Exploring microbial biodiversity through genome analysis using ResolveDNA™ Microbiome

Durga M. Arvapalli[1], Isai Salas-González[1], Jon S. Zawistowski[1], J. Blackinton[1], Gary L. Harton[1], Jason A.A. West[1]

[1]BioSkryb Genomics, Inc., Durham, NC.

ABSTRACT

Genomic surveillance of microbes is crucial to understand the genetic contributions to host/bacteria interactions. In contrast to metagenomic approaches, whole genome sequencing (WGS) of single bacteria is required to fully reveal the diversity and evolutionary trajectory of the microbiome. Here, we utilized the ResolveDNA™ Microbiome kit to generate high quality genome assemblies from two bacterial species. The workflow utilizes Primary Template-directed Amplification (PTA), a novel technology for low-input and single cell genome amplification with unprecedented coverage uniformity. Gram-negative (*E. coli*), gram-positive (*B. subtilis*) single bacterial cells and mixed bacteria (*E. coli* & *B. subtilis*) were sorted using fluorescence-activated cell sorting (FACS). The single sorted bacteria were lysed and amplified using the ResolveDNA™ Microbiome kit with amplification yields between 50-200 ng and average amplicon size of 1000 bp. ResolveDNA™ sequencing libraries were constructed, followed by Illumina MiniSeq sequencing. The reconstructed genome assemblies approximated the empirically estimated genome sizes for each species. Additionally, by employing an unbiased *de novo* phylogenetic approach we quantitatively validated the high-quality of the assemblies and calculated that the assembly completeness was over 95% for each single bacterium sequenced. Collectively, the data provides evidence that minute numbers of bacteria can be separated, amplified and prepared for next generation analysis (NGS) analysis. Having successfully analyzed single bacterial genomes, currently we are expanding these findings to computationally deconvolve multi-species bacterial cocktails and to generate datasets from environmental and clinical microbiome samples. We have demonstrated here that ResolveDNA™ Microbiome platform is robust to reconstruct high quality genomes for gram-negative and gram-positive bacteria at single-cell resolution, which allows for analysis of complex microbial populations, including those not culturable.

BACKGROUND

The human microbiome constitutes a complex collection of microbes that inhabit the human body and their interactions with the host influences overall human health (1). Traditional population-based microbial studies are confined to analysis of cells in bulk with lack of insight into cellular heterogeneity among individual cells (2). Single-cell omics has shed light on how individual cells perceive, respond, and adapt to the environment and has revolutionized the way of approaching genomic heterogeneity of microbial systems at a finer scale (3).

Conventional whole genome amplification (WGA) methods such as multiple displacement amplification (MDA) have the potential to amplify down to the femtogram level of genomic DNA present in a single microbial cell but often results in an uneven genome coverage (4,5). To overcome the amplification bias, we used BioSkryb Genomics Primary Template-directed Amplification (PTA) technology to amplify the genomes of single bacteria. The ResolveDNA™ Microbiome kit was used for amplification, followed by library preparation and next generation sequencing.

METHODS

Bacterial Sorting

Gram negative (*E. coli*) and gram positive (*B. subtilis*) bacterial stock cultures were cultured in LB broth at 37°C for 18 hours. The cultures were filtered for large cell clusters using a 20 µm mesh filter. The bacterial samples were centrifuged and resuspended in 1X DPBS. Single bacteria were sorted on Sony SH800 sorter equipped with a 130um sorting chip into 96-well plates containing 1 µL ResolveDNA Cell Buffer. For mixed species, both the bacterial samples were mixed at a 1:1 ratio. The sorted plates were vortexed, spun down and kept on dry ice, followed by storage at -80°C.

Bacterial DNA amplification

ResolveDNA amplification was performed on the sorted bacterial plates using the ResolveDNA Microbiome kit. Following amplification, the DNA was purified with ResolveDNA bead purification kit and Qubit quantified. The libraries were made using ResolveDNA Library Preparation kit and Qubit quantified, and the sizes were analyzed using TapeStation 4200 (Agilent).

Sequencing and data analysis

The libraries were pooled and sequenced on the Illumina MiniSeq platform at 2 million paired end reads per library. The sequencing data was analyzed for quality, contig assembly, and alignment using BioSkryb BaseJumper microbiome module. Raw fastq files were read filtered and trimmed followed by genome assembly and deconvolution. The deconvoluted samples were assessed for genome completeness.

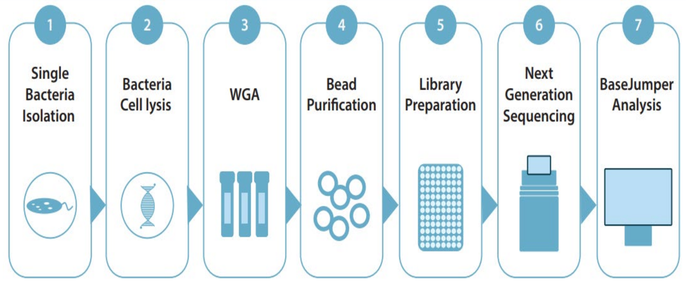


Figure 1. ResolveDNA Microbiome Workflow. FACS isolated single bacterial cells are lysed, followed by amplification with PTA (Primary Template-directed Amplification. Amplified DNA product is then purified and quantified before library preparation followed by sequencing and data analysis with BaseJumper module.

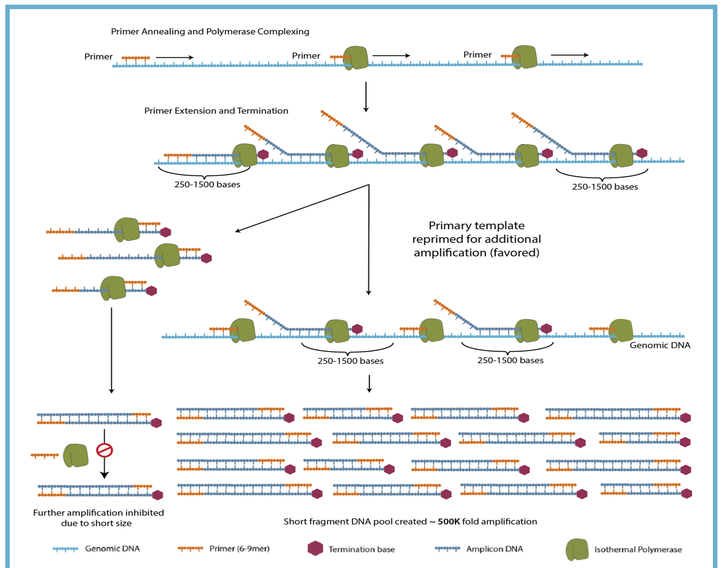


Figure 2. ResolveDNA™ PTA Amplification. Primary Template-derived Amplification (PTA) is a quasi-linear strand displacement process, where exonuclease-resistant terminators are incorporated into the reaction to generate small double-stranded amplification products (6). The small amplicons have a lower propensity to serve as templates for further amplification and thus primers are redirected to the main template, decreasing error propagation and resulting in uniform genome coverage.

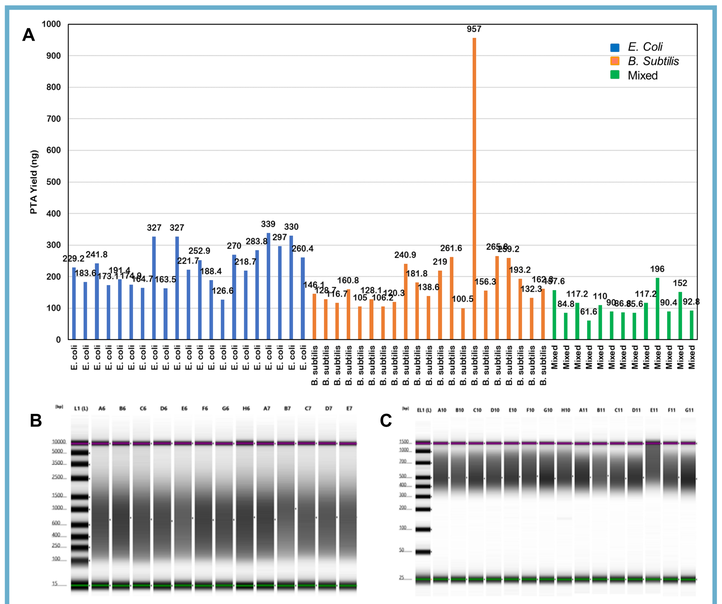
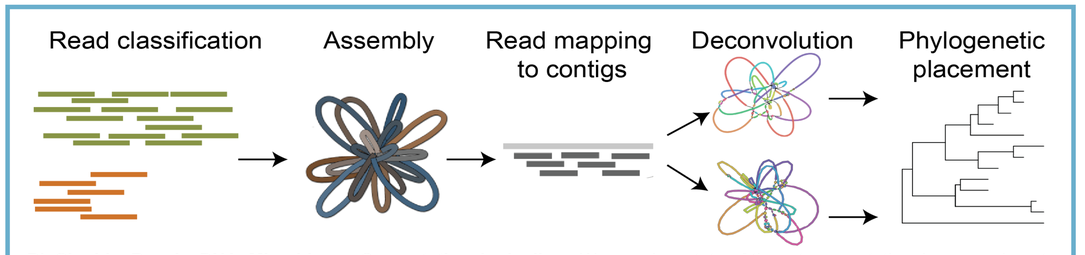


Figure 3. ResolveDNA Microbiome Amplification and Library Preparation. Isolated bacterial samples (*E. coli, B. subtilis* and mixed) were amplified using ResolveDNA, followed by bead purification and quantified whereby yields had an average range of 50-200 ng (A). The amplified samples were analyzed on TapeStation and the product sizes averaged 950 bp (B). Libraries were made using 100 ng of ResolveDNA amplification product as input, using the ResolveDNA Library Preparation kit. The library sizes (~550 bp) were analyzed using D1000 tape on TapeStation. The libraries were pooled and sequenced on an Illumina MiniSeq using a high output 300 cycle kit and the data were analyzed.



BioSkryb's ResolveDNA Microbiome Computational pipeline. We employ state-of-the-art computational approaches to transform raw-data into high-quality genomes for which genomic mining can be performed. Briefly, raw sequencing reads are filtered and processed using Fastp followed up by an optimized iterative genome assembly using Spades. Raw reads are mapped back into the obtained optimized assembly using BWA-Mem and then deconvoluted using the alignment information via Metabat2. In parallel, read-level taxonomic assigned is performed using Kaiju. Finally, full assembly and deconvoluted bins are taxonomically placed in the GTDB reference tree and evaluated for completeness using Checkm and GTDB-tk.

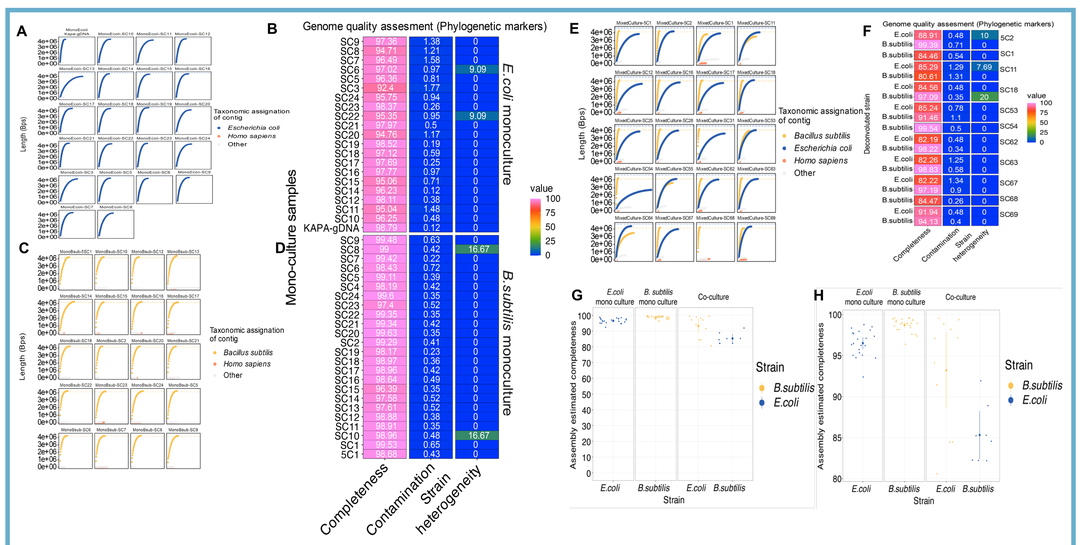


Figure 4. Determining Taxa composition of bacterial samples. Contig cumulative length plots of respective bacteria representing their taxonomic classification based on taxonomic read assignation (A, C, E). Dashed horizontal lines in the plot denote the empirically estimated genome size of the bacterial species in the experiment based on representative complete genomes downloaded from NCBI RefSeq. Respective metrics obtained from each bacterial assembly using phylogenetic marker approach showing the completeness, levels of contamination and strain heterogeneity (B, D, F). Plots depicting the assembly estimated completeness of each bacterial sample respectively (G, H (zoomed completeness)).

RESULTS

The ResolveDNA Microbiome successfully amplified bacterial DNA (*E. coli*, *B. subtilis* and mixed) with yields averaging 50-200ng (Figure. 3A). The analysis of the bacterial samples detected near complete genomes using BaseJumper microbiome module. *E. coli* (22 samples) showed a coverage of almost 95% (Figure. 4A), while the *B. subtilis* (20 samples) showed complete 100% coverage (Figure. 4C). In the mixed cultures (13 samples), we detected both *E. coli* and *B. subtilis* genomes (Figure. 4E). This likely is due to the dispensing of more than one cell into the wells during sorting. One of the panels showed detection of *B. subtilis* alone emphasizing the assay specificity.

Trace amounts of *Homo sapiens* and unidentified species DNA was detected in these samples. Figure. 4 (panels A,C,E) shows the level of contamination found to be minimal with no interference in identifying the bacterial species/genus of origin.

The de novo phylogenetic based assembly estimated the completeness of the bacterial samples with corresponding contamination and strain heterogeneity metrics (panels B,D, F).

SUMMARY

- The ResolveDNA Microbiome kit successfully amplified (yields~ 50-200 ng) single FACS sorted bacteria
- The BaseJumper module analysis effectively detected both single strains of bacteria (Gram positive and Gram negative) and mixed population (co-culture) with high level of completeness.
- Unified efforts of ResolveDNA Microbiome and BaseJumper analysis can be further extended in assessing the genomes of uncultivated microbes and to determine genomic variation between species.

REFERENCES

1. Rackaityte, E., Lynch, S.V. The human microbiome in the 21st century. Nat Commun 11, 5256 (2020). https://doi.org/10.1038/s41467-020-18983-8
2. Hatzenpichler, R., Krukenberg, V., Spietz, R.L. et al. Next-generation physiology approaches to study microbiome function at single cell level. Nat Rev Microbiol 18, 241–256 (2020). https://doi.org/10.1038/s41579-020-0323-1
3. Blainey P. C. (2013). The future is now: single-cell genomics of bacteria and archaea. FEMS microbiology reviews, 37(3), 407–427. https://doi.org/10.1111/1574-6976.12015
4. Ruan, Q., Ruan, W., Lin, X., Wang, Y., Zou, F., Zhou, L., Zhu, Z., & Yang, C. (2020). Digital-WGS: Automated, highly efficient whole-genome sequencing of single cells by digital microfluidics. Science advances, 6(50), eabd6454.https://doi.org/10.1126/sciadv.abd6454
5. Dean, F. B., Nelson, J. R., Giesler, T. L., & Lasken, R. S. (2001). Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. Genome research, 11(6), 1095–1099. https://doi.org/10.1101/gr.180501
6. Gonzalez-Pena V, Natarajan S, Xia Y, et al. Accurate genomic variant detection in single cells with primary template-directed amplification. Proc Natl Acad Sci U S A. 2021;118(24):e2024176118. doi:10.1073/pnas.2024176118