APPLICATION NOTE

16S rRNA sequencing on NextSeq™ 1000 & NextSeq 2000 Systems

Efficient, high-throughput characterization of microbial populations

illumina®

Sequencing 16S ribosomal RNA for microbial population studies

The 16S ribosomal RNA (rRNA) gene is involved in the translation of RNA to protein and is conserved across species. The gene itself is approximately 1550 bp with a mix of conserved and variable regions that facilitate sequencing applications for phylogenetic classification.[1] This application note describes a comprehensive workflow for sequencing the V3 and V4 variable regions of the 16S gene using the NextSeq 1000/2000 Reagent 600-cycle kits. Sequencing performance is compared to the well-established performance of the MiSeq™ Reagent v3 Kit (600 cycles) on the MiSeq System.

These NextSeq 1000/2000 Reagent 600-cycle kits expand the capacity and sequencing output of the NextSeq 1000 and NextSeq 2000 Systems. The kits are ideal for 16S rRNA sequencing because they enable labs to attain a higher sample read depth and output in less time than other next-generation sequencing (NGS) platforms and methods. The streamlined 16S metagenomics workflow integrates library preparation, proven Illumina NGS technology, and secondary data analysis through the 16S Metagenomics Labs app available on BaseSpace™ Sequence Hub (Figure 1).

Methods

Libraries for 16S analysis were prepared from microbial genomic DNA samples as described below and sequencing performance was compared between the NextSeq 2000 and MiSeq Systems. Using IDT for Illumina DNA/RNA UD Indexes Sets A to D during library preparation allows users to generate 384 16S libraries. Running 384 16S libraries on the NextSeq 1000/2000 P2 300M Reagents Kit (600 cycles) with standard SBS chemistry or the NextSeq 1000/2000 P2 XLEAP-SBS™ Reagent Kit (600 cycles) generates 100,000 to 200,000 reads per sample, which is sufficient for classification at the genus level. Users can generate 300M total reads on the NextSeq 1000/2000 P2 300M Reagents Kit (600 cycles) and 400M reads on the NextSeq 1000/2000 P2 XLEAP-SBS Reagent Kit (600 cycles). Sequencing run time is 34 hours for the NextSeq 1000/2000 P1 XLEAP-SBS Reagent Kit (600 cycles) and 42 hours for the NextSeq 1000/2000 P2 XLEAP-SBS Reagent Kit (600 cycles). In comparison, sequencing run time on the MiSeq System with MiSeq Reagent Kit v3 (600 cycles) is ~56 hours.

Samples

Microbial genomic DNA samples were obtained from two sources. The American Type Culture Collection (ATCC) 20 Strain Staggered Mix Genomic Material (ATCC, Catalog no. MSA-1003) is a mock microbial community comprised of a staggered distribution of genomic DNA prepared from bacterial strains selected based on attributes such as Gram stain, GC content, and sporulation attributes.

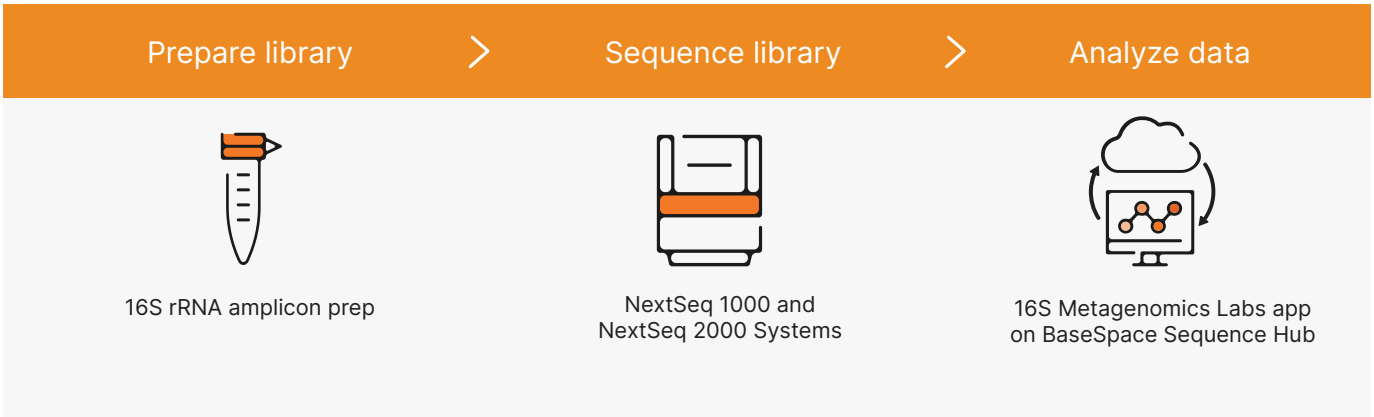

Figure 1: 16S metagenomic sequencing workflow—The 16S sequencing workflow includes library preparation, sequencing, and secondary data analysis. Sequencing run time for the NextSeq 1000/2000 P1 XLEAP-SBS Reagent Kit (600 cycles) is 34 hrs and run time for the NextSeq 1000/2000 P2 XLEAP-SBS Reagent Kit (600 cycles) is 42 hrs.

In addition, Illumina partnered with Zymo Research, Irvine, CA, to run a comparison using ZymoBIOMICS Fecal Reference with TruMatrix Technology (Zymo Research Catalog no. D6323), a real-world representation of fecal samples that is composed of stool from healthy donors. This fecal reference is used for quality control, process validation, assay development, and proficiency testing.

ATCC samples

Library preparation

Libraries for 16S analysis were prepared following the established 16S metagenomics sequencing library workflow. V3 and V4 regions of the 16S rRNA gene were PCR amplified using a bacterial primer pair selected from the scientific literature (Table 1).[2] Next, 10 µl Illumina sequencing adapters and unique dual-index barcodes (IDT for Illumina DNA/RNA UD Indexes Set A, Tagmentation, Catalog no. 20027213) were added in a second PCR, generating amplicons compatible with the NextSeq 2000 System. We recommend using Illumina Purification Bead, 100 ml (Illumina, Catalog no. 20060057) to purify the 16S libraries. The resulting libraries were manually normalized and pooled for sequencing (Figure 2). This method can also be used to target other amplicons of 16S RNA or genes throughout the genome.[3] The adapter sequence must be added to the locus-specific primer for the region to be sequenced.

Table 1: Primer sequences for 16S V3 and V4 amplicon sequencing

Primer name	Sequence[a,b]
16S amplicon PCR forward primer	5'–TCGTCGGCAGCGTCA GATGTGTATAAGAGACAG-CCTACGGGNGGCWGCAG–3'
16S amplicon PCR reverse primer	5'–GTCTCGTGGGCTCGGA GATGTGTATAAGAGACAG-GACTACHVGGGTATCTAATCC–3'

a. International Union of Pure and Applied Chemistry (IUPAC) nucleotide nomenclature: N = any base; W = A or T; H = A or C or T; V = A or C or G.
b. Primer sequence before the hyphen is the Illumina overhang adapter sequence. Primer sequence after the hyphen corresponds to locus-specific sequence.

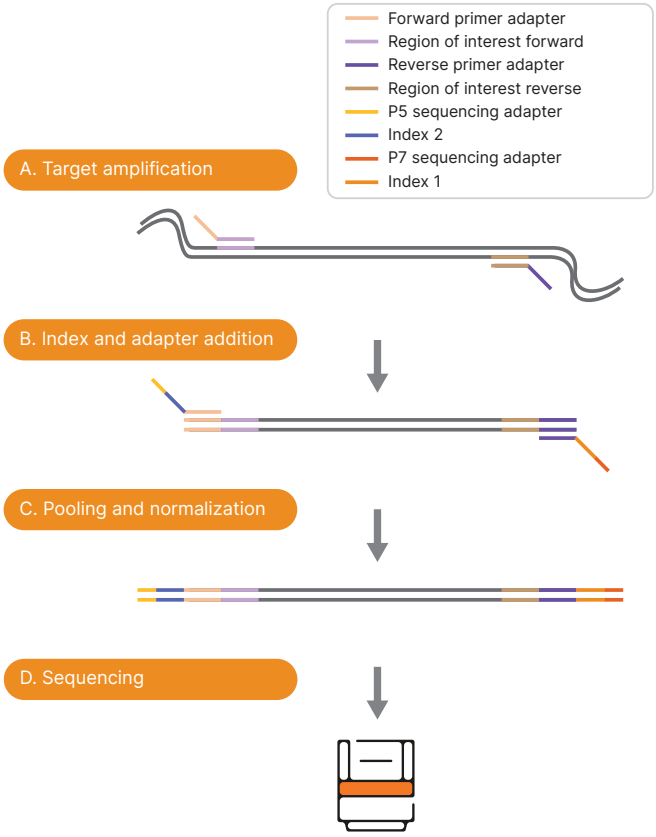

Figure 2: 16S V3 and V4 library generation—(A) Genomic DNA is amplified using 16S primers that include adapters for sequencing. (B) Indexes and sequencing adapters are added to the amplicon using IDT for Illumina DNA/RNA UD Indexes. (C) Libraries are pooled and normalized. (D) Sequencing is performed on a benchtop sequencing system.

Sequencing

Standard SBS 600-cycle kits

Following library preparation, 20 µl of 1000 pM 16S library with with 40% PhiX Control spike-in were sequenced with the NextSeq 1000/2000 P2 300M Reagents Kit (600 cycles) or 32% PhiX Control spike-in with the MiSeq Reagent v3 Kit (600 cycles). Sequencing was performed on the NextSeq 2000 System and the MiSeq System, respectively. Representative sequencing runs and analysis data are published on the BaseSpace demo data page.

XLEAP-SBS 600-cycle kit

Libraries were prepared from 96 technical replicates of the 20 Strain Staggered Mix Genomic Material (ATCC, Catalog no. MSA-1003) as described above. Following library preparation, 16S libraries were sequenced on the NextSeq 2000 System* with the NextSeq 1000/2000 P2 XLEAP-SBS Reagent Kit (600 cycles), including a 15% PhiX Control spike-in. Read length was 2 × 301 bp with dual indexing.

Data analysis

The 16S Metagenomics Labs app on BaseSpace Sequence Hub was used for downstream analysis exploring taxonomic classifications.

Zymo Research samples

Library preparation

Libraries for 16S analysis were prepared by the Microbiome Sequencing Service at Zymo Research. DNA from the ZymoBIOMICS Fecal Reference Material (Zymo Research, Catalog no. D6323) or ZymoBIOMICS Microbial Community Standard (Zymo Research, Catalog no. D6300) was extracted using the ZymoBIOMICS 96 MagBead DNA kit (Zymo Research, Catalog no. D4308). Libraries for 16S analysis were prepared using Quick-16S NGS Library Prep Kit (Zymo Research, Catalog no. D6400) with the V3-V4 primer. The primer was custom-designed by Zymo Research to provide sufficient coverage. The result library was auto-normalized during the targeted amplification step and pooled by equal volume.

* NextSeq 1000/2000 Reagent 600-cycle kits share performance specifications when used with either the NextSeq 1000 or NextSeq 2000 Systems, delivering high Q30 metrics and excellent uniformity.

Sequencing

Following library preparation, 16S libraries were loaded into either a NextSeq 1000/2000 P1 Reagents Kit (600 cycles) (with 30% PhiX Control spike-in), NextSeq 1000/2000 P1 XLEAP-SBS Reagent Kit (600 cycles) (with 15% PhiX Control spike-in), or a MiSeq Reagent v3 Reagent Kit (600 cycles) (with 15% PhiX Control spike-in). Sequencing was performed on the NextSeq 2000 System and the MiSeq System, respectively.

Data analysis

Unique amplicon sequences variants were inferred from raw reads using the DADA2 pipeline.[4] Potential sequencing errors and chimeric sequences were also removed with the DADA2 pipeline. Taxonomy assignment was performed using Uclust from QIIME v.1.9.1[5] with the Zymo Research Database, an internally designed and curated 16S rRNA database, as a reference. Composition visualization, alpha-diversity, and beta-diversity analyses were performed with QIIME v.1.9.1.

Results

Improved sequencing data quality

Samples sequenced using the NextSeq 1000/2000 P2 300M Reagents Kit (600 cycles) and the NextSeq 1000/2000 P2 XLEAP-SBS Reagent Kit (600 cycles) run on the NextSeq 2000 System both show improved Q30 read quality scores (Q-scores) compared to the MiSeq Reagent v3 Kit (600 cycles) run on the MiSeq System (Figure 3). Our results demonstrate that sequencing with XLEAP-SBS reagents delivers high-quality data, especially at the end of read length. Even with less than half the PhiX Control spike-in compared to the standard SBS NextSeq 1000/2000 P2 300M Reagents Kit (600 cycles), the NextSeq 1000/2000 P2 XLEAP-SBS Reagent Kit (600 cycles) still has the highest quality scores, which becomes more prevalent in the last 10 cycles (Figure 3). By reducing the percentage of PhiX Control spike-in required, using XLEAP-SBS reagents for sequencing on the NextSeq 2000 System provides ~7× more usable reads passing filter compared to sequencing on the MiSeq System using standard SBS reagents (Figure 4).

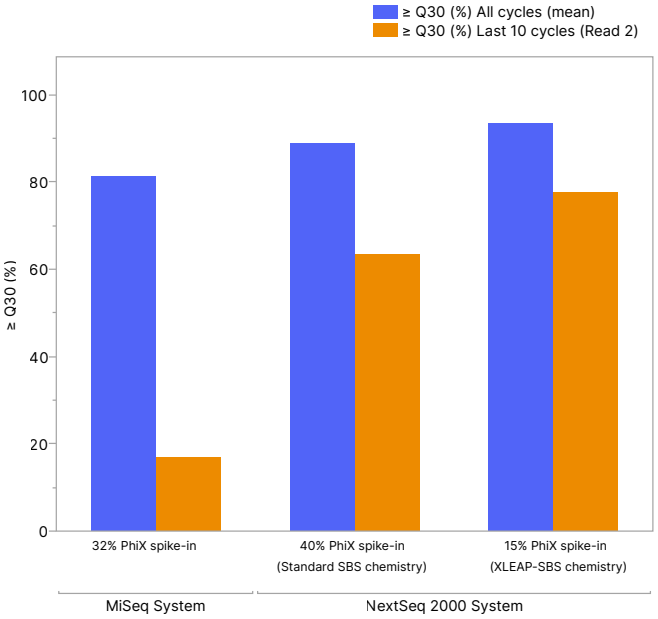

Figure 3: Primary sequencing performance metrics with ATCC control samples—Compared to 16S rRNA sequencing on the MiSeq System, the NextSeq 1000/2000 P2 300M Reagents Kit (600 cycles) and the NextSeq 1000/2000 P2 XLEAP-SBS Reagent Kit (600 cycles) run on the NextSeq 2000 System offer higher Q30 read quality scores, especially at the end of read length.

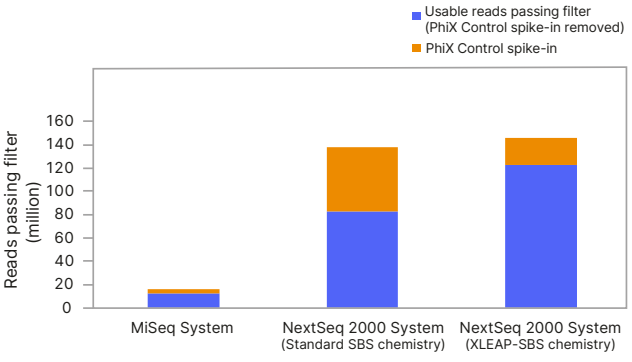

Figure 4: Improved data quality from ZymoBIOMICS Fecal Reference Material sequenced on the NextSeq 2000 System using the NextSeq 1000/2000 P1 XLEAP-SBS Reagent Kit (600 cycles)—A reduction in PhiX Control spike-in with the NextSeq 1000/2000 P1 XLEAP-SBS Reagent Kit (600 cycles) provides significantly more usable reads passing filter compared to the NextSeq 1000/2000 P1 Reagent Kit (600 cycles) and the MiSeq Reagent Kit v3 (600 cycles).

ATCC samples 16S analysis

To compare performance across systems and chemistries, the 20 Strain Staggered Mix Genomic Material (ATCC, Catalog no. MSA-1003) was sequenced on the NextSeq 2000 System with standard SBS and XLEAP-SBS reagents, or on the MiSeq System. Analysis of the 16S sequencing results identified all expected members of the bacterial community and showed comparable results between the NextSeq 2000 System (using both standard SBS and XLEAP-SBS chemistries) and the MiSeq System (Figure 5). The community profiles of all samples tested were also highly concordant between the NextSeq 2000 and MiSeq Systems (Figure 6). These results further reinforce the parity of performance between NextSeq 1000 and NextSeq 2000 Systems, with both standard SBS and XLEAP-SBS reagents, and the MiSeq System for 16S meta-genomics applications.

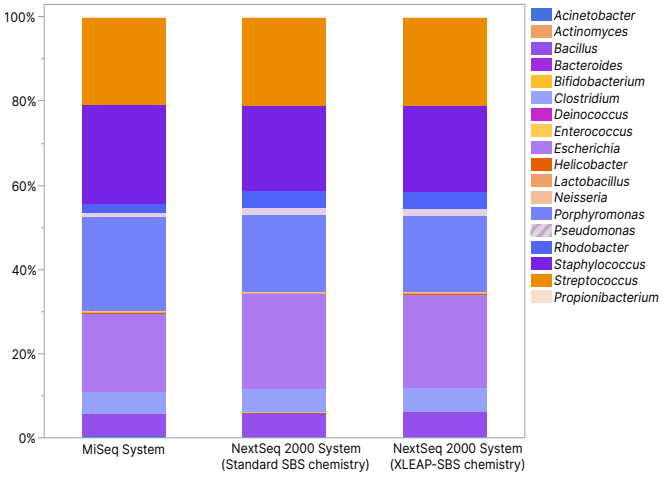

Figure 5: Comparative analysis of microbial composition of ATCC samples sequenced on NextSeq 2000 and MiSeq Systems—Analysis of microbial composition of ATCC samples with the NextSeq 2000 System using either the NextSeq 1000/2000 P1 Reagents Kit (600 cycles) or the NextSeq 1000/2000 P1 XLEAP-SBS Reagent Kit (600 cycles) demonstrates similar excellent genus-level coverage obtained by sequencing on the MiSeq System using the MiSeq Reagent v3 Kit (600 cycles).

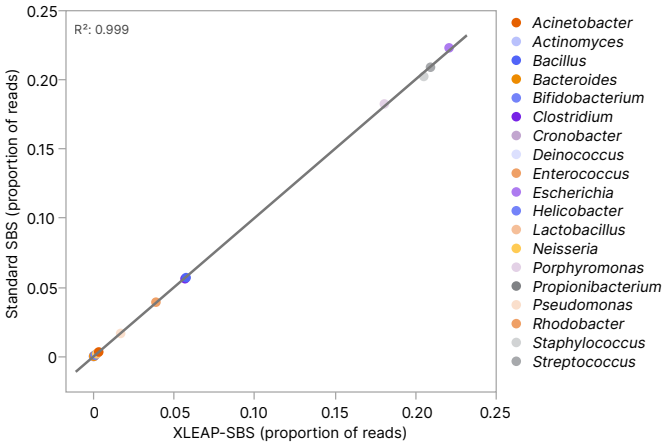

Figure 6: Comparative 16S analysis of the microbial composition of ATCC samples using standard SBS or XLEAP-SBS chemistries on NextSeq 2000 System—The proportional representation of bacterial genera quantified from samples run using the NextSeq 1000/2000 P2 300M Reagent Kit (600 cycles) with standard SBS and the NextSeq 1000/2000 P2 XLEAP-SBS Reagent Kit (600 cycles) is plotted. Analysis of samples for bacterial representation is highly concordant (R^2 = 0.999) between the two reagent kits.

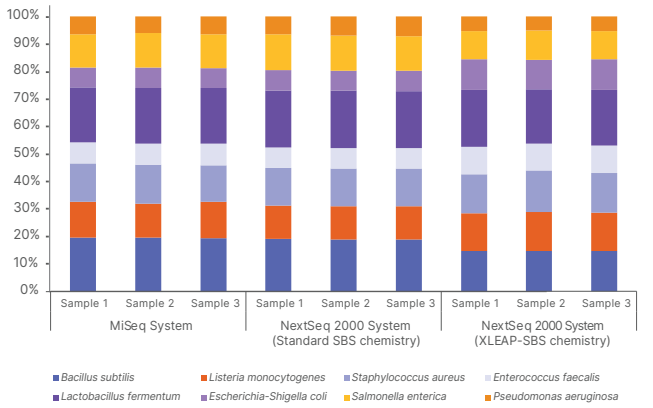

Figure 7: Comparative representation of eight highly represented 16S microbial sequences identified in the ZymoBIOMICS Microbial Community Standard using NextSeq 2000 and MiSeq Systems—Analysis of microbial composition of three samples with the NextSeq 2000 System using either the NextSeq 1000/2000 P1 Reagents Kit (600 cycles) or the NextSeq 1000/2000 P1 XLEAP-SBS Reagent Kit (600 cycles) demonstrates similar excellent species-level coverage obtained by sequencing on the MiSeq System using the MiSeq Reagent v3 Kit (600 cycles).

Zymo Research samples 16S analysis

The ZymoBIOMICS Community Standard with eight bacterial species from Zymo Research was used to compare performance of the NextSeq 2000 and MiSeq Systems. Similar to the performance with the ATCC reference samples, the 16S community profiles of the ZymoBIOMICS Microbial Community Standard were highly concordant between the NextSeq 2000 (using both standard SBS or XLEAP-SBS reagents) and MiSeq Systems (Figure 7). These results further support the use of the NextSeq 1000 and NextSeq 2000 Systems for 16S metagenomics applications with samples from different sources.

Working with low-diversity 16S libraries

Sequencing 16S rRNA libraries presents some unique challenges due to the lack of diversity in base composition resulting in a large percentage of the clusters having the same base during each cycle. The signals from this imbalance result in lower quality scores and impedes further analysis. This imbalance can be addressed by adding a compatible, well-balanced library to the flow cell. These libraries can be sourced from other samples or controls and can be used for error rate calculations.

Using XLEAP-SBS chemistry for sequencing on the NextSeq 2000 System, we demonstrate that high read quality can be achieved with a PhiX Control spike-in as low as 15% (Figure 3, Figure 8), compared to 40% with standard SBS chemistry. Users can determine loading concentration and PhiX Control spike-in based on experimental needs. For example, a 15% targeted spike-in may be a good fit if there is a need for higher output and the demonstrated Q30 is acceptable. If higher or maxmium Q30 is desirable, then a higher targeted spike-in should be used. Users should perform their own validation of libraries and confirm adapter compatibility. Note that percent occupied and percent loading concentration metrics are not relevant for unbalanced libraries, such as 16S amplicons.

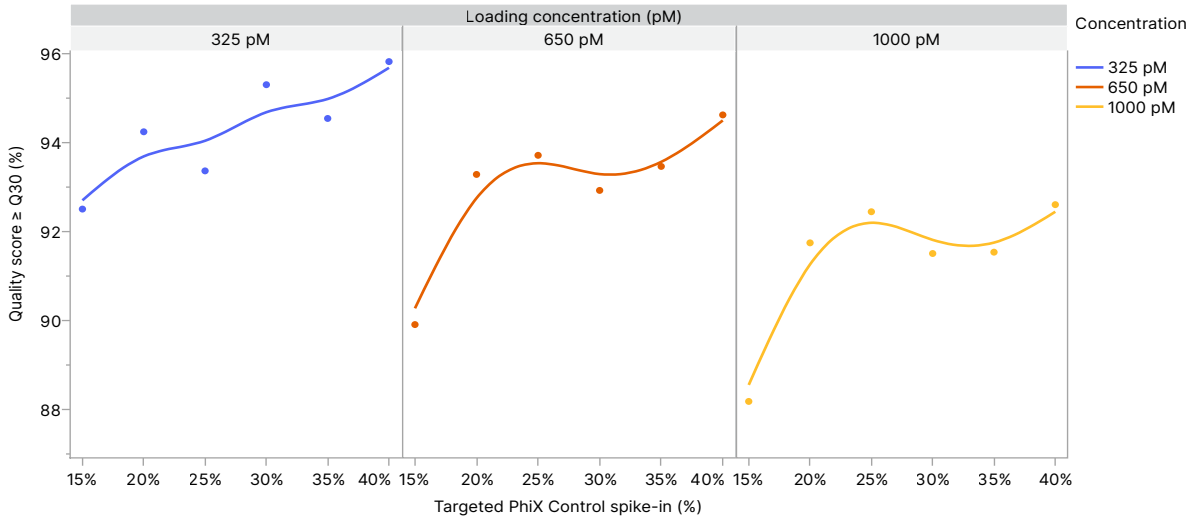

Figure 8: Read quality scores using varying PhiX Control spike-in percentages and loading concentrations— Loading ratios between PhiX Control spike-in and low sequence-diversity amplicons like 16S rRNA libraries can be adjusted to obtain the highest quality data. ATCC samples were sequenced on the NextSeq 2000 System with the NextSeq 1000/2000 P2 XLEAP-SBS Reagent Kit (600 cycles). High-quality data were obtained across a range of loading concentrations and PhiX Control spike-in percentages. Users can consider the balance of quality and yield to achieve desired experimental outcomes.

Summary

The results in this application note demonstrate comparable sequencing performance of the NextSeq 2000 and MiSeq Systems for 16S rRNA sequencing. The NextSeq 1000/2000 P2 300M Reagents Kit (600 cycles) and the NextSeq 1000/2000 P2 XLEAP-SBS Reagent Kit (600 cycles) are equivalent in read length to the MiSeq Reagent Kit v3 (600 cycles) that is commonly used for 16S NGS analysis. However, with 4× more sequencing output and a shorter run time, the NextSeq 1000 and NextSeq 2000 Systems significantly improve sequencing depth, turnaround time, quality scores, and scalability of 16S metagenomics studies. Additionally, using XLEAP-SBS reagents for sequencing on the NextSeq 1000 and NextSeq 2000 Systems lowers PhiX Control spike-in requirements, which provides more usable reads per flow cell while still delivering excellent data quality.

Learn more

NextSeq 1000 and NextSeq 2000 Systems

NextSeq 1000/2000 reagents

Demo data for 16S rRNA sequencing on the NextSeq 2000 System, run link, project link

16S Metagenomics Labs app

References

1. Clarridge JE 3rd. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev*. 2004;17(4):840-862. doi:10.1128/CMR.17.4.840-862.2004
2. Klindworth A, Pruesse E, Schweer T, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res*. 2013;41(1):e1. doi:10.1093/nar/gks808
3. Illumina. 16S metagenomic sequencing library preparation. support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf. Accessed February 9, 2023.
4. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13(7):581-583. doi:10.1038/nmeth.3869
5. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7(5):335-336. doi:10.1038/nmeth.f.303

illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-GL-01146 v2.0