Illumina 16S Metagenomic Sequencing Protocol

Frequently asked questions about the library preparation protocol for 16S metagenomics studies.

Metagenomics studies are useful for identifying the microbial species present in a sample. They commonly involve sequencing the 16S ribosomal RNA (rRNA) gene for taxonomic classification. The demonstrated Illumina protocol for 16S metagenomic sequencing library preparation targets the V3 and V4 variable regions of the 16S rRNA gene. It involves a two-step, tailed PCR approach that generates ready-to-pool amplicon libraries. This document provides guidelines and best practices for designing 16S metagenomic sequencing projects following the Illumina protocol.

Library Preparation

What are the advantages of a two-step, tailed amplicon approach over single, long PCR primers?

Locus-specific primers contain sequence tails that allow for a second PCR to add Nextera® XT indexed adapters. Tailed primers increase melting point, efficiency, and specificity while avoiding the disadvantages of long primers, such as hairpins, self-dimers, primer dimers, and chimeras.

Can I combine the two-step PCR into a single reaction?

Combining all four primers into a single reaction can result in cross-binding among primer sequences and spurious PCR product synthesis. Illumina does not recommend mixing the primers into a single reaction as it can result in non-specific products and poor library yield.

Can I use high-performance liquid chromatography (HPLC)–purified oligos?

Illumina recommends using standard desalting purification when ordering oligo primer sets. The purity required for a specific application depends on the potential complications caused by the presence of truncated oligomers, which must be evaluated empirically on a small batch of samples. Oligo DNA synthesis often results in a mixture of full-length products and truncated sequences, and the contamination can compromise amplification of the full-length product during PCR.

What resources and databases can I use to find validated oligos?

For more information about validated oligos, consult:

- Klindworth et al. (2013)[1]—An evaluation of 16S rDNA primers for sequencing
- Baker et al. (2003)[2]—A review and reanalysis of domain-specific 16S sequencing primers
- Probe Match[3]—A 16S rDNA sequence database
- TestPrime[4]—A 16S rDNA primer evaluation with *in silico* PCR

Can I use DesignStudio™ software to design and order 16S rRNA oligos?

Do not use DesignStudio software to design primers for two-step PCR. Use a dedicated oligo synthesis company to order 16S rRNA oligos. Further guidelines and recommendations are provided in the 16S Metagenomic Sequencing Library Preparation Guide[5].

Will using PCR additives compromise sequencing quality?

PCR additives such as betaine, dimethyl sulfoxide (DMSO), and 7-deaza-dGTP will not compromise downstream amplicon sequencing as long as they are used at concentrations that do not inhibit PCR amplification. Determine the appropriate concentration empirically.

What primer sequences are used in the protocol?

The primer sequences used in the protocol are:

- PCR1_Forward (50 bp): 5'–TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG–3'
- PCR1_Reverse (55 bp): 5'–GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC–3'

What multiplexing capabilities are available with the protocol?

Nextera XT Index Kits use a dual-index strategy involving 12 (i7) and 8 (i5) indexes, for a total of 20 indexes. Together, they can generate 96 (12 × 8) different index combinations. For higher levels of multiplexing, Nextera XT v2 Index Kits allow up to 384 samples to be pooled in one flow cell lane. Nextera XT v2 Index Kits are available in four sets of 96 index combinations each, for a total of up to 384 samples per set.

How many samples can I pool per run?

The number of samples that can be pooled per run depends on the scope and goal of a given sequencing project. Generally, 15,000–100,000 reads per sample are sufficient for classification. For information about pooling strategies for a few samples, refer to the Nextera Low Plex Pooling Guidelines Technical Note[6]. Illumina recommends creating a sample sheet in Illumina Experiment Manager (IEM) before preparing libraries to confirm that the appropriate index combinations are used. If using MiSeq Control Software (MCS) 2.4.1 or 2.5 with real-time analysis (RTA) software 1.18.54, index confirmation is not necessary.

Sequencing

What are the recommended sequencing guidelines?

Illumina recommends using MiSeq® v3 sequencing chemistry. MiSeq v2 reagents are not recommended for 16S metagenomics studies, as they will affect quality scores and other metrics. A 5% PhiX spike-in is required with MiSeq v3 chemistry. Table 1 shows the expected yield for 16S sequencing runs using MiSeq v3 chemistry. With MiSeq Reporter software (2.3 or later), 50 bp of read overlap is required to take advantage of the read-stitching feature.

Table 1: MiSeq v3 Chemistry Specifications

Parameter	Value
Paired-end reads	44–50 M
Output	13.2–15 Gb
Quality scores	> 70% of bases higher than Q30 at 2 × 300 bp

Why are Q-scores lower in low-diversity libraries? How can I judge run performance?

Sequencing shotgun libraries generates intensity profiles, and the clusters on a flow cell represent roughly even proportions of A, C, G, and T at every cycle. In low-diversity libraries, such as those used for amplicon or 16S sequencing, the proportion of bases is unbalanced, so a large percentage of the clusters show the same base during each cycle. The high signals caused by the imbalance affect the base calling algorithm, resulting in low Q-scores even though the base calling accuracy is not necessarily poor. The 5% PhiX spike-in enables error rate calculations that allow verification of base calling accuracy over the course of the run, for all PhiX clusters, which can be extrapolated to the samples.

Does 16S sequencing require separate sequencing primers?

The required sequencing primers for the Illumina 16S metagenomics workflow are included in the MiSeq reagent cartridge and in HiSeq® 2500 Rapid-Run reagents. If using the HiSeq 2000 or HiSeq 2500 System for 16S sequencing, the TruSeq® Dual Index Sequencing Primer Box is required.

Is there a minimum percentage of sequence identity required for genus- and species-level classifications using the Illumina 16S metagenomics workflow?

The percentage of sequence identity needed depends on the organism under study. For more divergent species, longer reads can aid assembly and classification.

Is there a minimum coverage or percentage of reads required for genus and species identification following the Illumina 16S metagenomics workflow?

Only one read is required for the single-read classification algorithm. Consider sample diversity when determining the number of reads to sequence.

What are possible sources of bias and error in metagenomic analyses?

Bias is defined as misrepresentation of the relative abundances of microbial populations in a sample. Error is defined as misrepresentation of a sequence due to PCR amplification and sequencing. When performing 16S sequencing, researchers can use a mock community with known composition to determine error rates[7].

Sources of bias:

- The method of DNA extraction and purification
- PCR primer selection
- Cycling conditions
- Community composition
- Number of 16S rRNA copies per genome

Sources of error:

- Polymerase error rate
- Chimera formation from incomplete PCR products
- Errors introduced during sequencing

Where can I find a control metagenomics sample for 16S rRNA sequencing?

Illumina recommends using the microbial mock community B sample available from BEI Resources (HM-276D)[8].

Analysis

What analysis software can I use?

This application is supported on the MiSeq System using the 16S Metagenomics workflow in MiSeq Reporter (v2.3) software or the BaseSpace® (v2.2) environment. Alternatively, data can be analyzed using third-party software. For more information, refer to the MiSeq Reporter Metagenomics Workflow Reference Guide[9], the Metagenomics Workflow Quick Reference Card[10], or the Third-Party Analysis Software and Utilities Technical Note[11].

Does the Illumina 16S rRNA database exclude species that are difficult to identify (e.g., *Bacillus anthracis*, *Bacillus cereus*, *Bacillus globisporus*, *Bacillus psychrophilus*)? Do I exclude these reads from my data?

No, there is no need to filter raw data before analysis. MiSeq Reporter software uses a version of the Greengenes database (13_5) to perform taxonomic classification. If the algorithm cannot classify a sequence read at the species level, then it will classify the read at higher taxonomic levels.

References

1. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, et al. (2013) Evaluation of a general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41: e1.
2. Baker GC, Smith JJ, Cowan DA (2003) Review and re-analysis of domain-specific 16S primers. J Microbiol Methods 55: 541–555.
3. Probe Match (rdp.cme.msu.edu/probematch/search.jsp) Accessed 13 August 2014.
4. TestPrime (http://www.arb-silva.de/search/testprime) Accessed 13 August 2014.
5. Illumina (2013) 16S metagenomic sequencing library preparation guide. (support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf)
6. Illumina (2011) Nextera low plex pooling guidelines. (www.illumina.com/documents/products/technotes/technote_nextera_low_plex_pooling_guidelines.pdf)
7. Schloss PD, Gevers D, Westcott SL (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS One 6: e27310.
8. BEI Resources (www.beiresources.org/Catalog/otherProducts/HM-276D.aspx) Accessed 13 August 2014.
9. Illumina (2014) MiSeq Reporter Metagenomics workflow reference guide. (support.illumina.com/documents/documentation/software_documentation/miseqreporter/miseq-reporter-metagenomics-workflow-ref-guide-15042317-c.pdf)
10. Illumina (2013) Running the MiSeq Metagenomics workflow. (support.illumina.com/documents/documentation/system_documentation/miseq/qrc/running_miseq_metagenomics_workflow_15040362_b.pdf)
11. Illumina (2013) Third-party analysis software and utilities. (support.illumina.com/documents/documentation/software_documentation/technote_thirdparty_software.pdf)

Illumina • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY

Pub. No. 1270-2014-003 Current as of 16 September 2014

illumina®