APPLICATION NOTE

16S rRNA sequencing on NextSeq™ 1000 and 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 P1/P2 600-cycle kits. Sequencing performance is compared to the well-established performance of the MiSeq™ Reagent v3 600-cycle kit on the MiSeq System.

The 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 16S metagenomics workflow integrates library preparation, proven Illumina NGS, and push-button secondary data analysis through the 16S Metagenomics Labs app available on BaseSpace™ Sequence Hub (Figure 1).

Methods

IDT for Illumina DNA/RNA UD Indexes Sets A to D allow users to generate 384 16S libraries. Running 384 16S libraries on the NextSeq 1000/2000 P1 Reagents 600-cycle kit will generate between 100,000 and 200,000 reads per sample, which is sufficient for classification at the genus level. Up to 100M total reads can be generated on the NextSeq 1000/2000 P1 Reagent 600-cycle kit and up to 300M reads on the NextSeq 1000/2000 P2 Reagent 600-cycle kit.

Library preparation

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. In addition, real-world stool samples were obtained for analysis through collaboration with researchers at Stanford University, Stanford, CA.[2]

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).[3] Next, 10 µl Illumina sequencing adapters and unique dual-index barcodes (IDT for Illumina DNA/RNA UD Indexes Set A, Tagmentation,

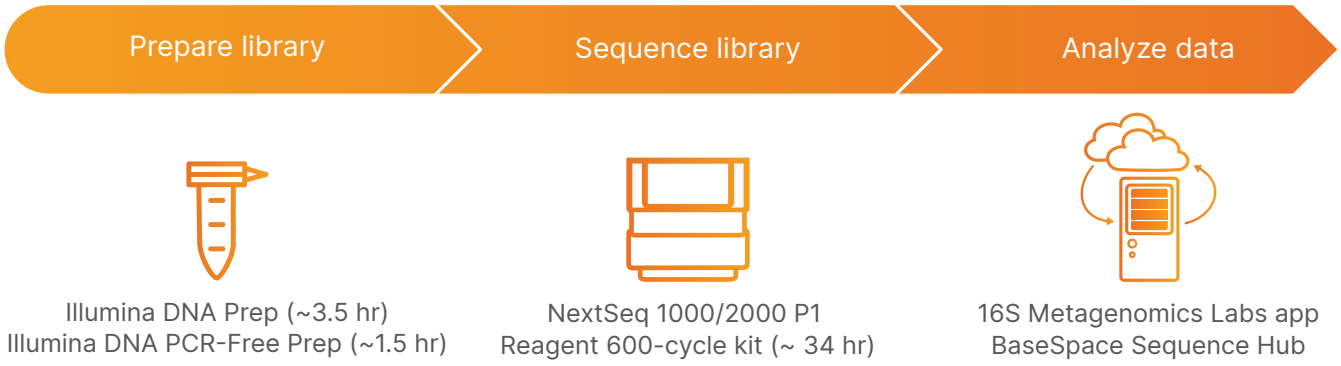

Figure 1: 16S metagenomic sequencing workflow—The 16S sequencing workflow includes library preparation, sequencing, and secondary data analysis. Sequencing run time for the NextSeq 1000/2000 P1 Reagents 600-cycles kit is ~34 hr and run time for the NextSeq 1000/2000 P2 Reagents 600-cycles kit is ~44 hr.

Catalog no. 20027213) were added in a second PCR, generating amplicons compatible with the NextSeq 2000 System. We recommend that customers use Illumina Purification Bead, 100 ml (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.[4] The overhang adapter sequence must be added to the locus-specific primer for the region to be targeted.

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 Illumina overhang adapter sequence. Primer sequence after the hyphen corresponds to locus-specific sequence.

Sequencing

Following library preparation, 20 µl of 1000 pM 16S library with 40% (v/v) PhiX spike-in library loaded onto flow cells from either a NextSeq 1000/2000 P1 Reagents 600-cycle kit or a MiSeq Reagent v3 600-cycle kit. 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.

To assess base diversity spike-in, an independent 40% Illumina DNA Prep well-balanced library was prepared in two technical replicates using three genomic bacterial isolates (*E. coli*, *B. cerus*, and *R. sphaeroides*; ATCC, Catalog no. 700926, 10987, and 17023, respectively) and added to a separate 16S pool.* Libraries were prepared with Illumina DNA Prep, (M) Tagmentation (24 Samples, IPB) (Illumina, Catalog no. 20060060) and IDT for Illumina DNA/RNA UD Indexes Set A, Tagmentation (96 Indexes, 96 Samples) (Illumina, Catalog no. 20027213). Representative sequencing runs and analysis data are published on the BaseSpace demo data page.

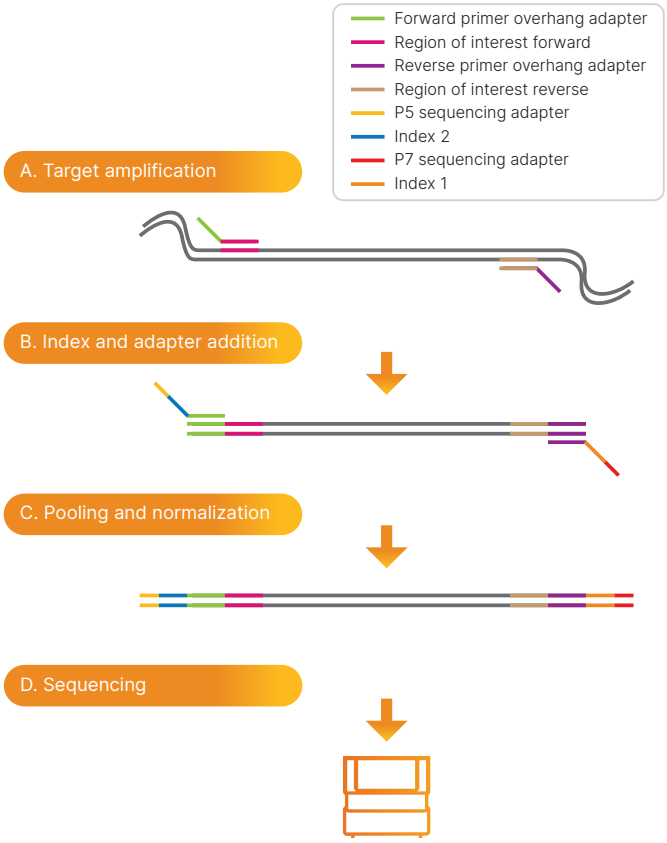

Figure 2: 16S V3 and V4 library generation—(A) Genomic DNA is amplified using 16S primers that include overhang adapters for library generation. (B) Indexes and sequencing adapters are added to the amplicon using IDT for Illumina DNA/RNA UD Indexes. (C) Libraries are pooled and normalized. (D) Sequencing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

Results

Improved primary sequencing metrics

The NextSeq 1000/2000 Reagent 600-cycle kit on the NextSeq 2000 System shows improved Q30 read quality scores (Q-scores) compared to the MiSeq Reagent v3 600-cycles kit run on the MiSeq System. The NextSeq

* Check index compatibility before adding other well-balanced libraries. Individual evaluation is required to ensure each sample has enough reads.

2000 System also provided four-fold more sequencing output over the MiSeq System and in ~ 20 hr less time per sequencing run (Figure 3).

ATCC samples 16S analysis

To compare performance across systems, the 20 Strain Staggered Mix Genomic Material (ATCC, Catalog no. MSA-1003) was sequenced on the NextSeq 2000 and MiSeq Systems. The 16S Metagenomics Labs app on BaseSpace Sequence Hub was used for downstream analysis exploring taxonomic classifications. Analysis of the 16S sequencing results identified all expected members of the bacterial community and showed comparable results between the NextSeq 2000 and the MiSeq Systems (Figure 4). The community profiles of all samples tested were also highly concordant between the NextSeq 2000 and MiSeq Systems (Figure 5). These results further reinforce the parity of performance between the MiSeq and the NextSeq 1000 and NextSeq 2000 System for 16S metagenomics applications.

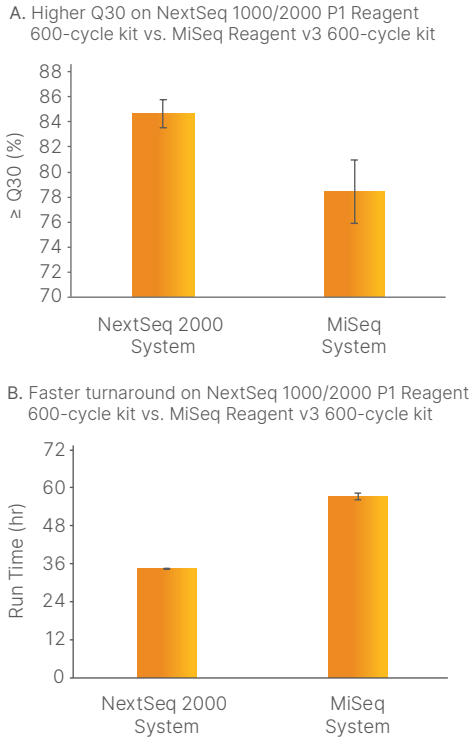

Figure 3: Primary sequencing performance metrics—Compared to 16S rRNA sequencing on the MiSeq system, the NextSeq 1000/2000 P1 Reagent 600-cycle kit on the NextSeq 2000 System offers (A) higher Q30 read quality scores and (B) ~ 20 hr shorter sequencing run times.

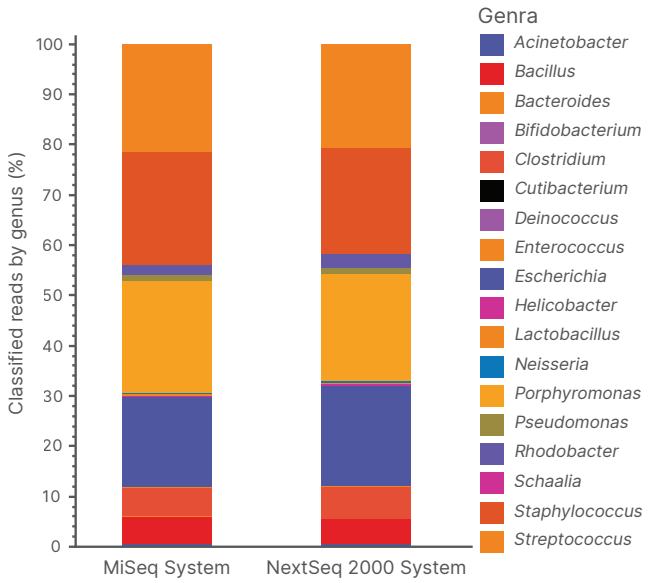

Figure 4: Comparative analysis of microbial composition of ATCC samples across NextSeq 2000 and MiSeq Systems—Analysis of microbial composition of ATCC samples with the NextSeq 2000 and MiSeq systems demonstrates similar excellent genera coverage.

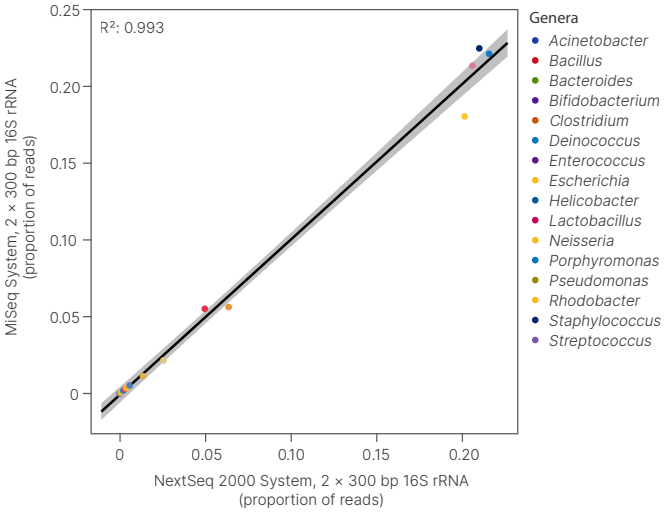

Figure 5: Comparative 16S analysis of the microbial composition of ATCC samples for NextSeq 2000 and MiSeq Systems—The proportional representation of bacterial genera quantified from samples run on each system is plotted. Analysis of samples for bacterial representation is highly concordant (R2 = 0.993) between the NextSeq 2000 and MiSeq Systems.

Stool sample 16S analysis

To demonstrate performance with real-world stool samples, 10 of the highest represented genera were compared across the NextSeq 2000 and MiSeq Systems. A representative sample is shown to demonstrate the similar distribution for the highest genera found on each system (Figure 6). Similar to the performance with the ATCC reference samples, the 16S community profiles of all real-world stool samples were highly concordant between the NextSeq 2000 and MiSeq Systems (Figure 7). These results further support the use of the NextSeq 2000 for 16S metagenomics applications with samples from different sources.

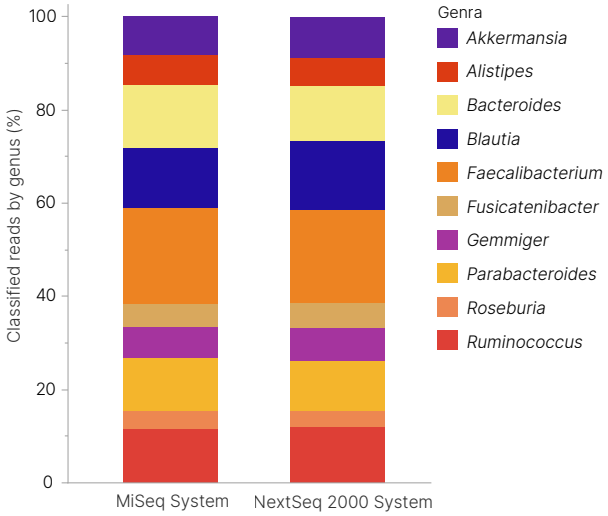

Figure 6: Comparative representation of ten highly represented 16S microbial sequences identified in stool samples using NextSeq 2000 and MiSeq Systems—Analysis of microbial composition of stool samples with the NextSeq 2000 and MiSeq Systems demonstrates similar excellent genera coverage.

Working with low-diversity 16S libraries

Sequencing low-diversity libraries, such as 16S rRNA sequencing libraries, presents some unique challenges due to the unbalanced base composition causing a large percentage of the clusters to show the same base during each cycle. The high signals caused by this imbalance results in low quality scores and impedes further analysis. This imbalance can be addressed by adding a compatible, well-balanced library to the flow cell. A well-balanced

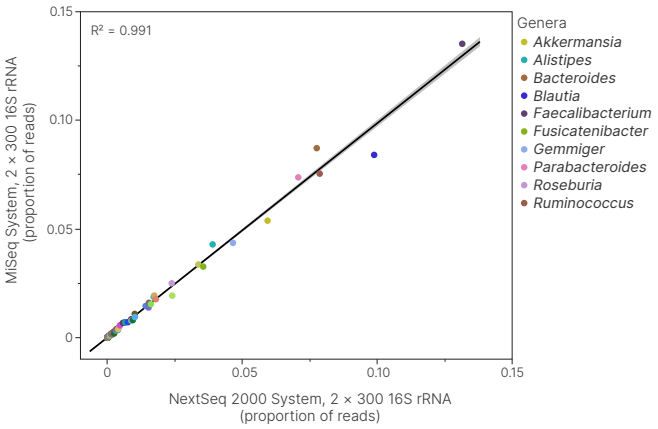

Figure 7: Correlation of 16S microbial composition identified in stool samples using NextSeq 2000 and MiSeq Systems —The proportional representation of bacterial genera quantified from samples run on each system is plotted. Analysis of samples for bacterial representation is highly concordant (R2 = 0.991) between the NextSeq 2000 and MiSeq systems.

spike-in library can be used to calculate a base call rate that can be extrapolated to the 16S samples. The well-balanced libraries can be sourced from other samples or controls and can be used for error rate calculations. The 40% library addition is also a starting recommendation and can be optimized by labs to meet their experimental needs. 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.

Even with the addition of a 40% (v/v) library for this run, the NextSeq 1000/2000 600-cycle kits allow for a high plexity of amplicon samples. For example, when using a 40% PhiX spike-in library, the NextSeq 1000/2000 P1 Reagent 600-cycle kit provides 60M reads for 16S samples (Table 2). Generally, 15,000–100,000 reads per sample are sufficient for 16S classification.[5–7]

Table 2: NextSeq 1000/2000 P1 Reagent 600-cycle kit reads/16S sample library when using 40% spike-in library

	Single reads (clusters) per sample	Paired-end reads per sample
96 samples	625,000	1,250,000
384 samples	156,250	312,500

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 P1 Reagent 600-cycle kit is equivalent in read length to the MiSeq Reagent v3 600-cycle kit that is commonly used for 16S NGS analysis. However, with four times more sequencing output and a ~20 hr shorter run time, the NextSeq 1000 and NextSeq 2000 Systems significantly improve on depth on sequencing, turnaround time, Q30 quality scores and scalability of 16S metagenomics studies. The NextSeq 1000 System functions the same as the NextSeq 2000 with the described 600-cycle kits and identical performance specifications between these instruments.

Learn more

NextSeq 1000 and NextSeq 2000 Systems

NextSeq 1000/2000 reagents

16S sequencing run with PhiX control library, project link, run link

16S sequencing run with Illumina DNA Prep control library, project link, run link

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. Maghini D, Dvorak M, Dahlen A, Roos M, Kuersten S, Bhatt AS. Achieving quantitative and accurate measurement of the human gut microbiome. *bioRxiv*. 2022;09.28.509972; doi: https://doi.org/10.1101/2022.09.28.509972
3. 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
4. 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.
5. Peterson D, Bonham KS, Rowland S, Pattanayak CW; RESONANCE Consortium, Klepac-Ceraj V. Comparative Analysis of 16S rRNA Gene and Metagenome Sequencing in Pediatric Gut Microbiomes. *Front Microbiol*. 2021;12:670336. Published 2021 Jul 15. doi:10.3389/fmicb.2021.670336
6. Sanchez-Cid C, Tignat-Perrier R, Franqueville L, Delaurière L, Schagat T, Vogel TM. Sequencing Depth Has a Stronger Effect than DNA Extraction on Soil Bacterial Richness Discovery. *Biomolecules*. 2022;12(3):364. Published 2022 Feb 25. doi:10.3390/biom12030364
7. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol*. 2013;79(17):5112-5120. doi:10.1128/AEM.01043-13

illumina®

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

M-GL-01146 v1.0