

Metagenomic insights of a complex soil sample using the PacBio Sequel II



Jack Royle¹, Apoorva Prabhu², Jiyuan An³, David Hawkes¹, Alistair Grinham⁴, Christian Rinke²

¹ Australian Genome Research Facility Ltd

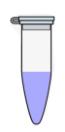
² Australian Centre for Ecogenomics, University of Queensland

³ Queensland University of Technology

⁴ School of Civil Engineering, University of Queensland.

De novo assembly of metagenomic samples has traditionally provided a challenge when dealing with highly diverse, complex sample types. Short read technology often produces highly fragmented contigs, which can result in assembly errors and complications during the binning process when clustered into metagenome-assembled genomes (MAGs). PacBio HiFi sequencing, with a low error rate and significantly longer read length (when compared to short-read data), therefore presents several advantages in the assembly and curation of full-length MAGs

Methodology



- Complex soil gDNA samples sheared to >8Kb in size.
- Prepared for Sequel II sequencing using PacBio TPK 2.0.



- Sequencing performed on single SMRT Cell 8M.
- High accuracy circular consensus reads (CCS) generated.

HiFi-MAG-Pipeline

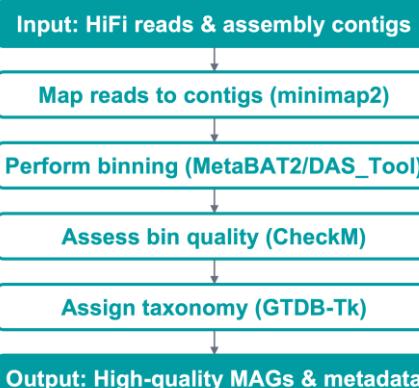


Fig. 1. HiFi MAG Pipeline workflow.

Results



Fig. 2. HiFi-MAG-Pipeline output. Medium quality: >50% completeness. High Quality: >70% completeness, <10% contamination. Near-complete: >90% complete, <5% contamination

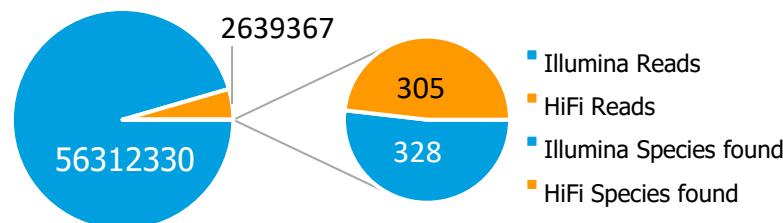


Fig. 3. Comparison of Illumina vs PacBio HiFi 16S profiling on the same metagenomic dataset using GraftM.

- Similar taxonomic identification levels with significantly less reads using the same bioinformatic approaches.

Summary

- Eighty-four quality MAGs assembled from 1 SMRT cell 8M.
- Four complete, single contig, circular MAGs generated.
- Significant MAG assembly and binning improvements compared to previous methods.
- Metagenome assemblies from complex soil samples is achievable.

