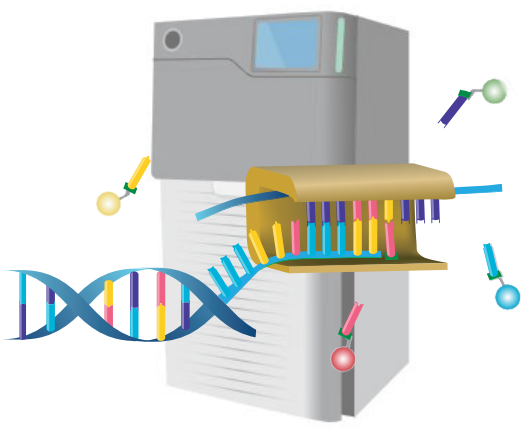
Microbiome Studies with SMRT & Nanopore Sequencing

Microorganisms are the most abundant and diverse forms of life on Earth. Characterizing species diversity and composition of microbes hosted by biota is revolutionizing our understanding of the impact on the maintenance of human health, crop improvement, and species conservation. The accessibility of high throughput sequencing has exponentially increased the number of microbial genomes. However, the commonly employed short-read strategies have extreme limitations, e.g., GC bias, difficulties in mapping repetitive elements, discriminating paralogous sequences, and phasing alleles.
Long read sequencing technologies, PacBio SMRT and Nanopore sequencing, have been seeing a dramatic increase in reading quality, throughput, cost-efficiency, and accuracy.

Pacbio SMRT Sequencing

PacBio Single-Molecule Real-Time Sequencing (SMRT) builds on two revolutionary inventions that overcome major challenges in current sequencing, known as zero-mode waveguides (ZMWs) and the utilization of phospholinked nucleotides. ZMWs provide the smallest available volume for light detection, ensuring a single-base resolution. Phospholinked nucleotides and immobilized DNA polymerases produce a completely natural DNA strand through fast, accurate, and processive DNA synthesis. SMRT sequencing has the capability to produce reads tens of thousands of nucleotides in minutes. This system provides a simple and rapid way to adoptions of metagenomics, epigenomics, and functional analysis.



ONT Nanopore Sequencing

Nanopore sequencing, another long-read single-molecule seq technology, can be used to assemble complete genomes and plasmids from metagenomic samples - resolving similar species and complex genomic regions. Due to the very small diameter of the nanopore, only a single nucleotide is allowed to pass through. While the charged nature of a single base in ATCG is different, the type of base can be detected through the difference of electrical signals, to achieve real-time single-molecule sequencing. Nanopore sequencing can be used for direct sequencing of RNA/DNA molecules, able to directly read out methylated cytosines, preserving the original base modification information. Now, Nanopore technology is applied across microbiology and microbiome research, such as antimicrobial resistance profiling, structural variant analysis, and reference alignment.

Technology Comparison

	PacBio SMRT Sequencing	Nanopore Sequencing	Short-read Sequencing
Principle of Sequencing	Sequencing by synthesis/DNA polymerase	Electronic signals sequencing/exonuclease	Sequencing by synthesis
Average Read Length	10-15 kb, up to 20 kb	up to 900 kb	50–300 bp
Single Pass Error Rate (%)	13	2-13	<1
Read Accuracy (%)	88–90/99.9 (CCS)	96–99	99.9
Bases Per Run	5 – 10 Gb	500 Mb	300 Gb
Runtime (h)	10–30	72	21
Advantages	Long average read length; No amplification of sequencing fragments; The longest individual reads approach 100 kb	Ultra-long reads; Electronic sequencing; Portable; No amplification of sequencing fragments; Accessing methylation data	Very high throughput; Low cost
Disadvantages	Low accuracy; Dependence on DNA polymerase activity	Related high sequencing error	Short reads, resulting in fragmented, partial genomes and ambiguous assembly; Amplification biased

Range of Applications

Long Read Amplicon (16S/18S/ITS) Sequencing

The amplicon (16S/18S/ITS) sequencing is limited to identifying complete and novel species because short-read approaches only obtain hypervariable regions (mean read length of 447 bp). Long-read sequencing offers complete, uniform, nonbiased coverage spanning long amplicons, with high basecalling accuracy (>95%) and strain-level taxonomic resolution. Its value has been assured in gaining more accurate classification and quantification in microbial community structure, abundance, and diversity evolutionary.

Long Read Whole Genome Sequencing

Long-read single-molecular sequencing technologies can retrieve metagenome-assembled genomes (MAGs) with high completeness and overall accuracy, which are not usually finished due to short-read limitations. It improves the contiguity and increases the retrieval of longer and circular contigs (>90% completeness and <5% contamination). Long-read sequencing, eliminating the need for massive computational calculations of genome assembly, enables us to span entire repetitive elements, transposons, and prophage sequences. Nanopore and PacBio long-read technologies have been applied to the gut, oral, stool and other microbiomes.

Long Read Sequencing in Epigenomics Analysis

DNA methylation plays an important role in the regulation of gene expression, virulence, and pathogen-host interactions. Microbial DNA contains three major types of methylation, N4-methylcytosine (4mC), 5-methylcytosine (5mC) and N6-methyladenine (6mA). Long read sequencing allows genome-wide analysis of DNA methylation at single-base resolution, facilitating understanding of the degree, evolutionary and physiological relevance of methylation.

CD Genomics aims at providing the research community with Illumina, PacBio SMRT and Nanopore platforms, and bioinformatics services. The utilization of real-time, long-read sequencing overcomes the challenges associated with traditional short-read sequencing technologies, to fully characterize microbial genomes - sheds new light on microbial evolution, pathogenicity, and antimicrobial resistance.