

Functional Profiling and Benchmarking of Soil Microbiomes Using HUMAnN and PICRUSt2

1. Introduction

This project investigated the functional potential of microbial communities in rice soil using shotgun metagenomics (HUMAnN 3.9) and predictive metagenomics (PICRUSt2). The goal was to benchmark both methods and interpret dominant functional pathways and enzymes within these microbiomes.

2. Materials and Methods

Data preprocessing included read quality control (fastp), merging (bbmap), taxonomic profiling (MetaPhlAn), and functional annotation (HUMAnN). Regrouping was performed using KEGG Ortholog (KO) and MetaCyc/EC mappings. PICRUSt2 functional predictions from 16S were used for benchmarking correlation and coverage tests.

3. Results

- HUMAnN identified 3,274 UniRef90 gene families. - Option A (KEGG Ortholog): 119 shared KOs between HUMAnN and PICRUSt2; Spearman $r = 0.186$, Pearson $r = 0.228$. - Option B (MetaCyc/EC): 227 MetaCyc reactions, 151 EC enzymes identified. - Dominant enzymes: oxidoreductases, transferases, hydrolases. - Differential analysis (PICRUSt2): rice soil enriched in carbon fixation and denitrification; shrimp soil enriched in sulfur metabolism.

4. Discussion

HUMAnN + MetaCyc/EC provides richer biological interpretability, whereas PICRUSt2 covers a broader range of KOs but with less quantitative accuracy. The benchmarking demonstrates complementary strengths—HUMAnN offers direct evidence while PICRUSt2 predicts unseen functions.

5. Conclusion

Combining HUMAnN (shotgun-based) and PICRUSt2 (16S-based) approaches yields a comprehensive view of soil microbiome functionality. MetaCyc and EC regrouping outperform KEGG in environmental coverage and interpretability.

6. Figures Summary

- Benchmarking_summary.png – KO correlation, top10 KOs, EC class. - KO_heatmap_methods.png – shared KO heatmap. - EC_top20.png, RXN_top20.png – top

enzymes and reactions. - Functional_differentiation_rice_vs_shrimp.png – condition-specific enrichment.

7. Future Work

- Process shrimp soil through HUMAnN for direct EC/RXN comparison. - Update UniRef90→KO mapping (v2024) to improve KEGG overlap. - Apply pathway-level differential abundance (LEfSe, DESeq2).