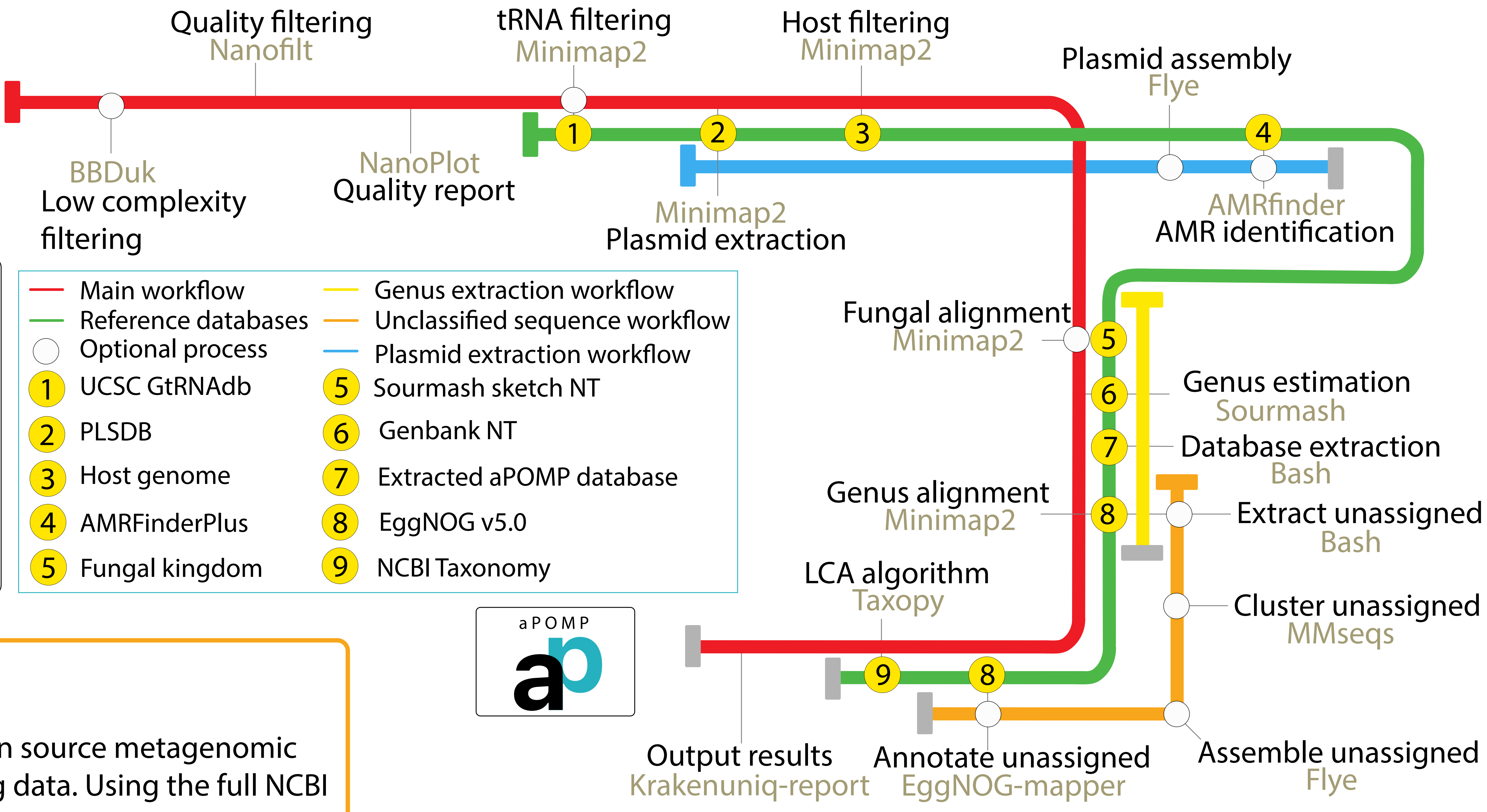
A **PO**retable **M**etagenomics **P**ipeline (**aPOMP**): Fast and Efficient Nanopore sequencing-based metagenomics

Vikas Peddu[1], Ben Bradley[2], Daniel Kim[1*], Miten Jain[3,4*]

[1] Department of Biomolecular Engineering and Bioinformatics, University of California Santa Cruz, Santa Cruz, CA, USA
[2] Department of Pathology, University of Utah School of Medicine, Salt Lake City Utah, UT, USA
[3] Department of Bioengineering, Northeastern University, Boston, MA, USA
[4] Department of Physics, Northeastern University, Boston, MA, USA

aPOMP is a nanopore specific metagenomics pipeline optimized to run efficiently on local machines without sacrificing reference database size, speed, or accuracy



Abstract

We present aPOMP: a Nextflow based open source metagenomic analysis pipeline for Nanopore sequencing data. Using the full NCBI NT database, aPOMP employs a pre-filtering step to estimate abundant genera, followed by directed alignments to the detected genera and LCA classification to resolve strain level specificity of the organisms in a sample. When requested, antimicrobial resistance genes can also be identified. Compared to existing long read metagenomic pipelines our approach performs well in recall, precision, F1-score, and speed. aPoMP can also be implemented offline in the field on resource limited computers running MinIONs, or with adequate computational resources for basecalling, profile reads in real time. We will present results from aPOMP applied to Nanopore sequencing data from a orthopedic device infection dataset[1].

Discussion

- aPOMP identifies a broad range of pathogenic and environmental species
- Runtimes < 1 day for MinION/PromethION data
- Classification is agnostic to Guppy models
- Robust F1 score, Precision, and Recall across all taxonomic ranks
- Timely classification using the cloud, or on premises compute
- aPOMP can be run in real time

R10.4 Zymobiomics Mock standard

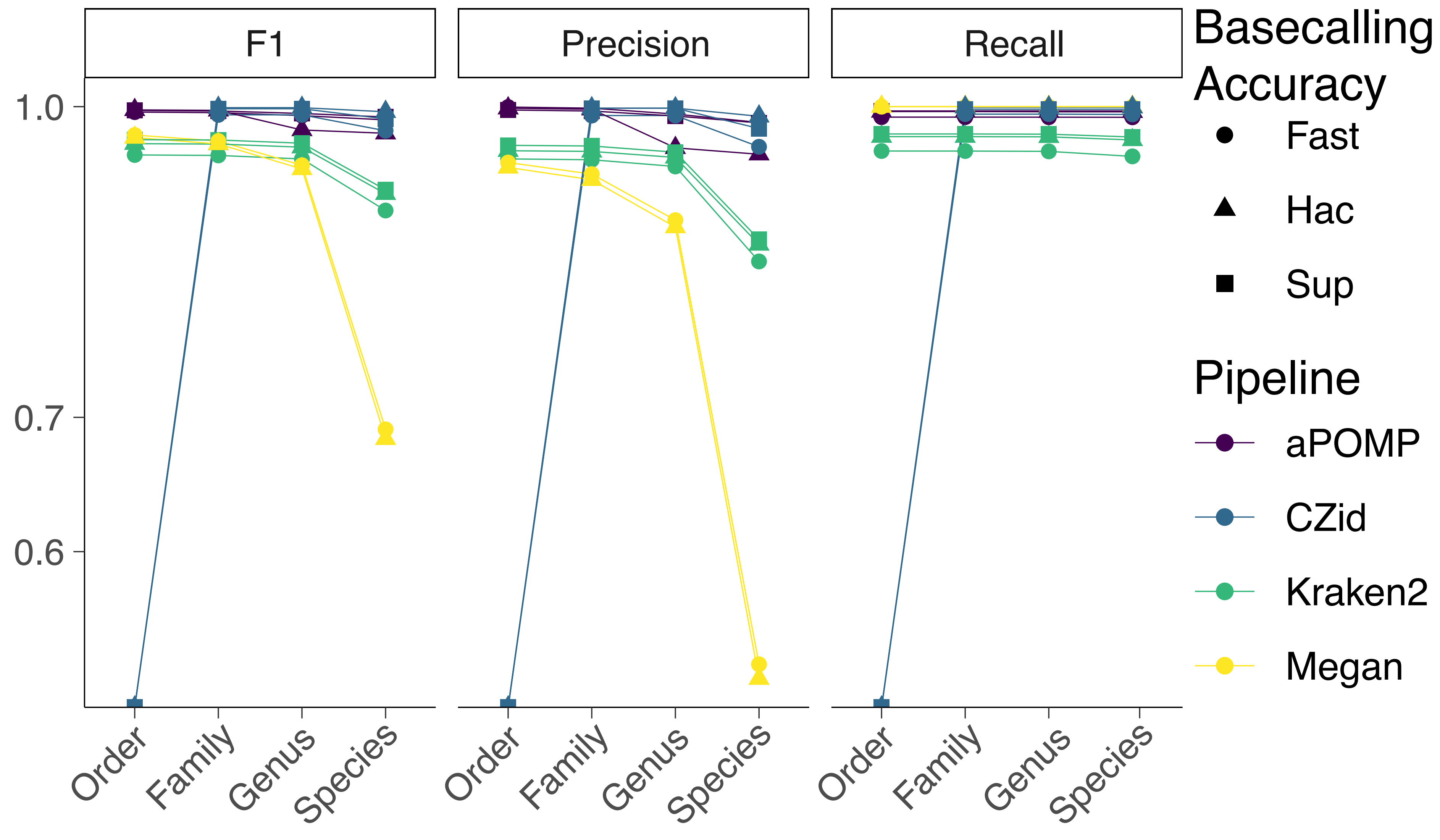


Figure 1: F1, Precision, and Recall scores for major taxonomic ranks for long and short-read metagenomics pipelines against the Zymobiomics D6300 Mock dataset[2,3]

Introduction

- ONT sequencing offers long reads, portable equipment, and a low financial barrier to entry
- aPOMP is an accurate and portable long read metagenomic pipeline for classification of known sequence and annotation of novel sequence.
- Species level alignments are resolved using a custom LCA algorithm
- Plasmids can be assembled and annotated for AMR genes
- Unassigned reads can be clustered and annotated using the eggNOG database
- aPOMP has optimizations for low-resource compute

Orthopedic device infection classification

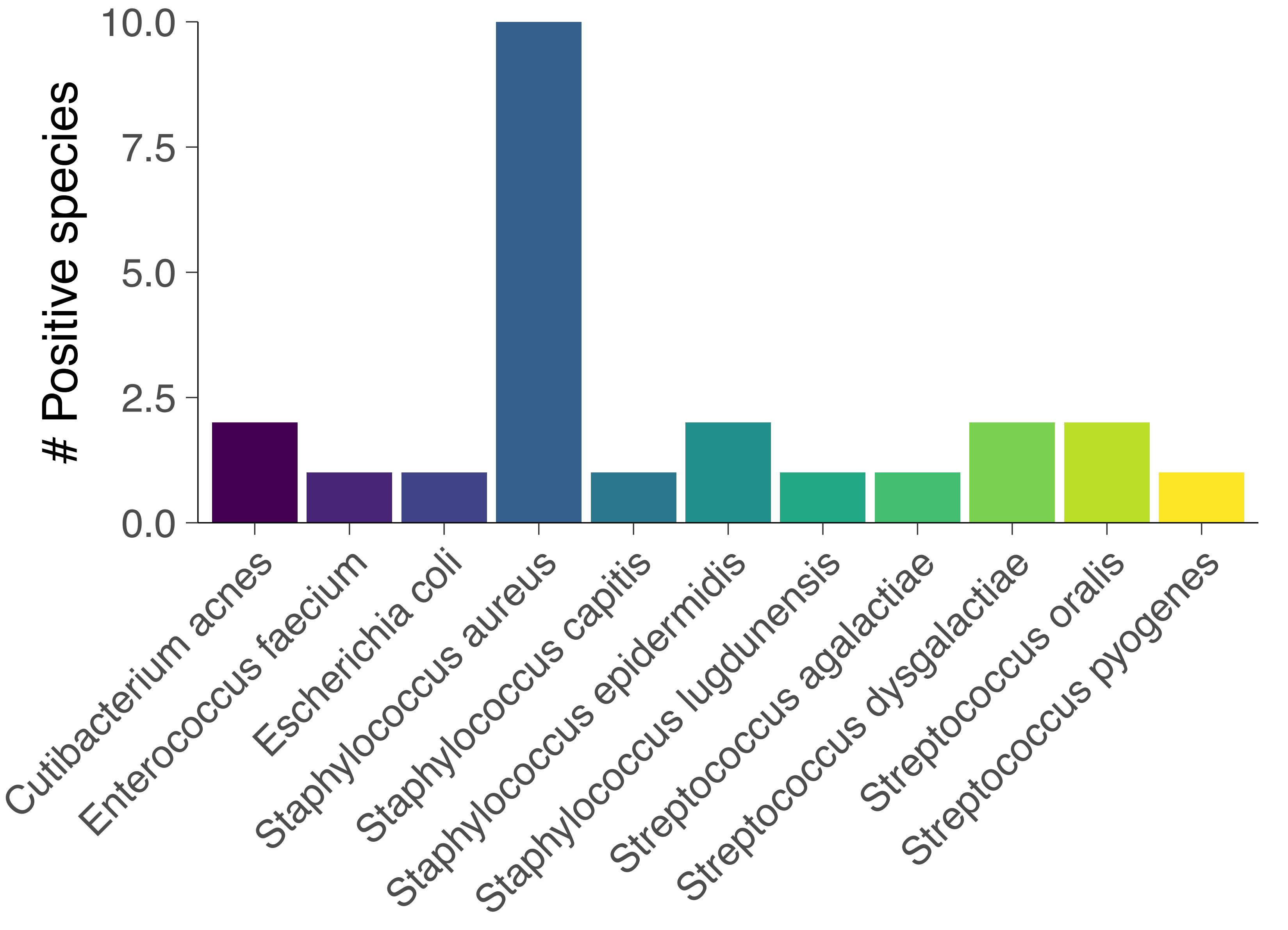


Figure 2: Species classified by aPOMP in a subset of samples from an orthopedic joint infection[1]

Bibliography

1.Street, T. L. et al. Clinical Metagenomic Sequencing for Species Identification and Antimicrobial Resistance Prediction in Orthopedic Device Infection. Journal of Clinical Microbiology 60, (2022).

2.Nicholls, S. M., Quick, J. C., Tang, S. & Loman, N. J. Ultra-deep, long-read nanopore sequencing of mock microbial community standard GigaScience 8, (2019).

3.Nanopore GridION and PromethION Mock Microbial Community Data Community Release. GitHub https://github.com/LomanLab/mockcommunity (2023)