID: PMC9262894.

Key NGS Takeaways:

- Sensitivity: Detecting low levels of contaminating plasmids (≥0.5% abundance) ensures that even trace co-transformants are identified. NGS detects low level co-transformants that Sanger misses.
- Scalability: NGS workflows handle hundreds to thousands of samples at once.
- Reliability: Ensures only truly clonal plasmids advance to costly downstream