Application Note

Scalable, True Multiplexed Library Preparation Enables High-Performance, High-Throughput Plasmid Sequencing

- Novel, truly multiplexed library preparation technology enables rapid generation of auto-normalized, sequencing-ready library pools in any laboratory setting.
- Even read distributions, longer and tunable insert sizes, and uniform coverage support high success rates in high-throughput *de novo* plasmid assembly.

plexWell Library Preparation Kits require only 1 hour of hands-on time per batch of 96 samples. Robust performance and flexible reagent formats facilitate implementation in high-throughput plasmid sequencing workflows.

Introduction

Plasmids are ubiquitously used in recombinant DNA technology for the cloning, engineering, transfer, and expression of DNA fragments or genes. Broad-based access to cost-effective, massively parallel, next-generation sequencing (NGS), and advances in synthetic biology have revolutionized the scale on which plasmids can be constructed, screened, and manipulated for wide-ranging applications, including biomarker discovery, antibody engineering, genome editing, and gene therapy.

NGS library preparation throughput, performance, and cost remain bottlenecks in high-throughput plasmid sequencing. Commercial library construction kits and homebrew methods typically require laborious quality control of input DNA and/or final libraries, are enzyme intensive, include multiple bead-based purification steps, and/or fail to provide a sufficient number of sequencing barcodes for optimal utilization of sequencing capacity on higher throughput sequencers. Furthermore, biases introduced during library preparation impact sequencing quality and coverage, leading to incomplete or low-quality plasmid assemblies.

seqWell's plexWell™ library preparation technology employs a transposase to selectively tag DNA samples with Illumina® adapters in a unique, sequential manner. The cost-effective, streamlined, and highly scalable workflow yields pools of normalized libraries with unique barcode combinations.[1] Uniform read distributions and high data quality further reduces overall sequencing costs.

In this study, we demonstrate the utility of plexWell Library Preparation Kits for high-quality, high-throughput plasmid sequencing on benchtop sequencers. Diverse sets of plasmids were processed to illustrate robust performance in real-life settings. plexWell Library Preparation Kits were also shown to outperform competitor transposase-based workflows in terms of ease-of-use, flexibility, and overall data quality.

For Research Use Only. Not for use in diagnostic procedures.

Materials and Methods

Plasmids – Data presented in Figures 2, 4, 5, and 6, and in Table 1 were generated from a set of 96 characterized, highly diverse plasmids (2.5 kb to 15.3 kb in length) obtained from a commercial supplier. This set included six viral plasmids in adeno-associated virus (AAV) vectors. Data presented in Table 2 and Figure 7 were generated from 960 (10 plates of 96) customer samples, submitted for seqWell's High-Throughput plexWell™ Plasmid Sequencing and *De-novo* Assembly Service.[2]

DNA quantification – Where applicable, input DNA and/or sequencing-ready libraries or library pools were quantified using an Infinite® F200 PRO microplate reader (Tecan) and Quant-IT™ PicoGreen dsDNA Assay Kit (ThermoFisher Scientific).

Library preparation with plexWell™ Library Preparation Kits – Plasmid DNA concentrations ranged from 2.8 – 24.4 ng/µL (median: 6.6 ng/µL). The global dilution factor (GDF) was calculated and was applied across all 96 samples to obtain an input DNA concentration range of 1.1 – 9.7 ng/µL (median: 2.6 ng/µL).

Of each diluted plasmid sample, 4 µL was used in the Sample Barcoding reaction (first library preparation step, see Figure 1). This translated to an DNA input range of 4.5 – 39.0 ng (median: 10.5 ng). Depending on the experiment, libraries were prepared with the plexWell 384 or plexWell Plus 24 Library Preparation Kit, according to standard protocols,[3,4] with twelve cycles of library amplification.

Additional library preparation – Eight of the 96 plasmid samples (B04, C05, F02, F07, F10, F11, G03, and H05) were also processed with a transposase-based library preparation kit from a different supplier. The manufacturer's recommended protocol requires accurate quantification and normalization of input DNA (1 ng), and provides for optional, bead-based library normalization. Two libraries were prepared from each of the eight samples: NXT libraries were quantified with the Quant-IT assay, and pooled for sequencing on an equal-mass basis; whereas NXT-Bead libraries were normalized and pooled using the bead-based procedure.

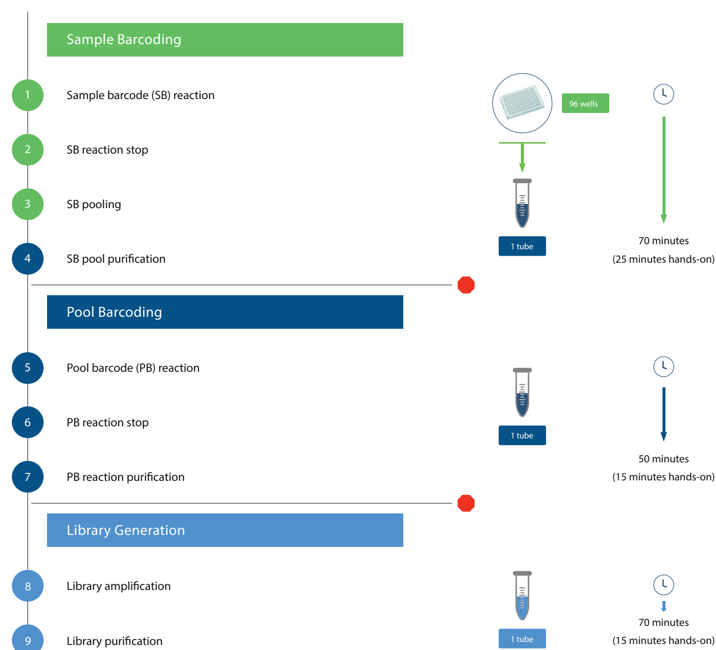


Figure 1. Overview of the plexWell library preparation workflow. The Sample Barcoding and Pool Barcoding steps are designed to fragment DNA, add adapters (including sequencing barcodes), and normalize libraries via sequential tagging with a transposase. A limited number of PCR cycles are performed in the Library Generation step to fill gaps, and produce a sufficient amount of the library pool for sequencing. Red hexagons designate safe stopping points. Integrated normalization eliminates laborious dilution of input DNA (a single global dilution factor is applied to all samples). Pooling of libraries in the first (Sample Barcoding) step significantly streamlines library construction by allowing for single-tube bead purifications. This supports higher throughput, facilitates automation, and simplifies processing of library pools for sequencing.

Sequencing – Paired-end sequencing (2 x 250 bp) was performed using an Illumina® MiSeq™ System. Library pools were diluted to a final concentration of 12 pM when using MiSeq Reagent Kits v2, or 20 pM for v3 reagent kits.

Data Analysis – Sequencing data were processed, and plasmids were assembled using the seqWell plasmid assembly pipeline. In short, sequencing data were demultiplexed, and fastq files generated with bcl2fastq.[5] This was followed by fastq preparation (quality trimming, read merging, and read normalization) using several packages from the bbtools suite.[6] The resulting merged/normalized fastq files (two per sample) were individually assembled using the Geneious circular assembler[7] with manual quality curation. The best assembly was selected for each plasmid.

For Research Use Only. Not for use in diagnostic procedures.

Results and Discussion

plexWell™ technology enables uniform read count distributions without normalization of input DNA

In the first party of this study, a batch of 96 diverse plasmids were processed with the plexWell 384 Library Preparation Kit. Despite an 8.7-fold range in original plasmid concentrations (2.8 – 24.4 ng/µL), all samples were diluted by the same global dilution factor (GDF = 2.5) to achieve the recommended median DNA input amount of approximately 10 ng.

Read fractions for the individual libraries in the pool of 96 (shown in Figure 2A), ranged 3.5-fold, from 0.44% to 1.53% (median: 1.05%); with a coefficient of variation (CV) of 17.75% across the data set. Neither the original plasmid concentration (Figure 2B), nor plasmid length (Figure 2C) impacted read distributions, suggesting robust library construction and normalization.

The last six libraries (H07 – H12) appeared to perform slightly worse than the rest of the batch, with a median fraction of reads of 0.62% for this subset. These libraries were prepared from the viral plasmid samples. Since AAV vectors/plasmids comprise a mixture of single- and double-stranded DNA species,[8] higher inputs are needed to achieve balanced read counts in sequencing batches comprised mainly of dsDNA plasmids.[9] Nevertheless, no special adjustments were made for this subset of plasmids in this particular study.

The plexWell workflow supports high-quality plasmid assembly

Overall coverage was highly uniform (98.6 ±4.7 X) across the set of 96 plasmids. Assemblies were obtained for 100% of the plasmids. Of these, 90 (94%) were regarded as high-quality (circular assemblies with no ambiguities). Three of the six assemblies for the viral samples were linear, whereas the remaining three contained ambiguities.

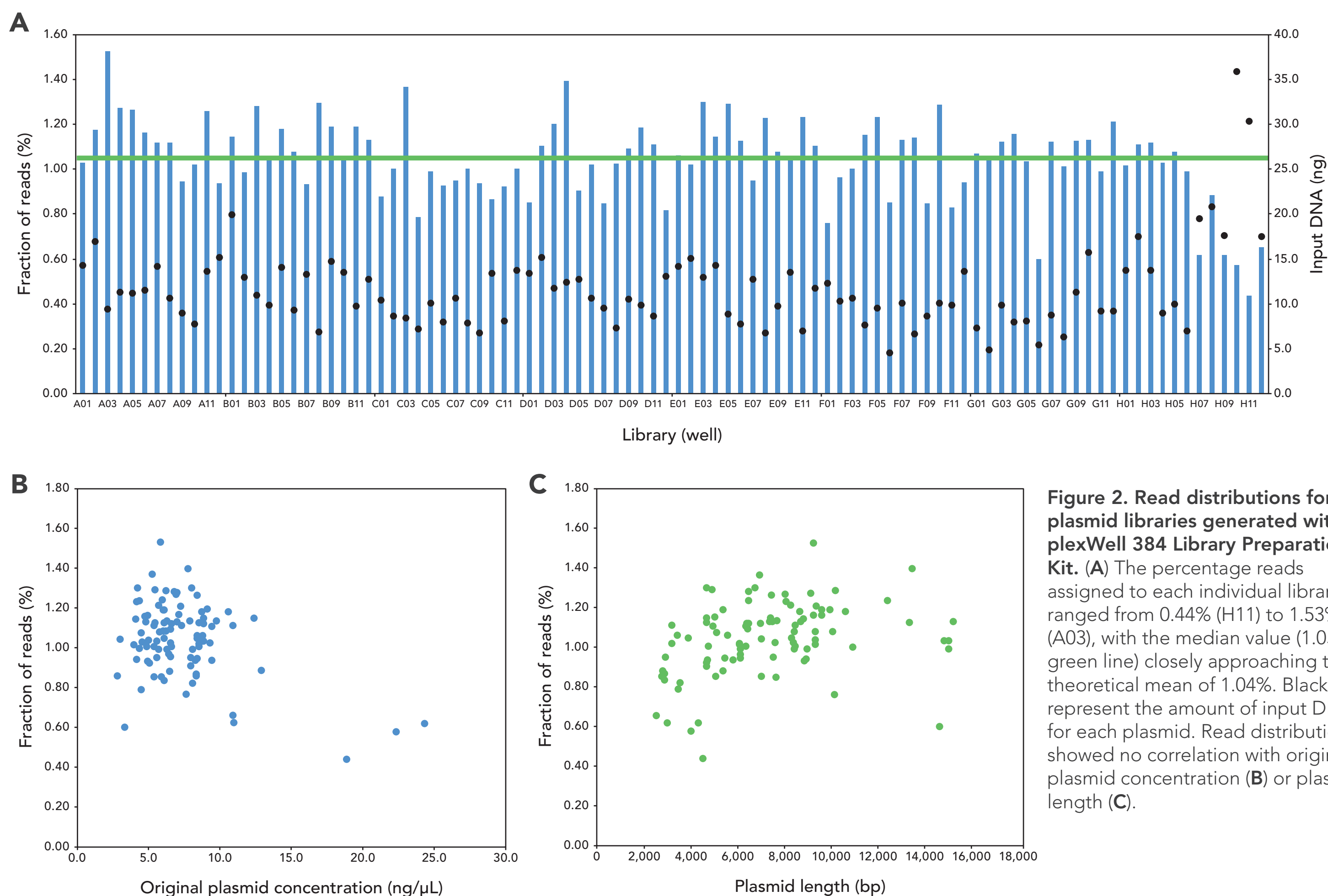


Figure 2. Read distributions for 96 plasmid libraries generated with the plexWell 384 Library Preparation Kit. (**A**) The percentage reads assigned to each individual library ranged from 0.44% (H11) to 1.53% (A03), with the median value (1.05%, green line) closely approaching the theoretical mean of 1.04%. Black dots represent the amount of input DNA for each plasmid. Read distributions showed no correlation with original plasmid concentration (**B**) or plasmid length (**C**).

plexWell™ outperforms competitor workflows with respect to flexibility and overall data quality

A side-by-side comparison of the three workflows (**plexWell** with early-stage pooling and integrated normalization, **NXT** with library quantification and pooling, and **NXT-Bead** with bead-based normalization and final library pooling) is given in Figure 3.

Read distributions for the subset of eight libraries prepared with all three workflows are given in Figure 4. Mean read fractions were identical for all three workflows, and matched the theoretical average.

The plexWell workflow performed slightly better with respect to median read count, whereas the additional effort dedicated to library quantification and mass-based pooling resulted in less variation with the NXT workflow.

Insert sizes for libraries generated with the three different workflows are given in Figure 5. The plexWell workflow yielded the most consistent library insert size. More importantly, insert sizes were significantly longer than those obtained with both the NXT and NXT-Bead workflows. This maximizes the amount of unique data obtained from paired 250-bp reads, and generally facilitates plasmid assembly.

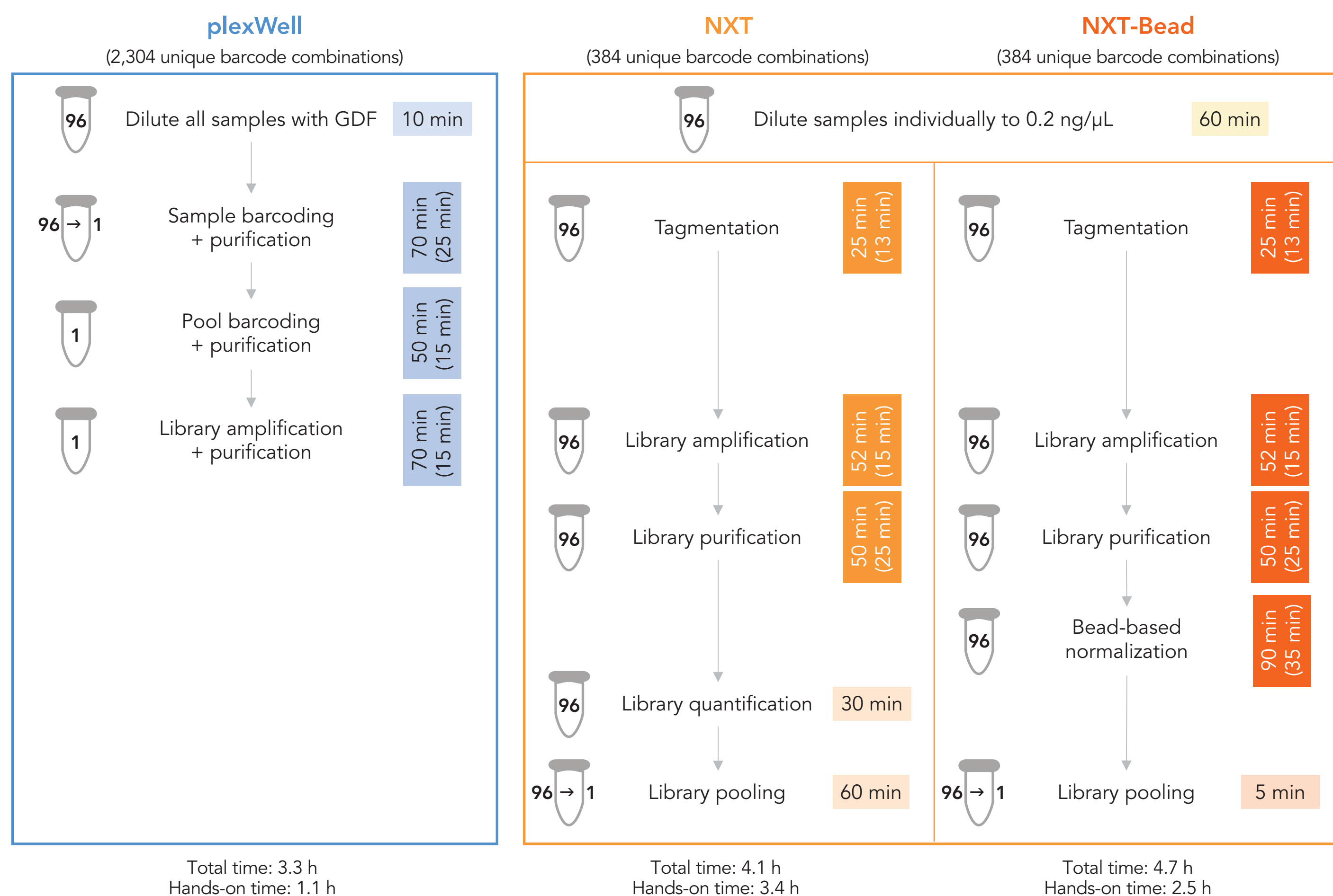
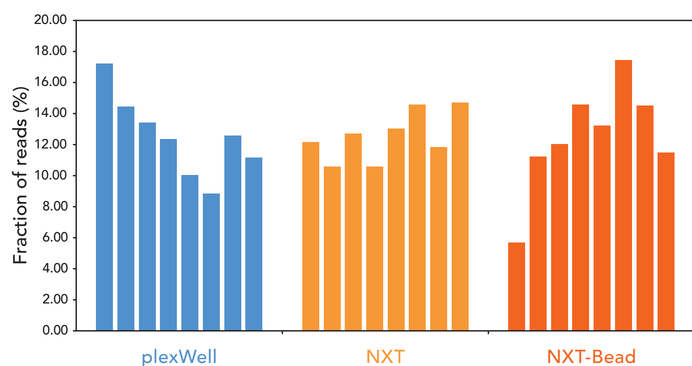


Figure 3. Comparison of three transposon-based library preparation workflows used in this study. The unique plexWell technology requires fewer and less laborious sample manipulations, and offers six times as many unique sequencing barcodes. This supports higher throughput irrespective of whether the workflow is executed manually or using automated liquid handlers. In addition, the plexWell kit's integrated normalization feature does not require careful dilution of input DNA, and obviates the need to quantify and pool individual libraries prior to sequencing. All times are for the processing of a batch of 96 samples, and have been verified by an experienced user. Since the hands-on time required for the plexWell workflow is only one-third of the total time, up to four batches of 96 libraries can easily be handled (in semi-parallel fashion) in a single workday by a skilled technician with access to multiple thermocyclers.

For Research Use Only. Not for use in diagnostic procedures.



	plexWell	NXT	NXT-Bead
Mean	12.50%	12.50%	12.50%
Median	12.45%	12.42%	12.63%
CV	21.00%	12.60%	27.30%
Ratio	1.94	1.39	3.05

Figure 4. Read distributions for the eight plasmid libraries generated with all three transposase-based workflows. Bars in each set correspond to libraries B04, C05, F02, F07, F10, F11, G03, and H05 (left to right). Plasmid DNA concentrations ranged between 2.5 and 2.6 ng/µL. DNA was used as is for the plexWell workflow, translating to a DNA input of 10.0 ±0.17 ng in each of the Sample Barcoding reactions. Plasmids were diluted further for the NXT and NXT-Bead workflows to meet the recommended input requirement of 1 ng. Key read metrics are summarized below the graph. *CV:* coefficient of variation. *Ratio:* ratio between the highest and lowest library read fraction for each workflow.

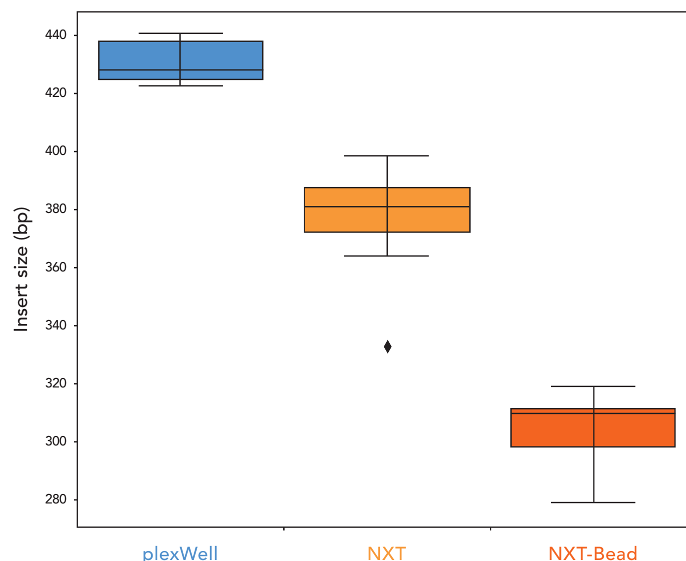


Figure 5. Fragment size distributions for libraries prepared with different transposase-based workflows. Insert sizes were calculated from demultiplexed, quality-trimmed, merged reads. The plexWell workflow produced the longest and most consistent insert sizes and offers tunable final library purification parameters to optimize library fragment size for sequencing read length. Bead-based purification in the NXT-Bead workflow appears to select for shorter library fragments, which are not optimal for the longer sequencing read lengths preferred for *de novo* plasmid assembly.

To further interrogate the quality of data generated with the three different workflows, data were progressively downsampled prior to plasmid assembly. Results are summarized in Table 1, and confirm that the plexWell technology supports high-quality plasmid assemblies from less sequencing data than other transposase-based kits. This was partially attributed to longer library fragment sizes, but also to more uniform coverage across the entire plasmid sequence (Figure 6). This suggests that the plexWell™ workflow introduces less bias during fragmentation, tagging, normalization, and/or library amplification than the other workflows.

Table 1. Percentage of high-quality assemblies from progressively downsampled data

Workflow	Number of read pairs		
	10,000	5,000	2,000
plexWell	100%	100%	100%
NXT	100%	87.5%	100%
NXT-Bead	100%	100%	75%

plexWell Library Preparation Kits support high-quality, high-throughput plasmid sequencing on benchtop sequencers

As shown in Figure 3, plexWell Library Preparation Kits enable processing of up to 384 (four batches of 96) plasmids in a standard workday by a single technician. Due to robust performance, less than 1.5 Gb of data per plate of 96 samples is typically required to achieve high-quality assemblies. As a result, the plexWell technology enables high-throughput plasmid sequencing without the need for sophisticated liquid handling, with low- and mid-throughput benchtop sequencers.

Results for 960 (10 plates of 96) customer-submitted plasmids, processed using seqWell's high-throughput plasmid sequencing and *de novo* assembly service pipeline, are given in Table 2 and Figure 7. Sequencing data were generated in single 2 x 250 bp Illumina® MiSeq™ run, yielding an average of 1.35 Gb of data per plate. Assemblies were obtained for 955 of the 960 plasmids (99.5%). Assemblies were free of ambiguities for 99.2% of the 960 samples.

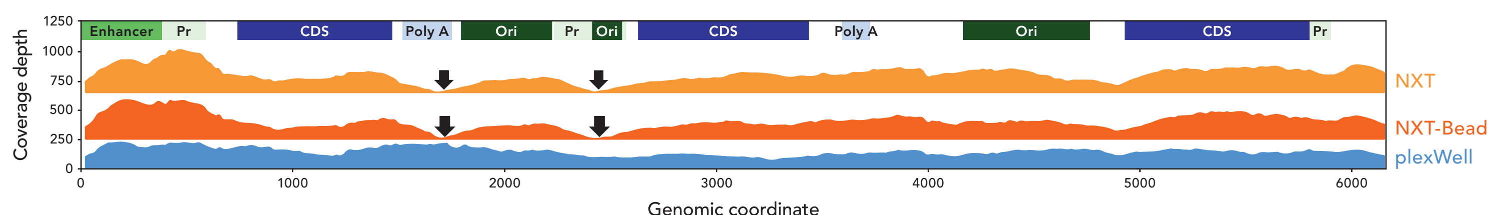


Figure 6. Example of local sequencing coverage for one of the plasmids sequenced with all three workflows. Data for all three workflows were downsampled to 2,000 reads to generate coverage profiles. The region of plasmid F02 shown here contains different functional elements, designated with bars of different shades of green and blue at the top of the image (*Pr:* promoter, *CDS:* coding sequence, *Ori:* origin of replication, *PolyA:* polyadenylation signal). The plexWell data shows more even coverage across the entire region. In the NXT and NXT-Bead data, some areas (e.g., the enhancer and promoter regions between coordinates 0 and 1,000) are relatively over-represented, at the expense of areas of very low coverage (such as those indicated with the black arrows). These areas of low coverage persist, even with five times as many reads (data downsampled to 10,000 reads; not shown).

Table 2. Sequencing metrics and assembly results for 960 plasmids sequenced in a single Illumina® MiSeq™ run

Plate	Sequencing metrics			Assembly results		
	Amount of data (Gb)	CV	Median reads/ plasmid	Plasmids assembled	Assemblies with no ambiguities	Circular assemblies with no ambiguities
1	1.999	18.61%	40,921	96 (100%)	96 (100%)	91 (94.8%)
2	1.373	22.29%	28,095	96 (100%)	96 (100%)	93 (96.9%)
3	1.764	25.15%	36,658	96 (100%)	94 (97.9%)	92 (95.8%)
4	1.634	24.93%	35,041	96 (100%)	95 (99.0%)	94 (97.9%)
5	0.895	12.88%	18,278	95 (99.0%)	95 (99.0%)	86 (89.6%)
6	0.997	21.62%	20,407	96 (100%)	96 (100%)	92 (95.8%)
7	1.008	18.75%	20,207	96 (100%)	96 (100%)	91 (94.8%)
8	1.212	82.91%	17,523	96 (100%)	96 (100%)	84 (87.5%)
9	1.076	47.68%	21,463	96 (100%)	96 (100%)	94 (97.9%)
10	1.556	33.65%	30,464	92 (95.8%)	92 (95.8%)	89 (92.7%)
Average	**1.351**	**30.85%**	**26,906**	**955 (99.5%)**	**952 (99.2%)**	**906 (94.4%)**

Samples were processed in 10 plates of 96 plasmids each. Plates are listed in random order. The run yielded a total of 13.5 Gb data after adapter- and quality trimming. *CV:* coefficient of variation across the 96 plasmids in a plate. Even though Plate 8 returned an unusually high CV, plasmid assembly rates were not impacted. The seqWell plasmid assembly pipeline assumes plasmid assemblies to be circular. However, the desired plasmid topology is not known for all customer samples.

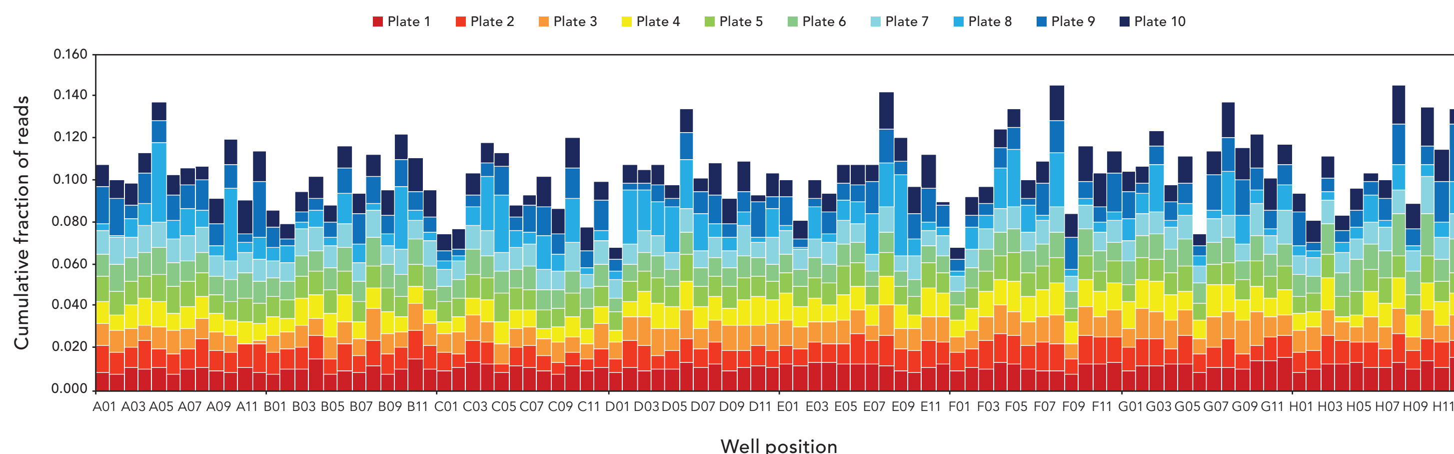


Figure 7. Read distributions by plate and well position for 960 plasmids sequenced in a single Illumina® MiSeq™ run.

For Research Use Only. Not for use in diagnostic procedures.

Conclusions

plexWell™ library preparation technology utilizes a unique, sequential, transposase-based tagging method to generate auto-normalized NGS libraries in true multiplexed fashion. This approach is particularly powerful in high-throughput sequencing of samples with relatively low sequence complexity (e.g., long PCR products, plasmids, and small genomes), where it is challenging to balance library construction capacity and cost with available sequencing capacity.

In this study, we have demonstrated the benefits of plexWell Library Preparation Kits for high-throughput plasmid sequencing, particularly on benchtop sequencers:

- The unique, streamlined workflow with integrated normalization significantly reduces hands-on time and enables the processing of up to 384 plasmids by a single technician in a standard workday.
- Effective auto-normalization results in even read distributions, without careful quantification and/or normalization of input DNA or sequencing-ready libraries.
- Longer and tunable fragment sizes and uniform coverage enable high *de novo* plasmid assembly quality and success rates, from less sequencing data than required for other transposase-based workflows.
- Plate-based reagents, flexible kit configurations, and up to 2,304 sequencing barcodes provide flexibility in terms of batch sizes, facilitate implementation in both manual and automated plasmid sequencing pipelines, and allow for optimal utilization of available sequencing capacity.

Robust performance and ease-of-use makes plexWell Library Preparation Kits ideally suited for high-throughput plasmid verification and screening in both research and commercial[10] settings.

References

1. https://seqwell.com/technology/
2. https://seqwell.com/plexwell-plasmid/
3. User Guide: plexWell™ 384 Library Preparation Kit and plexWell™ 96 Library Preparation Kit for Illumina® Sequencing Platforms. https://seqwell.com/wp-content/uploads/2021/06/plexWell_384_and_96_Library_Preparation_Kit_User_Guide_v20210609.pdf
4. User Guide: plexWell™ Plus 24 Library Preparation Kit for Illumina® Sequencing Platforms. https://seqwell.com/wp-content/uploads/2019/06/20210421-plexWell_Plus_24_Library_Preparation_Kit_User_Guide_v20210402.pdf
5. https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html
6. https://jgi.doe.gov/data-and-tools/bbtools/
7. https://www.geneious.com/features/assembly-mapping/
8. Guerin K, et al. *Human Gene Therapy* 2020, 31:664-678. doi: 10.1089/hum.2019.277
9. Internal data.
10. https://blog.addgene.org/plasmids-101-an-inside-look-at-ngs-plasmid-quality-control

Contact us at info@seqwell.com to learn how to optimize your plasmid sequencing and assembly pipeline

seqWell Inc.
66 Cherry Hill Drive
Beverly, MA 01915

+1 (855) SEQWELL (737-9355)
sales@seqwell.com

Learn more about
plexWell kits at
seqwell.com/products/

Unless otherwise stated in references, presented data on file.
For Research Use Only. Not for use in diagnostic procedures.
All product names and trademarks are the property of their respective owners.
© 2021 seqWell Inc. All rights reserved.

seqWell