

gtnet: A Python package for taxonomic labelling with the Genome Taxonomy Network

Andrew J. Tritt ¹

¹ Applied Math and Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: 

Submitted: 05 March 2024

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Summary

The field of metagenomics seeks to understand the genomic and functional diversity of microbial communities. Modern metagenomic sequencing pipelines produce unlabelled genomic sequences at an unprecedented rate. Processing of these sequences, i.e. contigs, involves labelling the taxonomy of these contigs. In recent years, the metagenomics fields has been coalescing around the use of the Genome Taxonomy Database ([Parks et al., 2021](#)), a phylogenetically informed taxonomy for consistently labelling microbial taxa. The Genome Taxonomy Network, GTNet, is a neural network capable of classifying metagenomic contigs with taxonomic labels from the Genome Taxonomy Database.

Statement of need

gtnet ([Tritt, 2024](#)) is a Python package and command-line utility built on top of GTNet. The purpose of this software is to make the predictive capabilities of the GTNet easily accessible to the metagenomics community.

In addition to deploying GTNet, the gtnet software seeks to address other outstanding issues in the field. Many taxonomic classification tools are still released as source code in tarball formats, require installation of third-party software that may no longer be maintained, or use application-specific output formats. These issues make existing tools cumbersome and difficult to use. By leveraging the existing Python ecosystem, we seek to make a tool that is easier to use and version for the sake of user-friendliness and reproducibility.

By releasing easily-installable and user-friendly software capable of generating GTDB taxonomies, we hope to lower the technical barrier to wide adoption of standardized taxonomy across the metagenomics field.

References

- Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P.-A., & Hugenholtz, P. (2021). GTDB: An ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research*, 50(D1), D785–D794. <https://doi.org/10.1093/nar/gkab776>
- Tritt, A. (2024). *The Genome Taxonomy Network* (Version 0.0.6). <https://github.com/exabiome/gtnet>