PacBio

Application note

Kinnex 16S rRNA kit for full-length 16S sequencing

Introduction

Microbes are an essential part of the ecosystem for human, plant, and animal species and the environments they live in. Microbes perform metabolic activities, produce and degrade compounds, and play a role in health, fitness, phenotype, and ecology. The human microbiome has been shown to be pivotal for human health, with dysbiosis in the gut microbiome having been linked to conditions such as inflammatory bowel disease (IBD), diabetes, cardiovascular disease, colon cancer, and neurological disease. Similarly, both soil and marine microbes play an active role in organism health. These relationships demonstrate a growing need and appreciation for more comprehensively characterizing the species within microbiomes and associating them with biological outcomes.

All bacteria have a 16S rRNA gene, making targeted 16S sequencing a reliable and cost-effective approach for assessing the composition of metagenomic communities. This is especially true for low bacterial biomass samples where amplicon sequencing is the best approach. However, the high similarity in the 16S rRNA genes among related bacteria mean that sequencing the entirety of the 16S gene (~1.5 kb) with high accuracy is essential for characterizing at the species or strain level.

Recent comparative studies have shown that PacBio® full-length 16S sequencing outperforms other sequencing methods (Notario et al., 2023, Figure 1). The *Kinnex™ 16S rRNA kit* takes amplified 16S gDNA amplicons as input and outputs a sequencing-ready library that results in a 12-fold throughput increase compared to other 16S libraries. Allowing up to 1,536-plex per library, Kinnex technology enables highly accurate, cost-effective 16S sequencing for microbiome studies.

102-326-601 REV01 DEC2023

PacBio

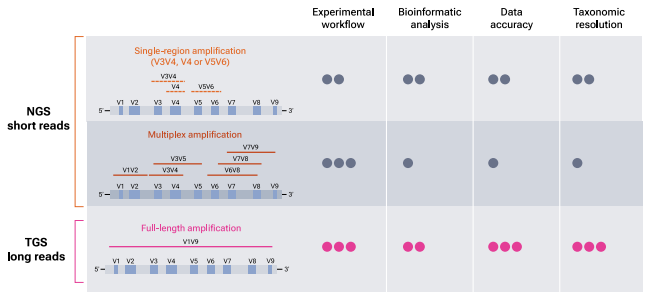

Figure 1. Full-length 16S rRNA sequencing with PacBio technology covers the entire 1.5 kb 16S gene with high accuracy, allowing for species- and strain-level identification in microbiome studies (adapted from Notario et al. 2023).

Full-length 16S rRNA gene sequencing with the *Kinnex 16S rRNA kit*

The *Kinnex 16S rRNA kit* uses the MAS-Seq method to increase throughput on PacBio long-read sequencers. MAS-Seq is a concatenation method for joining cDNA molecules into longer fragments (Al'Khafaji et al., 2023). HiFi reads generated from sequencing the concatenated molecules can then be broken up bioinformatically to retrieve the original amplicon sequences (Figure 2). The result is higher throughput and reduced sequencing needs for cost-effective 16S rRNA sequencing.

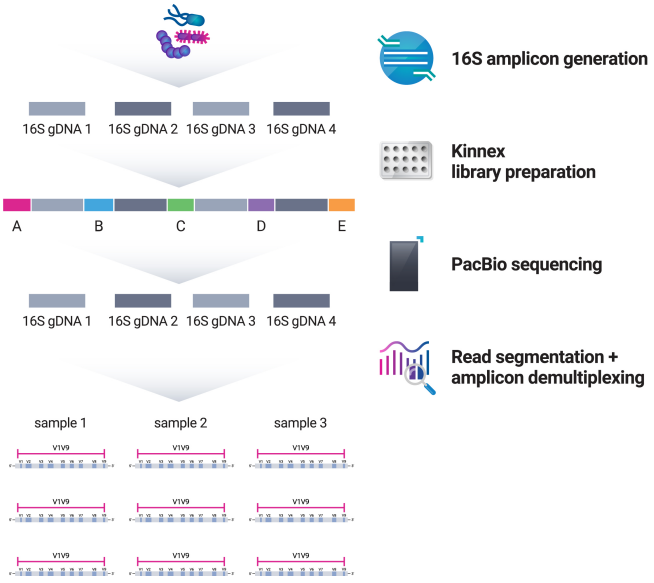

Figure 2. Kinnex 16S rRNA sequencing. Barcoded full-length 16S gDNA amplicons are concatenated into a large-insert library, sequenced, then processed using PacBio software.

Kinnex 16S rRNA library workflow

The Kinnex 16S rRNA workflow (Figure 3) begins with amplified 16S genes and produces a sequencing-ready library.

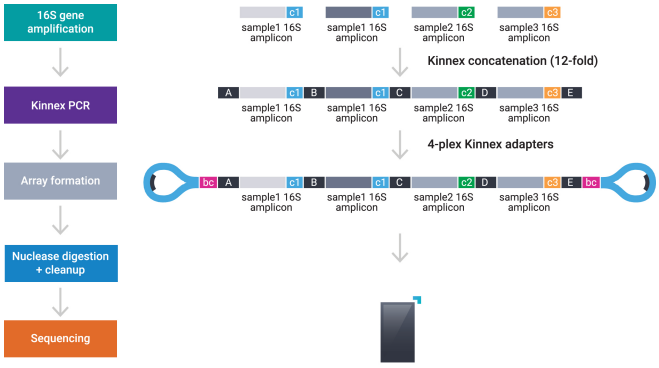

Figure 3. Kinnex 16S rRNA library workflow.

Full-length 16S gDNA is amplified with barcoded 16S gene-specific forward and reverse primers (based on the described protocol) compatible with Kinnex array formation. Using 12 forward primers, 32 reverse primers and the four Kinnex adapters allows up to 384-plex at the gDNA amplicon level and up to 1,536-plex when combined with all four barcoded adapters.

With proper full array formation and adequate sequencing, one SMRT® Cell on the Sequel® II, IIe, and Revio™ systems are expected to achieve 20–25 million and 50–60 million 16S sequences, respectively (Table 1).

Metric	Performance
Sample preparation time	2 days
Expected library size	18,000 bp
Target P1 loading	60–75%
Expected HiFi yield	1.5–2.5 M HiFi reads (Sequel II and IIe systems) 5–6 M HiFi reads (Revio system)
Expected full array %	80–90%
Expected read yield	20–25 M reads (Sequel II and IIe systems) 50–60 M reads (Revio system)

Table 1. Target Kinnex 16S rRNA library performance.

Kinnex 16S dataset release

PacBio has released public Kinnex 16S datasets consisting of several reference materials from Zymo Research and ATCC (American Type Culture Collection). We show Kinnex concatenation results in considerably more reads – allowing for higher multiplexing – while retaining the HiFi accuracy for each 16S read for species identification.

Figure 4 shows the throughput increase using the *Kinnex 16S kit* on Sequel II, IIe, and Revio systems. Without Kinnex, a regular full-length 16S library on the Sequel II and IIe systems achieve 2.1–3.3 million reads, compared to 19.6–28.4 million reads with Kinnex. On the Revio system the margin ranges from 8.5–10.1 million reads without Kinnex to 62.5–72.2 million reads with Kinnex. This difference allows for a higher read depth per sample. For example, a 384-plexity with Kinnex on the Revio system results in more reads than with 96-plexity without Kinnex on both Sequel II, IIe, and Revio systems (Figure 5).

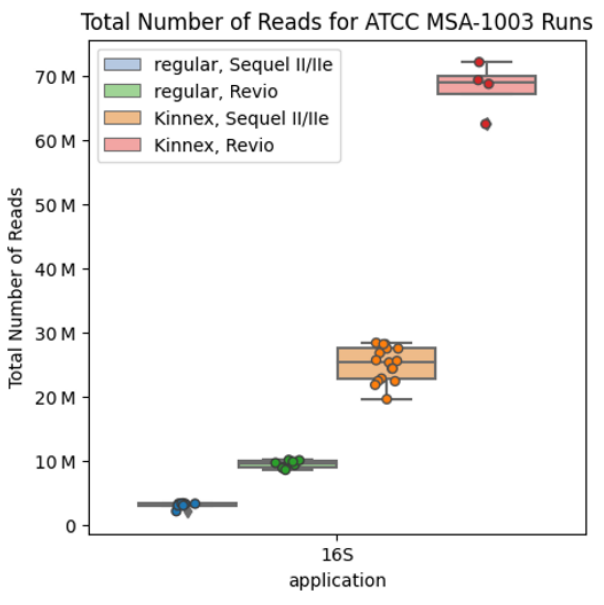

Figure 4. Kinnex concatenation results in 8- to 12-fold throughput increase for full-length 16S sequencing on Sequel II/IIe and Revio systems. Each datapoint in the boxplot represents a regular or Kinnex 16S library run on one SMRT Cell. The number of reads are the de-concatenated S-reads representing the original full-length 16S amplicon.

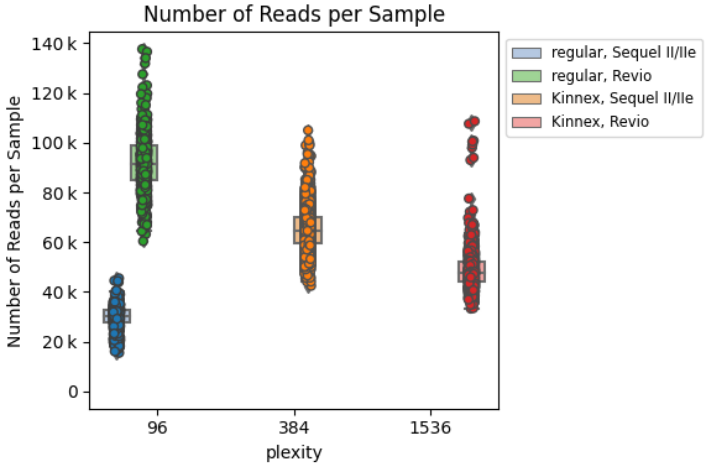

Figure 5. Increased read depth per sample as a result of Kinnex concatenation compared at different plexity.

Increased read depth also led to more identified species with better abundance concordance. Using a down-sampled Kinnex 16S dataset from the complex human fecal reference sample (Zymo D6323), we show that more input reads into the 16S pipeline, HiFi-16S-workflow (using --filterQ=30), results in more identified species (Figure 6) with greater fraction of reads assigned at the species level (Figure 7). On a 20-species microbial standard (ATCC MSA-1003), the Kinnex 16S library correlated better with the expected species representation than the regular 16S library due to higher read depth coverage (Figure 8).

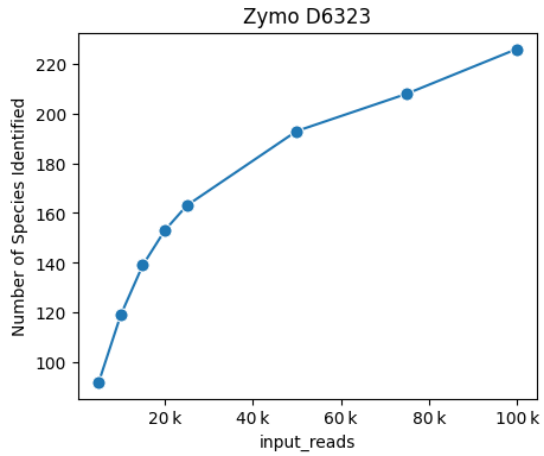

Figure 6. Increased read depth leads to more identified species using Kinnex 16S data based on down-sampled Kinnex 16S dataset. X-axis: Downsampled input S-reads with Q30 filter option in HiFi-16S-workflow. Y-axis: Number of species identified.

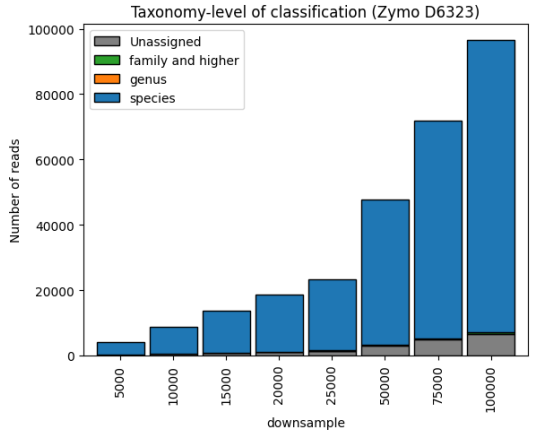

Figure 7. Full-length 16S sequencing with Kinnex allows for species identification across different read depths. X-axis: Downsampled input S-reads with Q30 filter option in HiFi-16S-workflow. Y-axis: Number of reads assigned to different taxonomic levels.

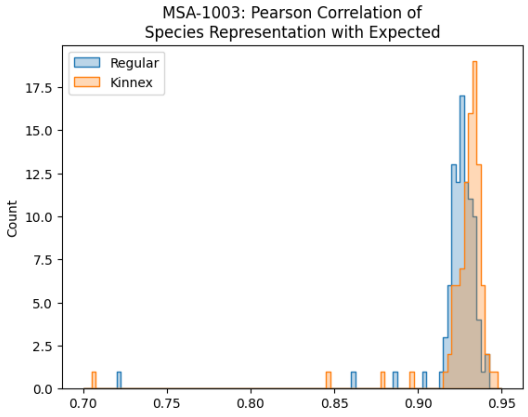

Figure 8. Kinnex allows higher multiplexing and captures better species representation in reference 16S materials. Kinnex: 96-plex of the same MSA-1003 (ATCC) sample, Regular: 96-plex of MSA-1003. Pearson correlation is calculated based on the fraction of reads assigned to species against the expected fraction.

Conclusion

Full-length 16S sequencing is imperative when species and/or strain characterization is needed, especially from low bacterial biomass samples where amplicon sequencing is the best or only option. The *Kinnex 16S rRNA kit* increases throughput by 12-fold using the MAS-Seq concatenation technology. We show that more read depth allows for higher multiplexing capacity (up to 1,536-plex) and allows more identified species per sample. Coupled with flexible multiplexing strategies and a SMRT® Link bioinformatics workflow, users can now achieve cost-effective full-length 16S rRNA gene sequencing for species- and strain-level identification for microbiome studies.

Bioinformatics recommendations for 16S analysis

Several full-length 16S pipelines such as the HiFi-16S-workflow can be used to analyze Kinnex 16S datasets after de-concatenation using Read Segmentation and per-sample demultiplexing. Note however that we recommend using --filterQ=30 to filter at the read level and to downsample to 100k S-reads (if exceeding this number of S-reads for low plexity samples) for optimal results.

Resources

Procedure & checklist – Preparing Kinnex libraries from 16S rRNA amplicons

HiFi-16S-workflow for HiFi full-length 16S analysis

Kinnex 16S rRNA datasets

Kinnex technology landing page

References

Al'Khafaji, A. M., et al. (2023). High-throughput RNA isoform sequencing using programmed cDNA concatenation. *Nature Biotechnology*, 1-5. https://doi.org/10.1038/s41587-023-01815-7

Notario, E., et al. (2023). Amplicon-based microbiome profiling: from second- to third-generation sequencing for higher taxonomic resolution. *Genes*, 14(8), 1567. https://doi.org/10.3390/genes14081567