Increasing the throughput of full-length 16S sequencing with Kinnex kits

Jeremy E Wilkinson,[1] Khi Pin Chua,[1] Siyuan Zhang,[1] Jason Underwood,[1] Minning Chin,[1] Wei-Shen Cheng,[1] Sian Loong Au,[1] Primo Baybayan,[1] Holly Ganz,[2] Guillaume Jospin,[2] Ye Tao,[3] Qin Lin,[3] Elizabeth Tseng[1]

1. PacBio, 1305 O'Brien Drive, Menlo Park, CA, USA 94025 2. AnimalBiome, 400 29th St., Ste 101, Oakland, CA, USA 94609 3. Biozeron Biotechnology Co., Ltd., Shanghai, 201800 China

jwilkinson@pacb.com
@jewilki

Background

In the past several years, the ability to capture the full-length (FL) 16S rRNA gene with PacBio HiFi sequencing has enabled researchers to profile microbiomes in significantly higher resolution. Only full-length and highly accurate 16S sequences can robustly identify the broad range of bacteria seen in complex microbial communities at the species level, without bias.

To further increase the cost effectiveness of FL 16S sequencing, we applied the Kinnex 16S rRNA kit, which is based on the multiplexed array sequencing (MAS-Seq) method (Al'Khafaji et al., 2023), to FL 16S amplicons. The MAS-Seq method is a versatile throughput increase method that takes advantage of the longer HiFi read lengths to concatenate amplicons into ordered arrays with programmable array sizes. We demonstrated that Kinnex 16S results in an ~8–12-fold throughput increase compared to standard FL 16S. We tested the method on a diverse range (11 types) of samples including mock communities, human and animal feces/guts, soil, sediment, rhizosphere, sludge, and water.

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 (figure 1).

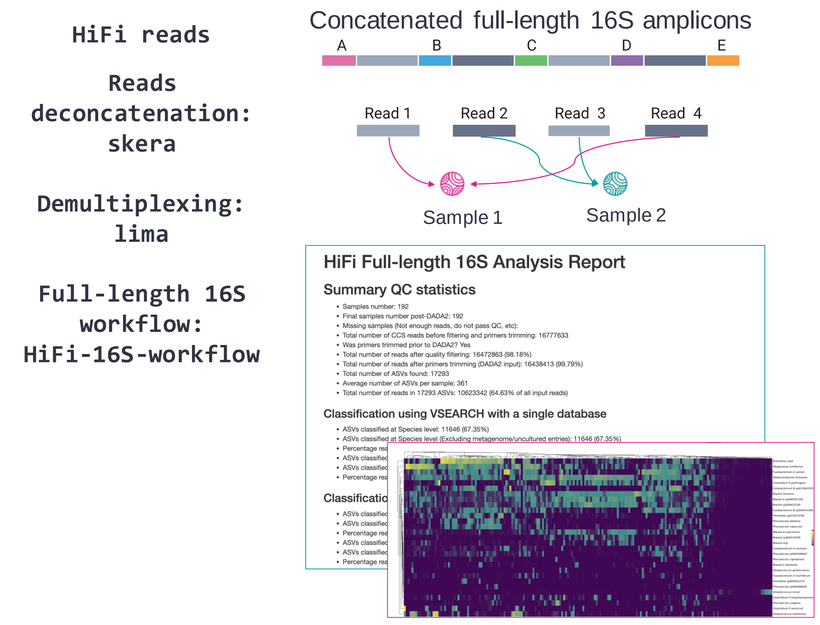

Figure 1. Workflow of Kinnex 16S sequencing. Full-length 16S amplicons are concatenated using the Kinnex 16S kit with an array size of 12. The reads are then deconcatenated into individual FL 16S reads (S-reads) followed by demultiplexing using samples' barcodes. Finally, the HiFi-16S-workflow is used to process and analyze the reads to generate amplicon sequence variants (ASV) and estimate the relative abundances of different species in the samples.

Comparing the Kinnex 16S to standard FL 16S datasets, we found no bias in community compositions and were able to assign up to ~90–99% of denoised reads to species. In addition, on the highly complex ZymoBIOMICS *Fecal Reference with TruMatrix Technology* (D6323) sample, we found Kinnex 16S to have high correlation to taxonomic abundances estimated from shotgun metagenomics sequencing using the same sample, emphasizing that it's possible to get shotgun metagenome taxonomic resolution at amplicon sequencing costs with FL 16S HiFi sequencing. Furthermore, with Kinnex 16S, researchers may now multiplex more samples to reduce cost per sample or to profile each sample deeper with more reads per sample.

SCAN ME

Application note — Kinnex 16S rRNA kit for full-length 16S sequencing

Kinnex 16S sequencing yields more HiFi reads

Using the PacBio Sequel IIe system, we collaborated with AnimalBiome and Biozeron Biotechnology to generate between 15 million and 34 million HiFi reads with one SMRT Cell 8M across five sequencing runs. The datasets include sample types such as human and animal feces/guts, food, soil, sediment, rhizosphere, sludge, seawater, and others. Applying the HiFi-16S-workflow to these datasets, we observed >80% reads assigned to species level for human and insect guts, fermentation, sediment, cat and dog fecal and oral sample types. All five sequencing runs achieved >15 M HiFi reads after deconcatenation and provided significant reads/sample at high multiplexing level (384 samples). The first 3 experiments (AnimalBiome and Biozeron) were early tests. The last two experiments (Zymo D6323 and mock communities) were after more optimization, resulting in a substantial increase in the number of HiFi reads (table 1).

Experiment	# samples	HiFi reads	Mean QV	S-reads*	Reads/sample at a 384-plex
AnimalBiome	192	1.62 M	Q35	17.15 M	**43,692**
Biozeron run 1	384	1.50 M	Q35	17.24 M	**44,883**
Biozeron run 2	384	1.45 M	Q33	15.20 M	**39,573**
Zymo D6323	1	2.55 M	Q33	30.47 M	**79,349**
Zymo and ATCC mocks	32	2.83 M	Q33	33.77 M	**87,943**

Table 1. Sequencing metrics for Kinnex 16S. HiFi sequencing data was generated for all five sequencing runs including many different sample types.
*S-reads are deconcatenated HiFi reads.

Using samples from AnimalBiome, we compared Kinnex 16S sequencing data to standard full-length 16S sequencing data and found >97% correlation in Shannon diversity between the two sequencing methods (fig. 2A). In addition, we compared relative abundance of species estimated based on shotgun metagenomics sequencing to that of Kinnex 16S sequencing and found high correlation between the two methods (fig. 2B).

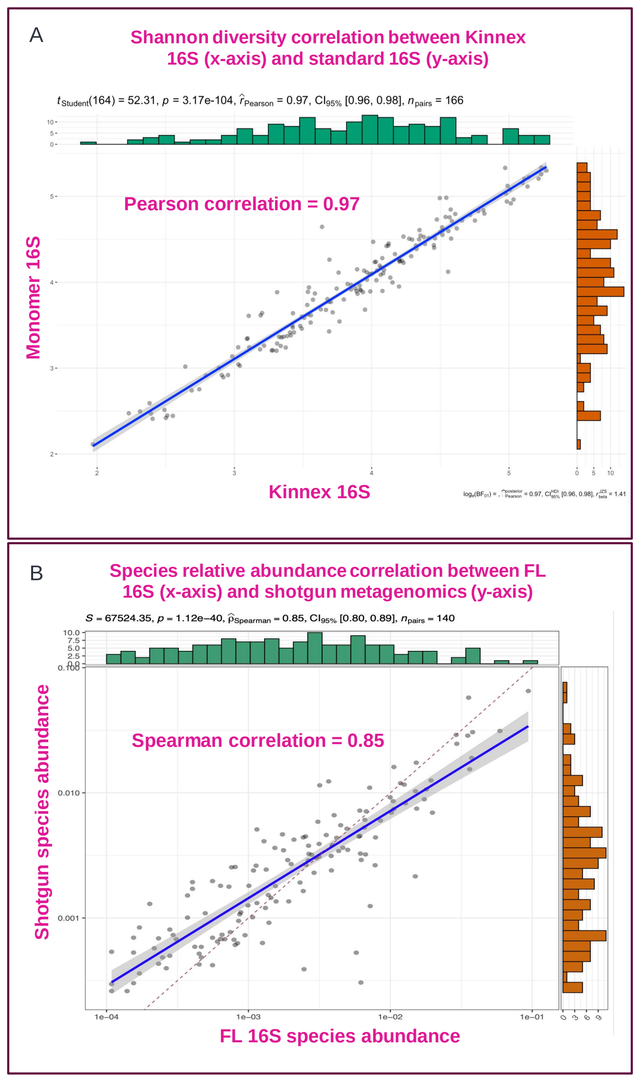

Figure 2. Checking for biases in full-length 16S data. (A) For each sample in the AnimalBiome dataset, Shannon diversity was calculated using data from Kinnex 16S and monomer (standard) FL 16S sequencing. Kinnex 16S samples' depths were downsampled to similar depth as standard FL 16S. Figure shows that Kinnex 16S provides diversity estimates close to standard FL 16S sequencing. (B) Figure shows high correlation of relative abundance estimates at species level between Kinnex 16S and shotgun metagenomics sequencing method using the Zymo D6323 sample.

The HiFi-16S-workflow

We developed an ASV-based analysis workflow written in Nextflow to analyze full-length 16S HiFi sequencing data, the HiFi-16S-workflow. The software used in the pipeline are containerized for ease of installation and include well-established tools such as Qiime 2 and DADA2. The HiFi-16S-workflow utilizes Nextflow to scale analysis of large number of samples on a high-performance computing cluster and can be easily adopted to the cloud. This workflow implements a simple VSEARCH-based (GTDB r207) and an exhaustive naïve-Bayes classification method (multiple databases, see GitHub FAQ for more details). We ran the HiFi-16S-workflow on Kinnex 16S data across different sample types and found a high percentage of reads assigned at the species level. The lower percentage of reads assigned at species level for complex environmental samples highlights the need for sample-type-specific databases which may improve taxonomy resolution (e.g., MiDAS database for activated sludge increased percentage of reads classified at species level from 38 to 89% using VSEARCH) (fig 3).

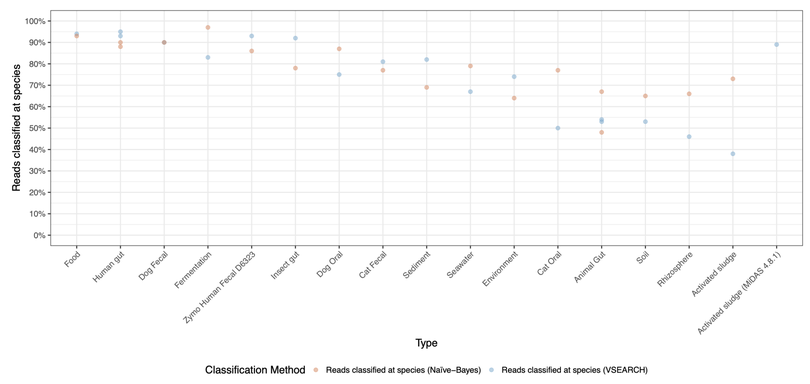

Figure 3. Percentage of reads classified at species level. After running the HiFi-16S-workflow, we calculated the percentage of denoised reads that belonged to species-level ASVs. For activated sludge, we also ran VSEARCH taxonomy classification using MiDAS 4.8.1.

Summary

- Kinnex 16S enables significant throughput increases, at ~10-fold, for full-length 16S HiFi sequencing.
- Kinnex 16S sequencing does not show bias compared to standard full-length 16S sequencing.
- Species relative abundances estimated based on Kinnex 16S sequencing are highly concordant with estimates based on shotgun metagenomics sequencing.
- Up to ~90–99% of denoised full-length 16S reads were able to be assigned to species level, dependent on the sample type and reference database.
- With Kinnex 16S, it's possible to multiplex more samples per SMRT cell and/or to profile each sample deeper with more reads per sample.

References / more information

- Al'Khafaji, A.M. et al. (2023). High-throughput RNA isoform sequencing using programmed cDNA concatenation. *Nature Biotechnology*. https://doi.org/10.1038/s41587-023-01815-7
- 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
- HiFi-16S-workflow: https://github.com/PacificBiosciences/HiFi-16S-workflow
- Microbiome and metagenome sequencing with HiFi reads: https://www.pacb.com/wp-content/uploads/Application-Brief-Metagenomic-Sequencing-with-HiFi-Reads-Best-Practices.pdf
- PacBio microbiome and metagenomics webpage: https://www.pacb.com/products-and-services/applications/complex-populations/microbial/