High-resolution microbiome species profiling at scale with the Kinnex kit for full-length 16S rRNA sequencing

Jeremy E Wilkinson[1], Jocelyne Bruand[1], Khi Pin Chua[1], Heather Ferrao[1], Davy Lee[1], Jeff Zhou[1], Kristopher Locken[2], Shuiquan Tang[2], Ethan Thai[2], John Sherman[2], Brett Farthing[2], Elizabeth Tseng[1]

[1]PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025, USA [2]Zymo Research Corporation, 17062 Murphy Avenue, Irvine, CA 92614, USA

jwilkinson@pacb.com
@jewilki

Full-length 16S sequencing yields better taxonomic resolution than short-read 16S

Targeted 16S sequencing is a cost-effective approach for assessing the bacterial composition of microbial communities. This is especially true for low bacterial biomass samples where amplicon sequencing is the best option. However, the high similarity between the 16S rRNA genes of related bacteria means that sequencing the entirety of the 16S gene (~1.5 kb) with high accuracy is essential for species- or strain-level characterization. Many recent comparative studies have shown that PacBio full-length (FL) 16S sequencing outperforms other sequencing methods for taxonomic resolution and data accuracy (Fig 1, Fig 2).

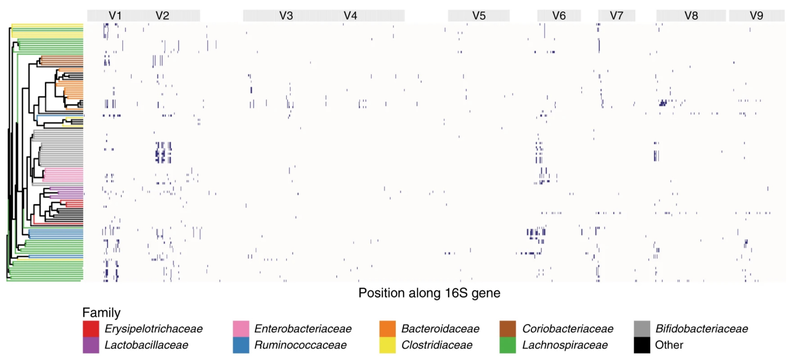

Figure 1. Intragenomic 16S gene polymorphisms in human gut microbiome isolates. Location of SNPs present in the 16S genes of individually cultured bacterial isolates. SNP locations were identified through full-length 16S gene sequences generated for each individual isolate. X-axis denotes position along the 16S gene. Y-axis denotes individual isolates clustered based on their inferred phylogeny. Dark blue region indicates the location of a polymorphism. (Figure 4a from Johnson et al., 2019).

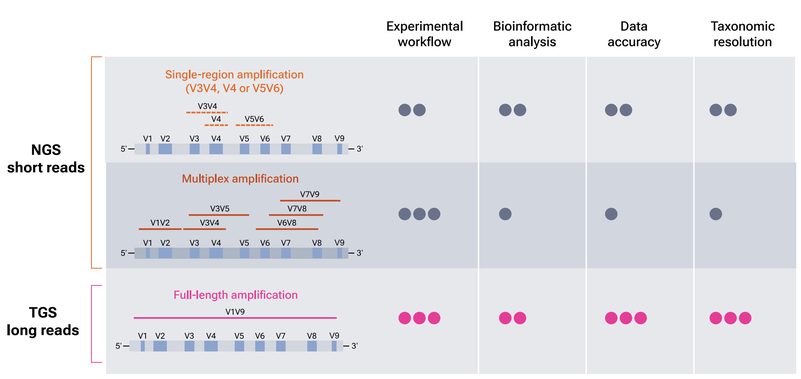

Figure 2. Comparison results of three different methods for 16S rRNA amplicon sequencing. Full-length 16S sequencing with PacBio technology covers the entire 1.5 kb 16S gene with high accuracy, allowing for species- and strain-level identification in microbiome studies and outcompetes the two short-read methods: single-region or 2-region partial 16S and multiplex amplification (adapted from Notario et al., 2023).

References

- Johnson, J. S., et al. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature communications*, 10(1), 5029.
- Notario, E., et al. (2023). Amplicon-Based Microbiome Profiling: From Second- to Third-Generation Sequencing for Higher Taxonomic Resolution. *Genes*, 14(8), 1567.
- Al'Khafaji, A. M., et al. (2024). High-throughput RNA isoform sequencing using programmed cDNA concatenation. *Nature biotechnology*, 42(4), 582–586.
- HiFi-16S-workflow: https://github.com/PacificBiosciences/HiFi-16S-workflow
- Application note — Kinnex 16S rRNA kit for full-length 16S sequencing: https://www.pacb.com/wp-content/uploads/Application-note-Kinnex-16S-rRNA-kit-for-full-length-16S-sequencing.pdf

The Kinnex 16S rRNA kit, method, and evaluation approach

The Kinnex 16S rRNA kit takes amplified 16S amplicons as input and outputs a sequencing-ready library that results in an up to 12-fold throughput increase compared to standard FL 16S libraries. The Kinnex 16S kit is based on the multiplexed array sequencing (MAS-Seq) method (Al'Khafaji et al., 2024) applied to FL 16S amplicons (Fig 3a). The result is significantly higher throughput and reduced sequencing needs for high accuracy, cost-effective FL 16S sequencing with the ability to multiplex up to 1,536 amplicon samples per PacBio SMRT Cell. We tested the Kinnex 16S rRNA kit on a diverse range of samples including mock community standards, stool, saliva, plant, soil, wastewater, and swabs (skin, oral, vaginal, and veterinary wound). We then analyzed the data using a user-friendly bioinformatics pipeline, HiFi-16S-workflow, that provides a FASTQ-to-report analysis solution for FL 16S HiFi reads (Fig 3b). We also examined the effects of read depth on the number of species detected among the sample types (Fig 4).

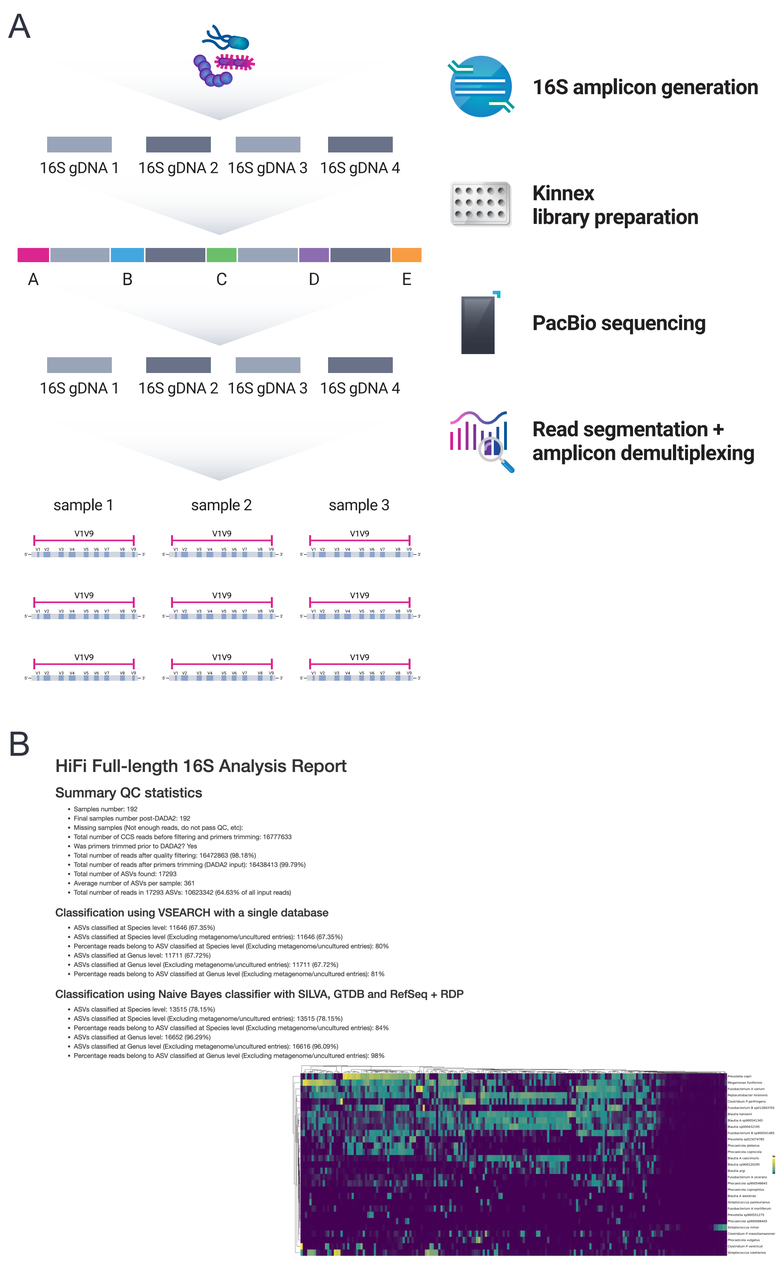

Figure 3. Kinnex 16S rRNA sequencing workflow. (A) A pooled library of barcoded full-length 16S amplicons from samples are concatenated into a large-insert library with the Kinnex 16S rRNA kit that uses 12 concatemers, sequenced on a PacBio HiFi long-read system, then processed using PacBio SMRT Link software. **(B)** The HiFi-16S-workflow is then used to analyze FL 16S reads to generate amplicon sequence variants and estimate the samples' species relative abundances.

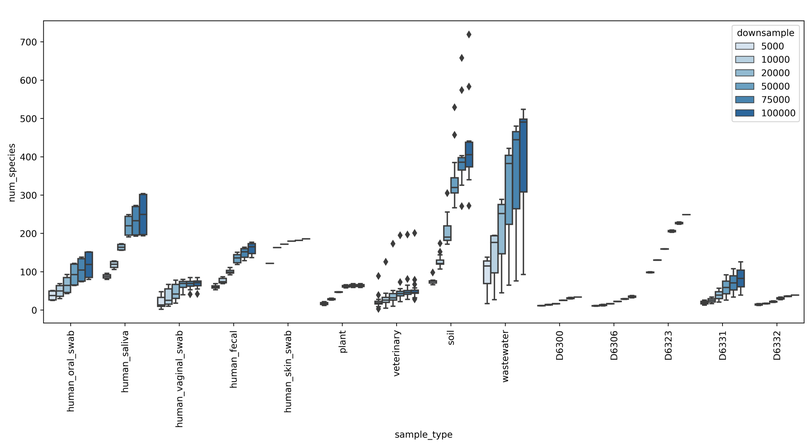

Figure 4. Effect of read depth on number of species detected among different sample types. Box plot of number of species detected among the different types of samples tested. Some sample types had multiple samples tested. Sample data was downsampled to see the effect of read depth on number of species observed. Outliers are indicated by diamonds.

Kinnex concatenation results in an up to 12-fold throughput increase for full-length 16S

Results show that Kinnex 16S sequencing can yield >30k average reads per sample on a single SMRT Cell at a 1,536-plex on the Revio system, at a 1,152-plex on the Vega system, or at a 768-plex on the Sequel IIe system. Without Kinnex, a regular full-length 16S library on the Sequel II systems achieves 2.1–3.3 million (M) reads, compared to 19.6–28.4 M reads with Kinnex, on the Revio system from 8.5–10.1 M reads without Kinnex to 62.5–94.4 M reads with Kinnex, and 41.1 M reads with Kinnex 16S on the Vega benchtop system (Fig 5a, Fig 5b). This difference with Kinnex 16S allows for more samples multiplexed per SMRT Cell and higher read depths per sample. On a 20-species microbial standard, the Kinnex 16S library correlated better with the expected species representation than the regular 16S library due to higher read depth (Fig 5c).

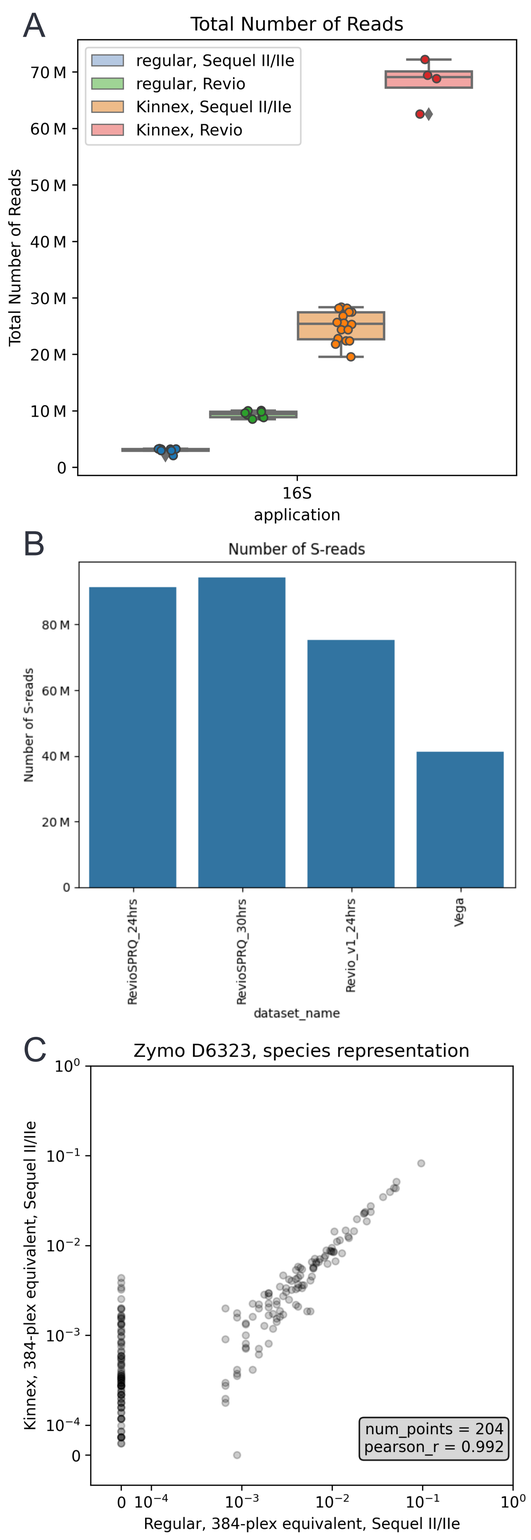

Figure 5. (A) Kinnex 16S results in an up to 12-fold throughput increase for full-length 16S sequencing on Sequel II/IIe and Revio systems. Each datapoint in the box plot represents a regular or Kinnex 16S library run on one SMRT Cell. The number of reads are the de-concatenated S-reads representing the original full-length 16S amplicon. Outliers are indicated by diamonds. **(B) Kinnex 16S is compatible with all PacBio HiFi long-read systems.** Number of deconcatenated HiFi reads (Segmented reads [S-reads]) from Kinnex 16S libraries run on one SMRT Cell on the Revio system with SPRQ chemistry for 24 and 30 hours, Revio system without SPRQ chemistry for 24 hours, and Vega benchtop system for 24 hours. **(C) Kinnex 16S is highly correlated to standard FL 16S for species composition.** 384-plex of the Zymo fecal reference sample (D6323) on both Kinnex 16S and regular FL 16S on the Sequel IIe system. Pearson correlation is calculated on the species relative abundances between the two datasets.

Summary

- **Kinnex 16S enables significant throughput increases**, up to 12-fold, for FL 16S HiFi sequencing on any PacBio HiFi long-read system.
- Due to the higher number of reads per sample, **Kinnex 16S allows for more recovery of lower abundance species.**
- Researchers may now **multiplex more samples to dramatically reduce cost** per sample or to **profile each sample deeper with more reads per sample**, with the Kinnex 16S rRNA kit.
- The additional reads per sample along with better taxonomic resolution is advantageous for numerous environmental sample types which are often highly diverse, containing many microbial species.