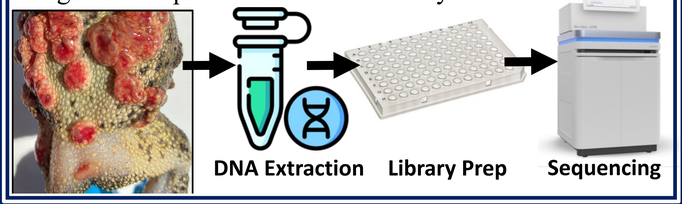
Emergence and Molecular Characterization of Protist *Amphibiothecum penneri* in American Toads (*Anaxyrus americanus*) in Northern New England

Julia Murray

Department of Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, Durham, NH 03824

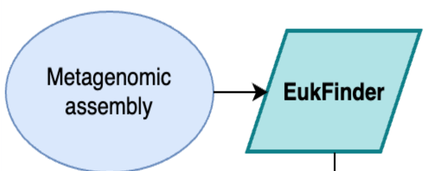
Introduction

- Protist are a growing, but poorly characterized threat to amphibian health
- *Amphibiothecum penneri* is a unicellular holozoan protist pathogen of American toads, that does not have a published genome.
- From 2022-2025 increasing numbers of toads in the Lamprey River Watershed were diagnosed with *A. penneri* infection by the MEED lab and HCGS.
- Our aim was to assemble the *A. penneri* genome from a clinical sample.
- Since host and pathogen are eukaryotes, it is challenging to parse out their genomes in a mixed sample.
- Targeted filtering and validation analyses were used to generate a protist-enriched assembly.



Methodology

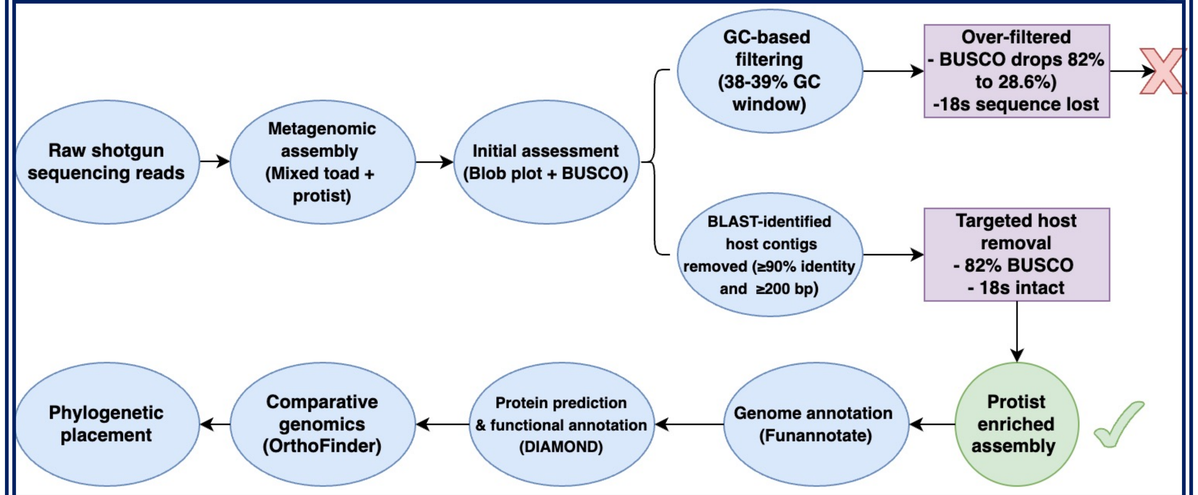
- Metagenomic contigs were classified using EukFinder, a taxonomic binning tool that assigns contigs based on sequence similarity and composition.
- Contigs were partitioned into Euk (known eukaryotic), EUnk (eukaryotic unknown), and Pro (prokaryotic) bins, with the target protist genome recovered from the EUnk bin.
- Assembly composition and completeness were evaluated pre- and post-filtering using BUSCO with eukaryota and tetrapoda lineage datasets to distinguish protist signal from host contamination.



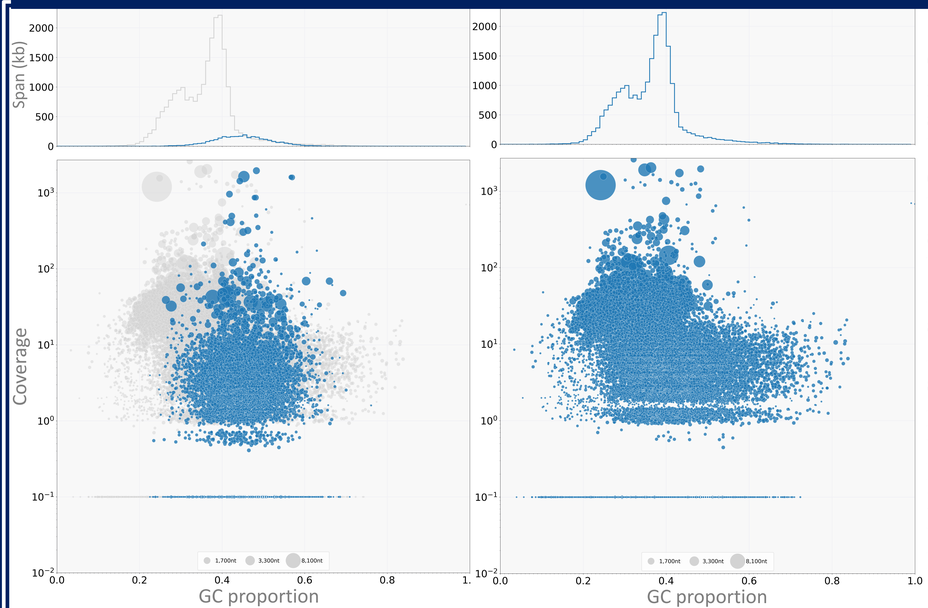
Bins

Euk	*EUnk*	Prok
➢Contigs that match known eukaryotic genomes ➢Mostly host ➢Possible conserved euk sequences	➢Contigs defined as eukaryotic with no close reference genome ➢Target protist ➢ Residual host contamination	➢Contigs matching bacterial/archaeal sequences ➢Not relevant to target DNA

Workflow



Evaluation



Blob plots show GC content versus coverage for contigs before (left) and after (right) BLAST-based host removal (≥90% identity, ≥200 bp). The unfiltered assembly forms a single broad cluster representing mixed host and protist sequences. After filtering, the assembly is enriched for protist-derived contigs with reduced host signal. Remaining clustering likely reflects GC variation and/or residual host DNA.

Fig 1 (above): Blob plot comparison before and after BLAST-based host removal.

BUSCO analysis (Euk and Tetrapod datasets) was used to assess genome completeness and host contamination. GC-based filtering reduced eukaryotic completeness, indicating over-filtering, whereas BLAST-based filtering preserved completeness while maintaining low tetrapod signal. The similarity between Initial and BLAST-filtered Euk results suggests that removed sequences may primarily represent conserved eukaryotic genes shared between host and protist.

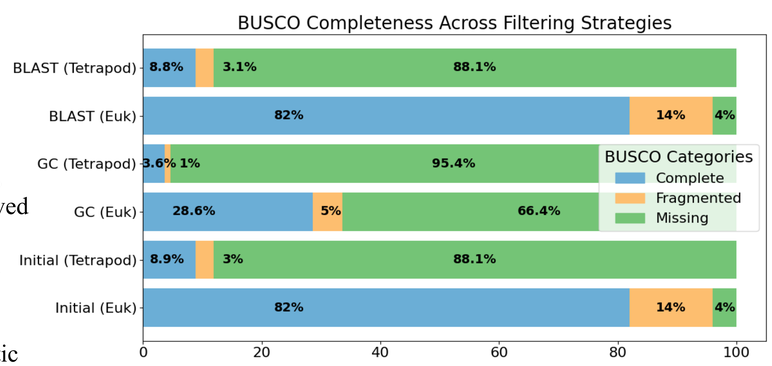
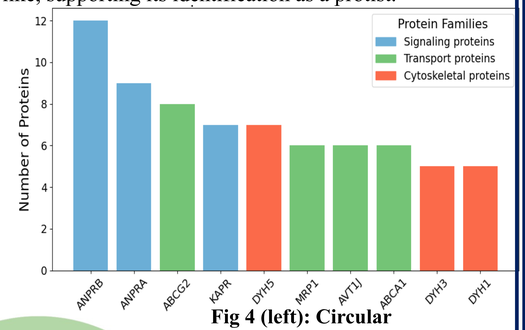


Fig 2 (above). BUSCO completeness between filtering strategies.

Results

- Protein homology searches using DIAMOND, a sequence alignment tool for comparing predicted proteins to reference databases, found matches across diverse eukaryotes. This reflects conservation of core eukaryotic genes and database bias toward well-annotated model organisms.
- Functional annotation revealed conserved protein families involved in transport, cytoskeletal dynamics, and signaling, consistent with a eukaryotic genome.
- Phylogenetic analysis using OrthoFinder, a tool that infers evolutionary relationships based on shared orthologous proteins, places *A. penneri* within holozoan protists, consistent with prior 18S rRNA-based identification. The well-annotated model organism *Xenopus laevis* (African clawed frog) does not cluster with the *A. penneri* proteome, confirming that the recovered genome is not amphibian-like, supporting its identification as a protist.

Fig 3 (right): Top annotated protein families identified by DIAMOND, color coded by protein function.



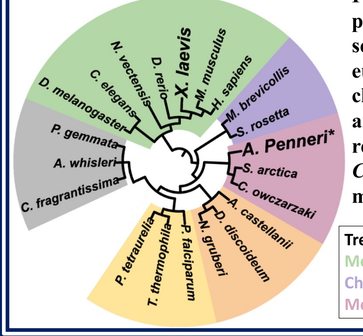


Fig 4 (left): Circular phylogenetic tree with colored sectors denoting major eukaryotic clades. *A. penneri* clusters within holozoan protists and groups with unicellular relatives such as *S. arctica* and *C. owczarzaki*, distinct from metazoan lineages.

Conclusions

- Separation of a co-extracted host and protist genome using EukFinder-based binning and targeted filtering recovered a high-quality draft genome of *A. penneri*.
- While sufficient for phylogenomic placement, the assembly remains below reference quality due to fragmentation and residual contamination.
- The recovered genome supports its classification as a holozoan protist and enables further analysis of pathogenic mechanisms, diagnostics, and mitigation strategies.

References

BUSCO: Manni et al., 2021, *Mol. Biol. Evol.,* DIAMOND: Buchfink et al., 2021, *Nat. Methods,* OrthoFinder*:* Emms & Kelly, 2019, *Genome Biol.,* GeneMark: Lomsadze et al., 2005, *Nucleic Acids Res,* Funnotate: Palmer & Stajich, 2020, *Zenodo,* Eukfinder (Roger Lab), available at: https://github.com/RogerLab/Eukfinder

This publication was supported by the NH-INBRE program and the Center for Integrated Biomedical and Bioengineering Research (CIBBR) through grants from the National Institute of General Medical Sciences of the National Institutes of Health under Award Numbers P20GM103506 and P20GM113131, respectively.