MedGenome Labs Ltd.
3rd Floor, Narayana Nethralaya Building, Narayana Health City,
#258/A, Bommasandra, Hosur Road, Bangalore - 560 099, India .
Tel : +91 (0)80 67154989 / 990, Web: www.medgenome.com

MEDGENOME

Whole Genome Metagenome Analysis

O753855

Submitted to

Mr. B Srikanth

MCC18355-Sree Biologix

Report by

Medgenome Labs Pvt Ltd

Bangalore-560099

23-12-2024

MedGenome Labs Ltd.
3rd Floor, Narayana Nethralaya Building, Narayana Health City,
#258/A, Bommasandra, Hosur Road, Bangalore - 560 099, India .
Tel : +91 (0)80 67154989 / 990, Web: www.medgenome.com

Contents

MedGenome Labs Ltd.
3rd Floor, Narayana Nethralaya Building, Narayana Health City,
#258/A, Bommasandra, Hosur Road, Bangalore - 560 099, India .
Tel : +91 (0)80 67154989 / 990, Web: www.medgenome.com

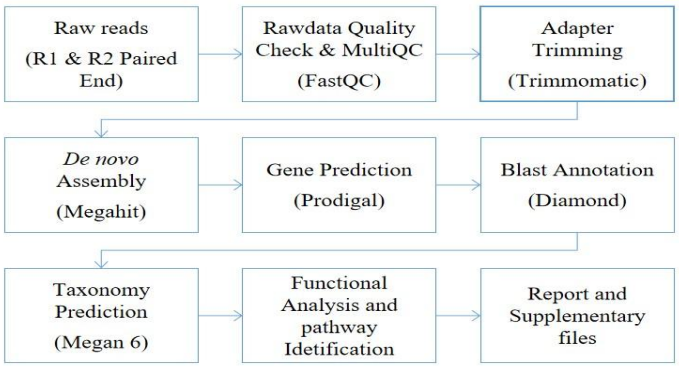
Overview

The given samples were sequenced using NovaSeqXplus with a read length of 151 bp. The samples were taken for whole genome metagenome analysis. The *de novo* assembly was carried out using the reads to obtain the scaffolds. These scaffolds were then used for gene prediction. The abundances at the phylum, genus and species level are given. The abundances in terms of SEED annotations are also given.

Table 1 : Samples Details

Sample name	Sample ID
8801433	FRW_2024_01

Figure 1: Bioinformatics workflow



Methodology

Read quality check

Initially, we checked the following parameters from the samples fastq file - Base quality score distribution, sequence quality score distribution, average base content per read, GC distribution in the reads, PCR amplification issue, overrepresented sequences and adapters.

Based on the quality report, the fastq files were trimmed to retain high-quality sequences and the low-quality sequence reads were excluded from the analysis. The adapter trimming was performed using the fastq-mcf tool (version- 1.04.803).

Denovo Metagenome assembly

Again, the Adapter trimed reads were *de novo* assembled using Megahit (version. 1.2.9).

ORF prediction and annotation

The Primarily assembled genome was used for ORF prediction and ORF annotation using Prodigal (v. 2.6.3).

Taxonomy classification and functional annotation

For predicted ORF the taxonomic classification and identification of SEED pathway was performed using Megan 6.

Data Summary

The given sample was taken for whole genome metagenome analysis and 3.7 GB data was generated for the samples . Q30% was above 92 % and GC% is 48 %. Kindly refer to the QC report for further details. The quality check for the samples after adapter trimming has been performed and the adapter trimmed sequences are taken for downstream analysis.

Table 1: Data summary

Sample_name	Data before AT (GB)	Data after AT (Gb)	Read Quality before AT	Read Quality after AT	GC% before AT	GC% after AT
FRW_2024_01	3.72	2.58	38.66	39.26	48.61	47.67

*AT - Adapter trimming

Bioinformatics Analysis Results

Figure 2: Data before and after adapter trimming

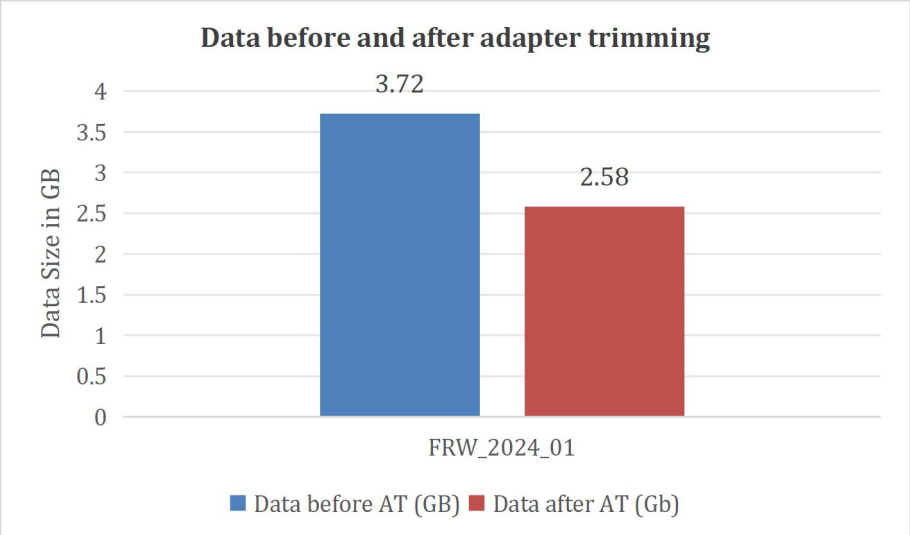


Figure 3: Average Read quality before and after trimming

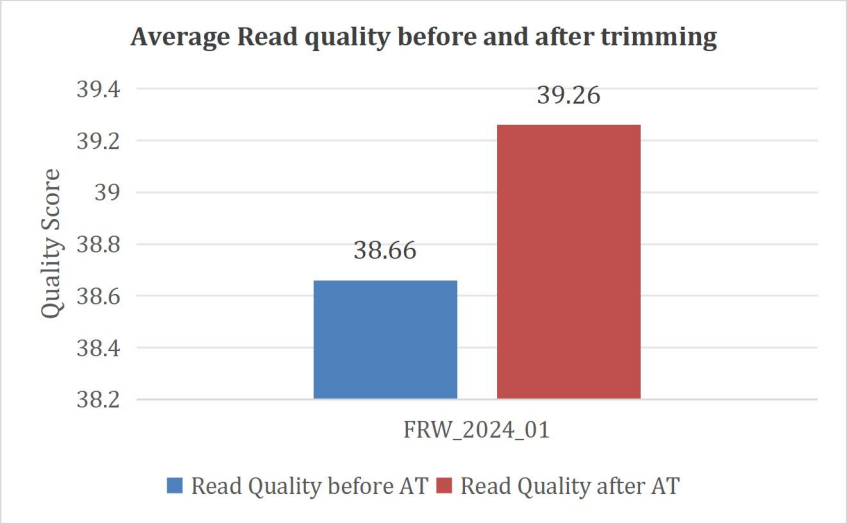
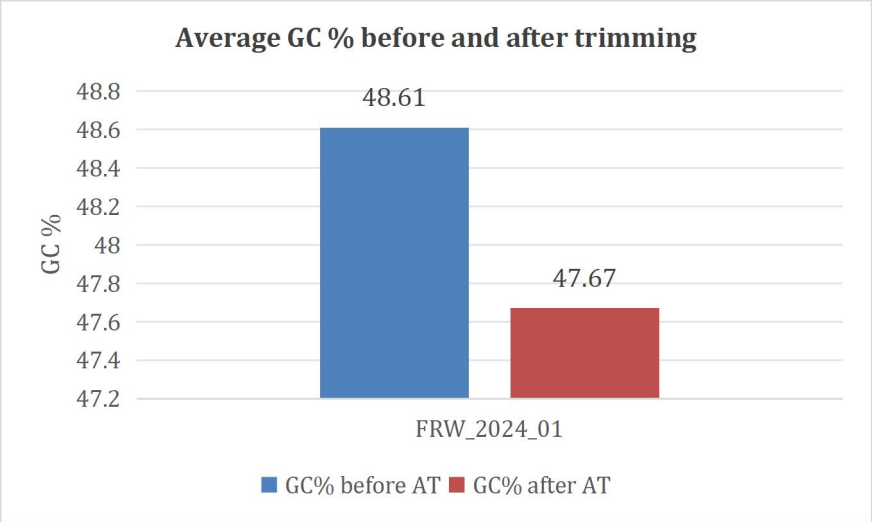


Figure 4: Average GC % before and after trimming



Denovo Metagenome assembly

Denovo metagenome assembly was carried out using megahit assembler (v1.2). Megahit makes use of succinct *de Bruijn* graphs which are compressed representations of *de Bruijn* graphs. The assembled scaffolds were taken for further downstream analysis. The distribution of scaffolds, N50 values are given in table 2.

Table 2: Denovo assembly stats

Sample name	contigs (>= 1000 bp)	Largest contig	N50	N75	L50
FRW_2024_01	9211	904399	11330	2084	909

ORF Prediction

Open reading frame (ORF) prediction is done using Prodigal (v2.6.3). The assembled scaffolds from the *de novo* assembly is taken for the ORF prediction. The obtained ORFs are filtered using in house Perl scripts. ORFs of length below 200 bp being filtered. Table 4 depicts the ORF length distribution.

Table 3: ORF length distribution

Sample name	Total genes	genes>200bp	500bp>genes<1000bp	1000bp>genes
FRW_2024_01	63728	25432	22594	15702

Microbial Diversity

Bacterial kingdom is abundant in all the samples. The taxa wise microbial abundance is given in the supplementary sheet.

Taxonomy level Comparison

Figure 6: Distribution of Phylum

Phylum **Proteobacteria** was the most abundant followed by **Firmicutes** in given sample.

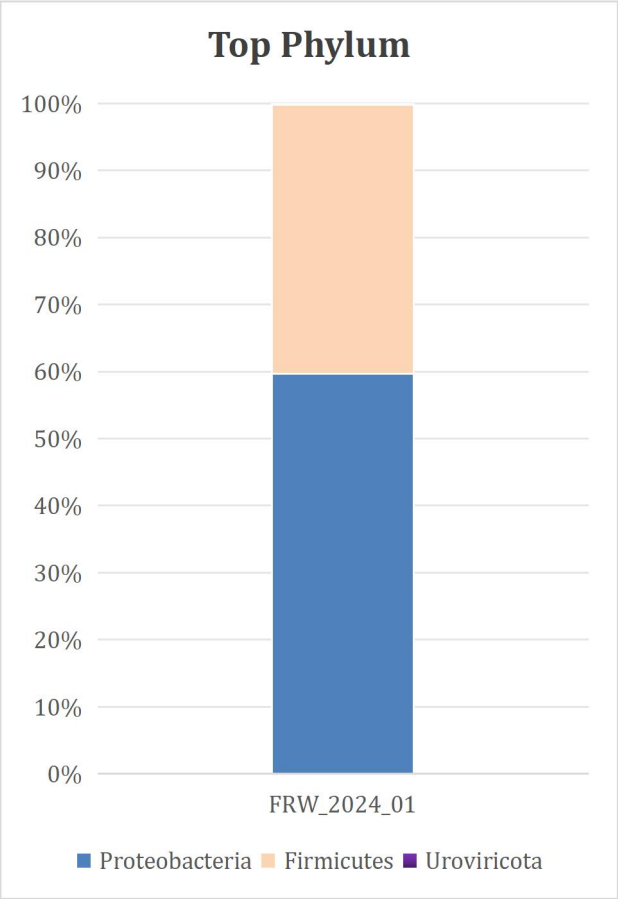
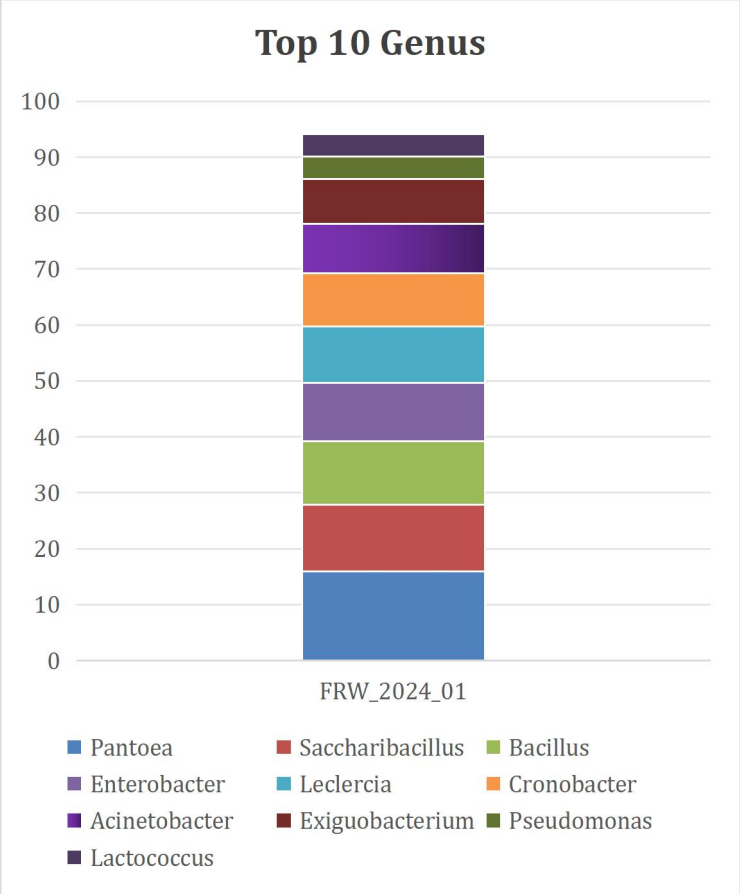


Figure7: Distribution of Genus

Genus **Pantoea** was the most abundant followed by **Saccharibacillus** in given sample.



Supplementary Files:

Supplementary1.1: Denovo assembly stats
Supplemenatry1.2: Gene prediction stats
Supplemenatry1.3: Microbial diversity
Supplemenatry1.4: Seed Pathways
Supplemenatry1.5: Denovo Assembled fasta
Supplemenatry1.6: Predicted ORF(fasta)
Supplemenatry1.7: Supplementary Images

References:

1. Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics, 25:1754-60. [PMID: 19451168]
2. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph - Dinghua Li, Chi-Man Liu, Ruibang Luo, Kunihiko Sadakane, Tak-Wah Lam Bioinformatics - 2015
3. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010;11:119. Published 2010 Mar 8. doi:10.1186/1471-2105-11-119