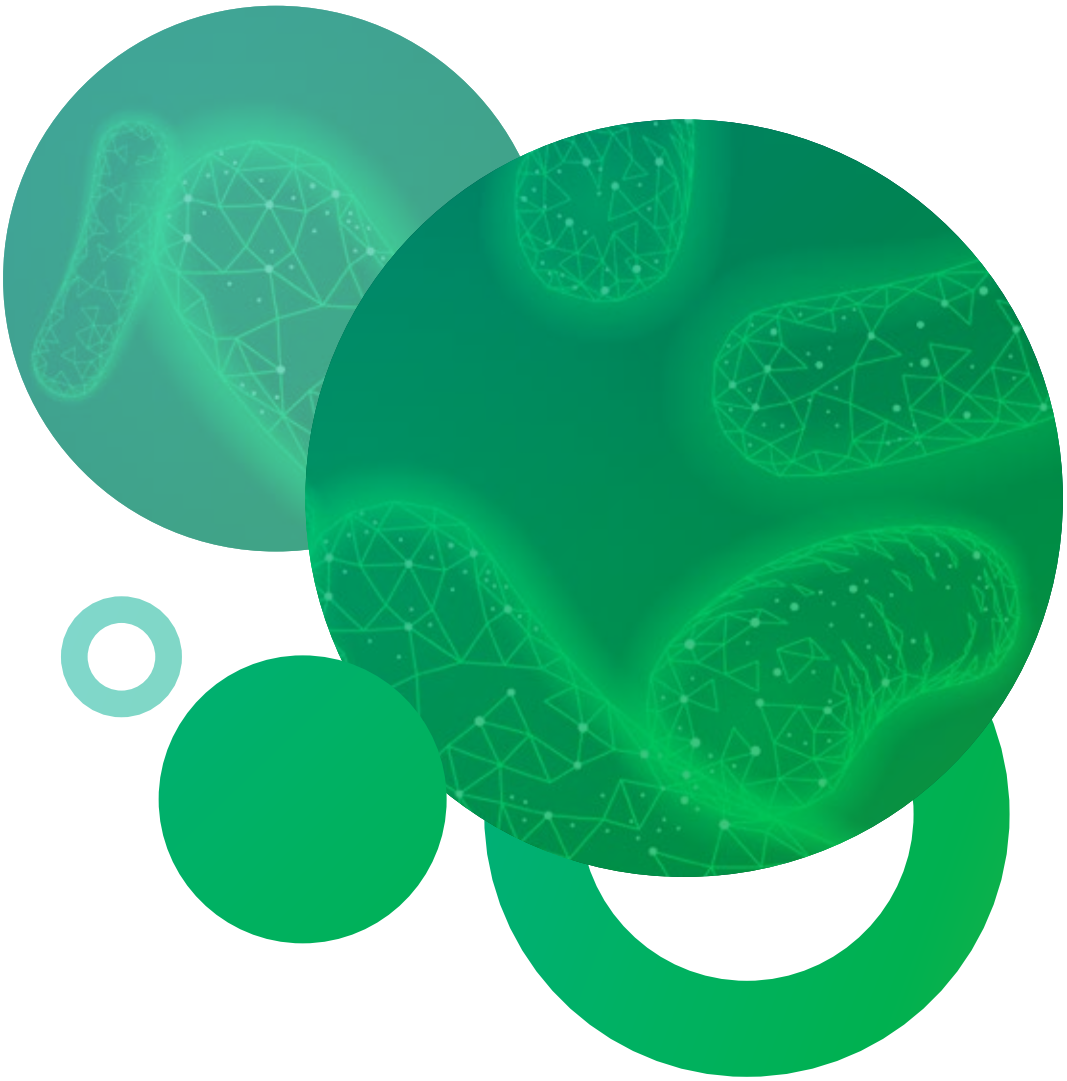
ZYMO RESEARCH

The Beauty of Science is to Make Things Simple

Considerations for Choosing a 16S rRNA Gene Library Preparation Protocol



www.zymoresearch.com

Since the beginning of the microbiome era decades ago, 16S rRNA gene-based amplicon sequencing has been widely adopted for microbial composition profiling. Many researchers might assume the 16S sequencing workflow is now well-optimized and selecting a heavily-cited protocol is trivial. However, there are aspects and challenges in 16S sequencing that should be considered before adopting a protocol.

Which 16S Hypervariable Region to Choose?

The 16S rRNA gene contains nine hypervariable regions (V1-V9) separated by conserved regions. These conserved regions provide primer binding sites that cover a broad range of phylogeny while the hypervariable regions contain species-specific sequence diversity used for taxonomic identification. This makes the 16S rRNA gene the perfect target for microbiome profiling with short read sequencing. However, each hypervariable region has different levels of conservation and sequencing different regions can cause different outcomes in the observed microbial profiles (Tables 1 & 2). A better understanding of these differences will help researchers determine the most suitable region(s) for their research needs. For example, targeting regions V1-V3 provides the most sequence variability, which means doing so has the best potential for taxonomy resolution (Figure 1). On the other hand, the V1-V3 region has poor coverage for Archaea and thus may not be suitable for some sample types, such as environmental samples. The 27f/8f forward primer (AGAGTTTGATCCTGGCTCAG), which is commonly used in V1-V2 and V1-V3 sequencing, also has 3 mismatches to the 16S sequence of *Bifidobacterium*, causing poor sensitivity for this genus and closely related microbes [6]. This is problematic for those researching the human gut microbiome since *Bifidobacterium* is an important constituent of the gut microbial community.

Table 1. Primers and amplicon lengths targeting 16S hypervariable regions

Region	Forward Primer(s)	Reverse Primer	Size	Reference
V3-V4 (Klindworth, et al)	341f: CCTACGGGNGGCWGCAG	806R: GACTACHVGGGTATCTAATCC	~460 bp	[1]
V3-V4 (Zymo Research)	341F: CCTACGGGDGGCWGCAG, CCTAYGGGGYGCWGCAG	806R: GACTACNVGGGTMTCTAATCC	~460 bp	Zymo Research
V4	515F: GTGYCAGCMGCCGCGGTAA	806R: GGACTACNVGGGTWTCTAAT	~292 bp	[2]
V1-V3	8F: AGRGTTTGATYMTGGCTYAG	534R: TBACCGCGGCTGCTGGCAC	~520 bp	Zymo Research
V4-V5	515F: GTGYCAGCMGCCGCGGTAA	926R: CCGYCAATTYMTTTRAGTTT	~410 bp	[3]
V6-V8	926F: AAACTYAAAKGAATTGACGG	1392R: ACGGGCGGTGTGTRC	~470 bp	[4]
V3	338f: ACWCCTACGGGNGGCWG	537R: GWNTTACCGCGGCKGCT	~200 bp	Zymo Research
V1-V2	8F: AGRGTTTGATYMTGGCTYAG	341R: CTGCWGCCNCCCGTAGG	~340 bp	Zymo Research

Table 2. Properties of 16S hypervariable regions for 16S rRNA gene sequencing. The coverage of the primers was determined using the TestPrime program from SILVA (https://www.arb-silva.de/search/testprime/) on August 10, 2021.

Region	Coverage (Silva)			Comments
	Bacteria	Archaea	Eukarya	
V3-V4 (Klindworth, et al)	82.80%	0.30%	0%	Very low coverage of Archaea
V3-V4 (Zymo Research)	84.20%	76.10%	0%	Excellent coverage of both Archaea and Bacteria. Better taxonomy resolution than V4 alone.
V4	83.60%	83.50%	0.10%	Excellent coverage of both Archaea and Bacteria. Limited taxonomy resolution.
V1-V3	82.60%	0%	0%	Best taxonomy resolution and good bacteria coverage, but no coverage of Archaea.
V4-V5	81.00%	84.50%	80.80%	Great coverage for all. But the good coverage of Eukaryotes is a double-edge sword.
V6-V8	78.50%	69.30%	71.30%	Great coverage for all. But the good coverage of Eukaryotes is a double-edge sword.
V3	86.70%	0%	0.10%	Small amplicon. Good for profiling fragmented DNA.
V1-V2	81.60%	0%	0%	Good taxonomy resolution and good bacteria coverage, but no coverage of Archaea.

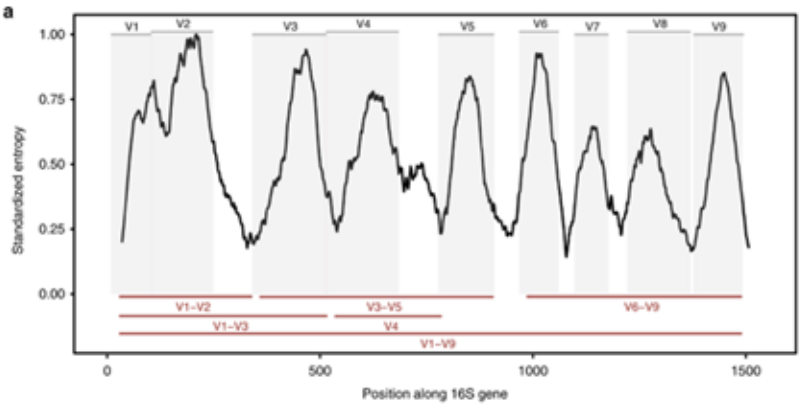


Figure 1. Entropy plot of 16S rRNA gene sequences [5]. The nine hypervariable regions along the 16S rRNA gene have differing levels of sequence diversity, which determines the identification resolution possible. The higher the curve is, the more variable the region is.

Alternatively, the V4 region provides the broadest coverage of Bacteria and Archaea, which is why the Earth Microbiome Project 16S protocol uses this region [2]. However, the V4 region is not as variable as other regions and therefore has limited taxonomy resolution (Figure 1). The V3-V4 region on the other hand, is popular due to its balance of taxonomy coverage and resolution [1,7]. And, adding just one additional primer (CCTAYGGGGYGCWGCAG) gives the V3-V4 region good coverage for Archaea as well. The other 16S variable regions are not as widely utilized but they each have pros and cons (Table 1).

The Dilemma of Controlling PCR Chimera

When amplification of more than one similar template occurs in one single well, chimeric products derived from different templates can be produced. These PCR chimeras are often produced in targeted 16S amplifications. The hybridization normally happens in conserved regions that are shared by different parent templates. Without proper control, chimeric sequences can represent >50% of all PCR products. These chimeric sequences can introduce noise or false positives in subsequent taxonomy analysis. Many factors affect the formation of PCR chimera including the DNA polymerase, template concentration, PCR cycles, the target 16S regions, and the diversity of the sample. In order to suppress PCR chimera formation, many 16S protocols recommend limiting template inputs and PCR cycles. For example, the Illumina 16S library prep protocol recommends 12.5ng of input microbial DNA and 25 PCR cycles [7]. However, for samples of lower microbial load, 25 PCR cycles is not enough to produce sufficient PCR product required for NGS. Additionally, choosing a fixed number of PCR cycles for all samples may overrun the PCR reaction and cause PCR chimera in samples of high microbial load.

How Accurate are the Various 16S Library Prep Protocols?

Protocols targeting the V1-V3 regions are attractive because they can achieve species-level resolution. On the other hand, targeting the V4 region provides superior coverage of Archaea. Both data accuracy and potential biases should be considered when deciding on a protocol. For example, many factors of the PCR process can cause uneven amplification, e.g. degeneracy of the primers, the sequence and size variation of the target region, the DNA polymerase, annealing temperature, and extension time. These errors may cause an underestimation or overestimation of the abundance of certain microbes. Without proper assessment and optimization, 16S library preparation protocols are likely to introduce unexpected biases.

Can the Normalization Process be Improved?

For most NGS library preparations, including 16S, a process called normalization is required prior to sequencing. Samples must be pooled to the same concentration, or "normalized", in the final library to ensure equal coverage during sequencing. This process typically consists of multiple steps: (1) the library of each sample undergoes a clean-up to remove enzymes, nucleotides, and primers leftover from PCR, (2) each clean library is then quantified by TapeStation®, Nanodrop™, Qubit® or qPCR, (3) a calculated volume of each library is pooled together based on equal molarity to form the final sequencing library. For a 96-well plate, this process can take several hours of hands-on time. Such a tedious process is prone to human errors and cross-contamination.

How Does the *Quick*-16S™ Plus NGS Library Prep Kit Address these Challenges?

The *Quick*-16S™ Plus NGS Library Prep Kit (V3-V4) features an innovative, auto-normalizing 1-step PCR amplification method that addresses the various challenges of 16S library preparation.

The choice of hypervariable regions to target was first carefully considered. The *Quick*-16S™ Plus NGS Library Prep Kit (V3-V4) uses primers targeting the V3-V4 region as this region gives the best balance between taxonomy resolution and coverage. Compared to the V4 region alone (popular in use for its broad coverage), the inclusion of the V3 region allows for better taxonomy resolution. Furthermore, compared to commonly used V3-V4 primer sets [1], an additional forward primer

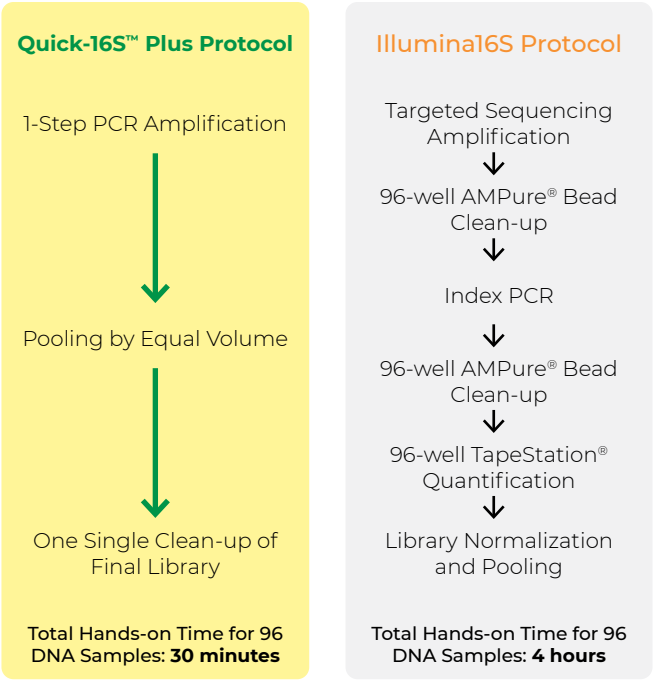


Figure 2. Quick-16S Plus workflow versus the Illumina 16S protocol [7]. Total hands-on time calculations are based on the preparation of 96 DNA samples.

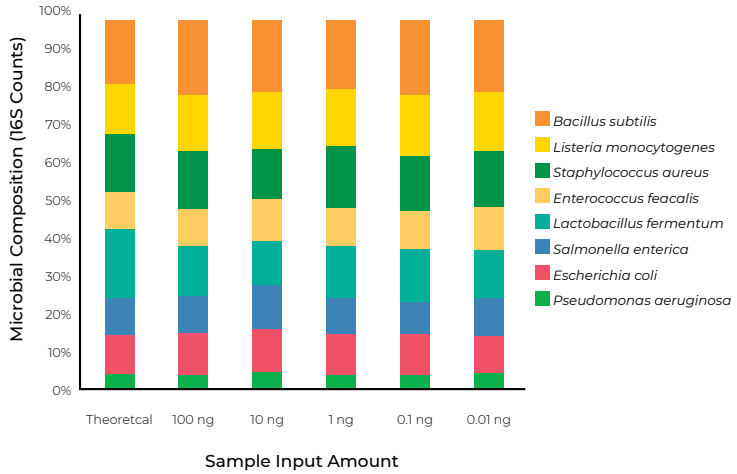


Figure 3. Benchmark the performance with ZymoBIOMICS Microbial Community DNA Standard. With the Quick-16S Plus NGS Library Prep Kit (V3-V4), microbial composition profiles are accurate from 100 ng down to 0.01 ng of the ZymoBIOMICS® Microbial Community DNA Standard as input.

(CCTAYGGGGYGCWGCAG) is included to add ~76% coverage of the Archaea kingdom (Table 1). This makes the *Quick*-16S™ Plus NGS Library Prep Kit (V3-V4) a great choice for profiling either human microbiome or environmental samples.

During development, many combinations of DNA polymerases, amplification buffers, and PCR conditions were tested in order to develop a single PCR protocol that could handle samples with a broad range of microbial load and minimize PCR chimera formation without controlling PCR cycle number. Out of this optimization came a new amplification system, named "Equalase", that addresses the challenges described earlier, including chimera formation, PCR cycle control, input template control, normalization, and accuracy. Equalase condenses the 16S library preparation process into one simple PCR reaction (Figure 2). With this new kit, the PCR reaction can be run for 40+ cycles without the worry of PCR chimera formation (typically <10%), which ensures excellent sensitivity even for samples with extremely low microbial load. With Equalase, microbiome profiling is very stable when amplifying from a wide input range of template DNA (from 0.01ng to 100 ng, Figure 3). Equalase is so named because it also automatically normalizes the PCR products during amplification (Figure 4). As a result, the traditional normalization process of pooling by equal molarity is not needed. After PCR amplification, all PCR products are simply pooled by equal volume. A single one-tube PCR cleanup is then performed prior to sequencing (Figure 2).

It is common practice to quantify each individual sample library before pooling together. This is also used as a QC measure to ensure each sample library amplified successfully. The *Quick*-16S™ Plus kit streamlines quantification while also adding the ability to measure absolute abundance of each sample. The Equalase mastermix contains a fluorescence dye which, in conjunction with included 16S rRNA gene standards, allows for qPCR quantification of total 16S rRNA gene copy numbers. (Figure 5). The combination of total 16S copy number quantification with the relative percentage abundance composition determined by 16S sequencing allows researchers to obtain the cell count of each microbe in a sample.

16S rRNA gene sequencing continues to be one of the leading methods for microbial community profiling. Although protocols have been well-established, it is important for researchers to consider the different aspects of the library prep workflow and how they can affect data output and accuracy. Additionally, the ease-of-use of a workflow should not be overlooked, as complicated protocols can lead to more human errors. After rigorous testing and optimization, the *Quick*-16S™ Plus NGS Library Prep Kit (V3-V4) simplifies the 16S library prep workflow and generates accurate and robust libraries with minimal PCR chimera. With only 30 minutes of hands-on time for 96 samples, the Quick-16S Plus NGS Library Prep Kit (V3-V4) features the most streamlined protocol for 16S rRNA gene sequencing library preparation.

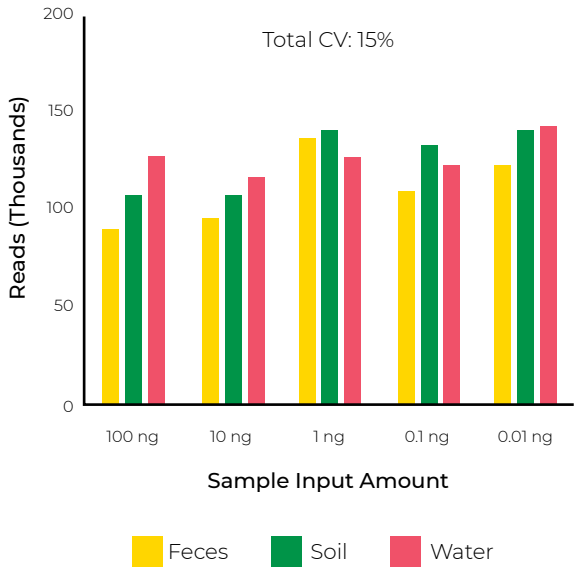


Figure 4. Auto-normalization across a wide range of DNA inputs by Equalase. 0.01-100 ng fecal, soil, and water DNA were used as inputs. Libraries were pooled by equal volumes (2 µl each) without further normalization and sequenced using MiSeq® Reagent Kit v3 (600-cycle). The CV (coefficient of variation) is the ratio of the standard deviation to the mean with lower values corresponding to less dispersion around the

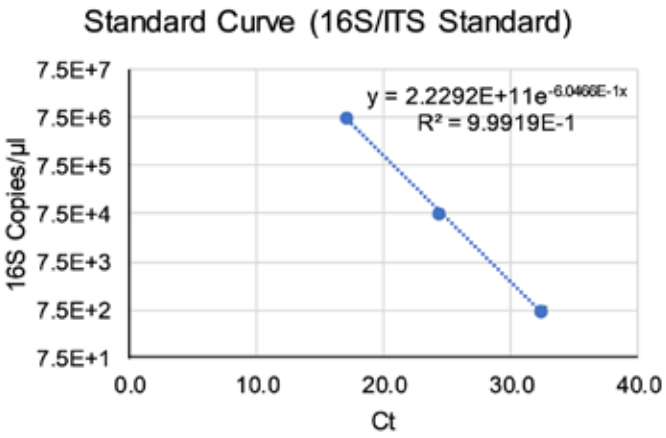


Figure 5. qPCR quantification embedded in 16S library preparation enables total 16S copy number quantification. Serial dilutions of a standard with quantified 16S rRNA gene copy numbers were used to generate a standard curve. The linear equation can then be used to calculated total 16S rRNA copy numbers of each sample based on Ct values.

References

1. Klindworth, A., E. Pruesse, T. Schweer, J. Peplies, C. Quast, M. Horn, and F. O. Glöckner. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic acids research 41:e1.
2. Thompson, L. R., J. G. Sanders, D. McDonald, A. Amir, J. Ladau, K. J. Locey, R. J. Prill, A. Tripathi, S. M. Gibbons, G. Ackermann, J. A. Navas-Molina, S. Janssen, E. Kopylova, Y. Vázquez-Baeza, A. González, J. T. Morton, S. Mirarab, Z. Zech Xu, L. Jiang, M. F. Haroon, J. Kanbar, Q. Zhu, S. Jin Song, T. Kosciolek, N. A. Bokulich, J. Lefler, C. J. Brislawn, G. Humphrey, S. M. Owens, J. Hampton-Marcell, D. Berg-Lyons, V. McKenzie, N. Fierer, J. A. Fuhrman, A. Clauset, R. L. Stevens, A. Shade, K. S. Pollard, K. D. Goodwin, J. K. Jansson, J. A. Gilbert, and R. Knight. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551:457-463.
3. Parada, A. E., D. M. Needham, and J. A. Fuhrman. 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environmental microbiology 18:1403-1414.
4. Guo, J., J. R. Cole, Q. Zhang, C. T. Brown, and J. M. Tiedje. 2016. Microbial Community Analysis with Ribosomal Gene Fragments from Shotgun Metagenomes. Applied and environmental microbiology 82:157-166.
5. Johnson J. S., Spakowicz D. J., Hong B., Petersen L. M., Demkowicz P., Chen L., Leopold S. R., Hanson B. M., Agresta H. O., Gerstein M., Sodergren E., Weinstock M. 2019 Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nature Communications 10:
6. Kuczynski, J., C. L. Lauber, W. A. Walters, L. W. Parfrey, J. C. Clemente, D. Gevers, and R. Knight. 2011. Experimental and analytical tools for studying the human microbiome. Nature reviews Genetics 13:47-58.
7. Illumina 16S protocol: https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf