Optimizing Long-Read 16S rRNA Sequencing for Accurate and Cost-Effective Plant Microbiome Profiling

Dalena Ong[2,3,#], Weilun Tan[2,4#], Jane Michelbach[2,4], Ana Almeida[1,2,4]

1. Associate Professor and Chair, Department of Biological Sciences, California State University, East Bay; Associate Director, Green Biome Institute, California State University, Hayward, California, USA
2. Green Biome Institute, California State University, East Bay, Hayward CA. 94542.
3. Pre-Professional Health Academic Program, California State University, East Bay, Hayward, California, USA
4. Bay Area RaMP Program in Microbiome Sciences, University of California, Berkeley, Berkeley, California, USA

#. These authors contributed equally

Introduction

Plant-associated microbiomes play a critical role in plant health, productivity, and ecosystem function by regulating nutrient cycling, stress responses, and host–microbe interactions. Precise characterization of these microbial communities is therefore essential for advancing our understanding of plant biology and for informing sustainable agricultural and conservation practices.

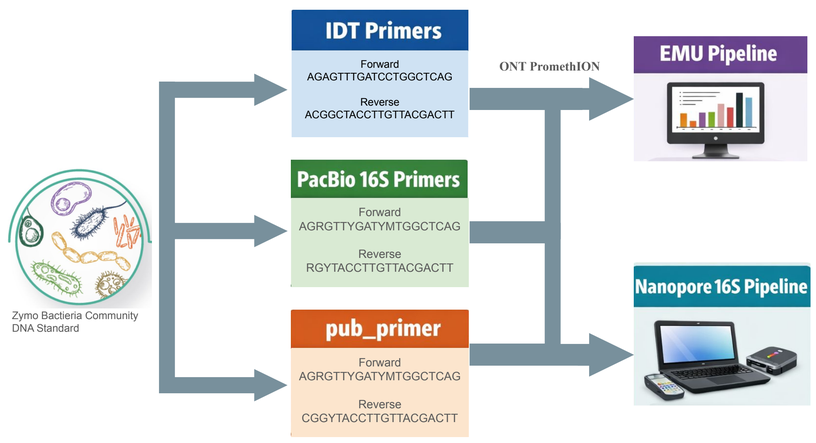
Despite significant methodological progress, accurate profiling of plant microbiomes remains technically challenging. One of the primary obstacles is host-derived DNA contamination, particularly from chloroplast and mitochondrial 16S rRNA sequences, which are frequently co-amplified by universal bacterial primers and can dominate amplicon libraries (1). This off-target amplification substantially reduces effective sequencing depth and obscures bacterial community signals. In addition, primer bias and the inherent limitations of short-read sequencing can further distort estimates of microbial diversity and relative abundance, complicating biological interpretation and comparisons across studies.

Long-read sequencing platforms such as Oxford Nanopore Technologies (ONT) enable recovery of full-length 16S rRNA gene sequences, offering improved taxonomic resolution and reduced ambiguity relative to short-amplicon approaches (2). However, the effective application of ONT sequencing to plant microbiome studies requires careful optimization of experimental and computational parameters, including primer selection, bioinformatic pipelines, error-correction strategies, and host DNA suppression methods such as chloroplast blockers, all of which strongly influence sequencing accuracy, throughput, and cost-efficiency.

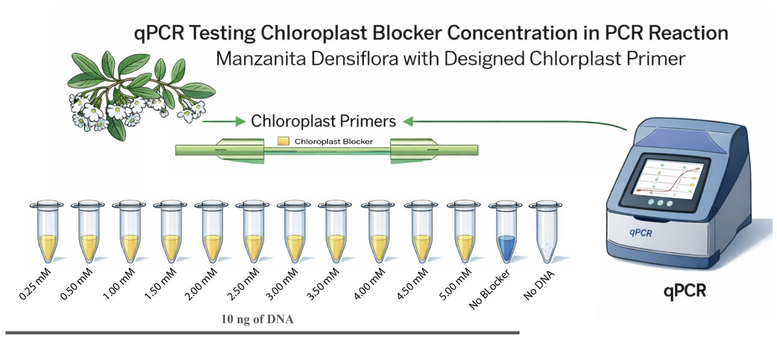
In this study, we systematically evaluate and optimize an ONT-based full-length 16S rRNA sequencing workflow for plant microbiome characterization. By benchmarking bioinformatic pipelines (EMU (4) vs. ONT pipeline (5)) using a defined mock community, testing multiple primer sets, implementing a UMI-based error-correction approach (3), and empirically optimizing chloroplast blocker (pCNA) concentrations using qPCR, we aim to establish a high-throughput, cost-efficient, and accurate protocol suitable for large-scale plant microbiome studies.

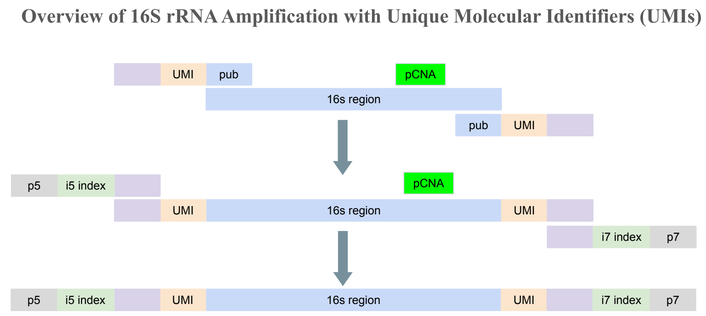
Methodology

Section 1: Comparing Primers and Bioinformatic Pipelines



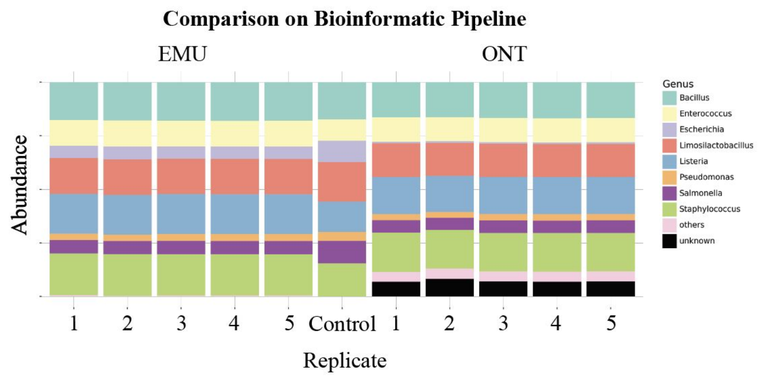
Section 2: Testing Chloroplast Blocker



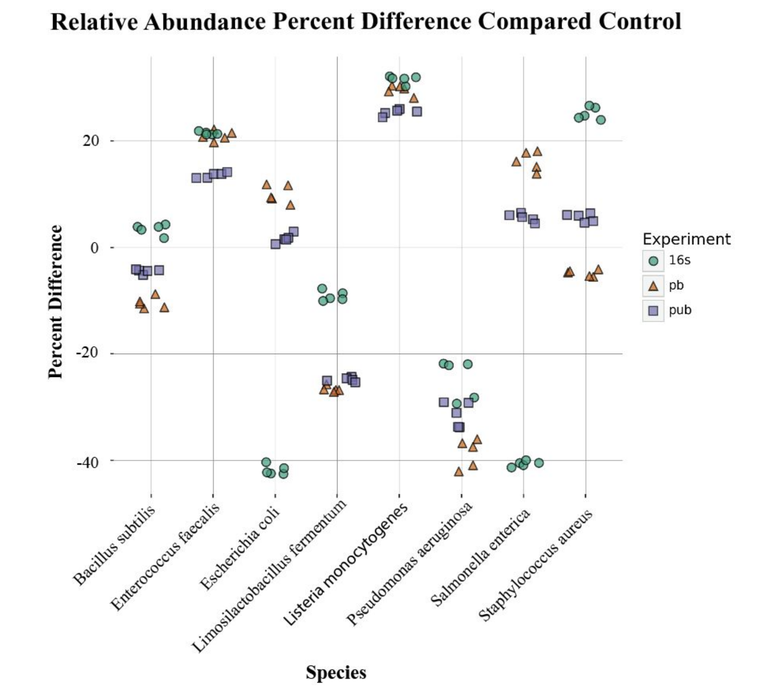


Results

Section 1: Comparing Primers and Bioinformatic Pipelines

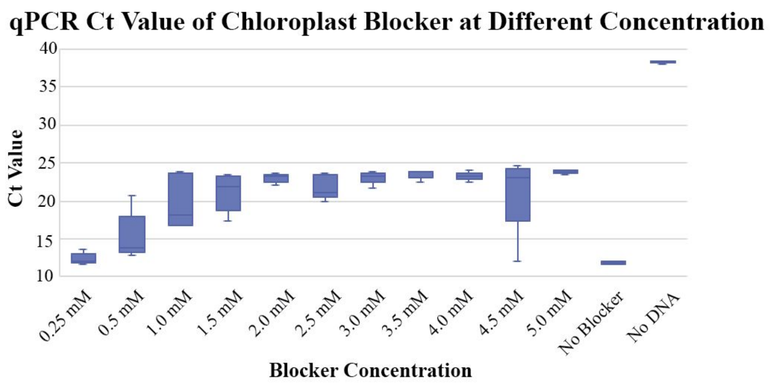


Compares genus-level relative abundance profiles generated by two bioinformatic pipelines, EMU and ONT, across five technical replicates and a control. Each bar represents the proportional composition of bacterial genera within a replicate, allowing direct visual comparison of taxonomic consistency, variation among replicates, and differences in genus assignment between the two pipelines, including the presence of "others" and "unknown" categories.

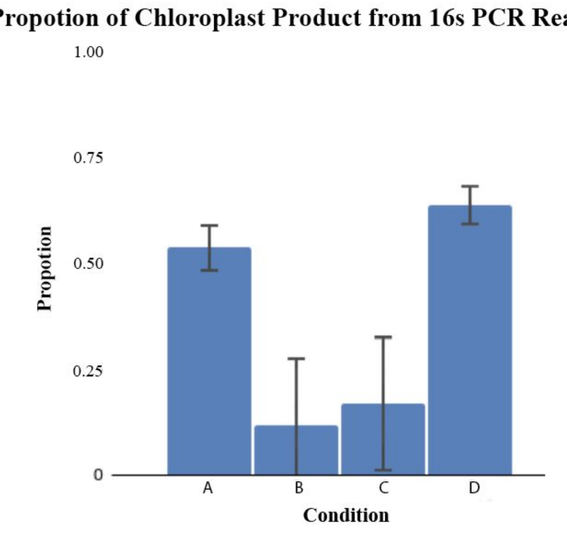


The percent difference in species-level relative abundance compared to the control across three primer sets (16S, pb, and pub). Each point represents an individual replicate, highlighting primer-dependent biases and variability in abundance estimates for each bacterial species relative to the control sample.

Section 2: Testing Chloroplast Blocker



qPCR Ct values across a gradient of chloroplast blocker concentrations, including no-blocker and no-DNA controls. Increasing blocker concentration generally results in higher Ct values, indicating effective suppression of chloroplast amplification, while controls validate assay specificity and background amplification levels.



the proportion of chloroplast-derived amplicons from 16S PCR reactions under four experimental conditions. A: adding chloroplast blocker at 1st PCR reaction; B: adding chloroplast blocker at 2nd PCR reaction; C: adding chloroplast blocker to both reaction; D: no chloroplast blocker added. Error bars indicate variability across replicates, highlighting differences in chloroplast amplification efficiency among conditions and the effectiveness of strategies to reduce host chloroplast contamination.

Conclusion

In this study, we developed and validated an optimized long-read 16S rRNA sequencing workflow for plant microbiome profiling that integrates primer evaluation, bioinformatic benchmarking, UMI-based error correction, and chloroplast blocker optimization. Using a defined mock community, we demonstrate that full-length 16S sequencing on the Oxford Nanopore platform can achieve improved taxonomic resolution while maintaining consistent genus- and species-level abundance estimates across pipelines and primer sets by using the EMU pipeline.

We further show that implementing a chloroplast blocker step is critical for plant-associated microbiome studies, as host-derived chloroplast DNA can otherwise dominate amplicon libraries, obscuring bacterial community signals. qPCR-based optimization revealed a concentration-dependent suppression of chloroplast amplification, highlighting the need for empirical calibration to balance effective host DNA reduction with preservation of bacterial amplification. Our results also indicate that the optimization of the timing of chloroplast blocker application can influence blocking efficiency and downstream community composition.

Taken together, these findings demonstrate that accurate and cost-effective plant microbiome profiling using long-read 16S sequencing requires an integrated experimental and computational framework. By combining optimized chloroplast blocker concentration, UMI-based error correction, and carefully selected primers and analysis pipelines, the workflow presented here enables high-throughput, high-resolution characterization of plant-associated microbial communities. This approach provides a robust foundation for future studies in plant health, agriculture, and conservation biology, where precise microbial profiling is essential for understanding host–microbe interactions and ecosystem function.

Citation

1. Hanshew, A. S., Mason, C. J., Raffa, K. F., & Currie, C. R. (2013). Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities. Journal of microbiological methods, 95(2), 149–155. https://doi.org/10.1016/j.mimet.2013.08.007

2. Lin, X., Waring, K., Ghezzi, H., Tropini, C., Tyson, J., & Ziels, R. M. (2024). High accuracy meets high throughput for near full-length 16S ribosomal RNA amplicon sequencing on the Nanopore platform. PNAS nexus, 3(10), pgae411. https://doi.org/10.1093/pnasnexus/pgae411

3. Amstler, S., Streiter, G., Pfurtscheller, C., Forer, L., Di Maio, S., Weissensteiner, H., Paulweber, B., Schönherr, S., Kronenberg, F., & Coassin, S. (2024). Nanopore sequencing with unique molecular identifiers enables accurate mutation analysis and haplotyping in the complex lipoprotein(a) KIV-2 VNTR. *Genome medicine*, *16*(1), 117. https://doi.org/10.1186/s13073-024-01391-8

4. Curry, K. D., Wang, Q., Nute, M. G., Tyshaieva, A., Reeves, E., Soriano, S., Wu, Q., Graeber, E., Finzer, P., Mendling, W., Savidge, T., Villapol, S., Dilthey, A., & Treangen, T. J. (2022). Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. Nature methods, 19(7), 845–853. https://doi.org/10.1038/s41592-022-01520-4

5. EPI2ME Labs. (2025). wf-16s. Retrieved from https://github.com/epi2me-labs/wf-16s

Acknowledgements:

The authors would like to thank the UC Berkeley RaMP Program for support to JM and WT (NSF award #2216550); the Green Biome Institute (GBI, CSUEB) for funding; CB for helpful insights and discussions.

Contact:

Dr. Ana Almeida
ana.almeida@csueastbay.edu
Weilun Tan
weilun.tan@csueatbay.edu