ENS-2002 Final Report

# An analysis on the functional relationships inside the skin microbiome of *Dendrobates tinctorius*

## Introduction

[background stuff, do some reading] “microbiomes are important because… (ref)” / “they have found to have many effects on their hosts, such as … (ref)” / “Chitridiomycosis is caused by Bd fungus, this is bad because…(ref)” [just some ideas]

[intro stuff, again, read] - “*Dendrobates tinctorius* is a model species for this analysis because it is vulnerable to chitridiomycosis, and … (ref)”/ “microbiomes have been found to relate to disease resistance(ref)”/ “an understanding of the functions and relationships at a genomic level is useful because… (ref)”

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Table showing the phylogenetic names as discerned by GTDB-TK on the using version, it could not find any specific species, possibly due to the MAG origin   | **Family** | **Genus** | **Number of samples** | | --- | --- | --- | | Sphingomonadaceae | *Sphingomonas* | 4 | | Microbacteriaceae | *Microbacterium* | 2 | | Brevibacteriaceae | *Brevibacterium* | 1 | | Dermabacteraceae | *Brachybacterium* | 1 | | Enterobacteriaceae | *Pantoea* | 1 | | **Total** | | **9\*** | | Samples given to me by Aaron Comeault and Alberto Orta on the GET INFO FROM AARON | | | | \*I received 10 samples, however GTDB-TK could not place sample 1Dt100h in a family, so I discarded it from further study | | | |

## Methods

Many computational programs were employed for this analysis. Free, online samples, taken from the NCBI website’s database, were used BECAUSE…

This includes: slurm, bash, R, BLAST, eggNOG-mapper, GTDB-TK, CheckM/CheckM2, etc to produce computational analysis of the genetic makup and relationships between 10 genera sampled by Scientists in the ECW labs at Bangor, and how they relate to online samples…

[**WWWWWH**/As mentioned above, I used \_\_\_\_package, in \_\_\_, to produce \_\_\_ because \_\_\_. I did this by \_\_\_, \_\_\_ files were created and can be viewed in the GitHub repo. Data was downloaded/ sites were accessed on \_\_\_]

### Complex computational techniques?

This course has made me learn how to do advanced code in R, involving for loops and if statements, as well as using functions in packages such as “tidyverse” (such as mutate() or select()) to do complex analysis. I have also leaned how to interface with slurm so as to work on the hawk supercomputer, as this permitted the use of eggNOG-mapper en masse.

I also got a lot of experience querying databases using API calls. This was necessary to pull down large amounts of data from the NCBI and KEGG websites swiftly. Both websites have a REST API and packages in R that allow me to call them and download the data from inside an R script.

### Initial genera analyses

### CheckM/2

### GTDB-TK

This programme was used to accurately identify the phylogentic identification of the Bangor-generated samples. Specifically, the de\_novo workflow was used. This was run on hawk using the de\_novo\_gtdbtk.sh slurm script. This produced a list of samples that i needed to download the data for.

### Phylogenetic trees

### Heatmaps

The main purpose of this module for me was to create as many heatmaps as necessary to analyse the genetic content and function of the main genera of focus (Sphingomonas, Pantoea, Brevibacterium, Microbacterium and Brachybacterium). Once I have created these to identify important Proteins and pathways I can then analyse them to see how the genera significantly differ from each other, this is all in service of seeing what impacts they may have on the bacterial skin microbiome of the host species. A specific protein / process was taken as significant if it was found in over 80% of the individuals of a genus, and less than 50% in each of the others.

## Results

GTDB-TK analysis confirmed many of the preconceived names given to the samples, however, it could not determine even a genus for sample 1Dt100h, so i removed it from further study, it also renamed sample \_\_\_ from *Enterobacter cancerogenus* to *Pantoea sp. (should i mention the two i flipped because they were switched?)*

[As shown by figures X and Y, the Bangor samples match the variation seen amongst the larger population] **[unsure]**. Species inside the genus *Sphingomonas* make up a large proportion of the samples, being 4 of the 9, this could concern oligocolonisation by this genera (Garud and Pollard, 2020). (im not sure about this one)

## Conclusions

[too early to really say anything other than “the samples are what we think they are”, still need to do heatmap and tree analysis]

[In Conclusion, \_\_]

[my next steps will be:

* doing the second heatmap of the inter-family stuff
* sourcing other points of interest
* waiting for Alberto to do more samples for me
* ]

## References

Garud, N.R. and Pollard, K.S. 2020. [Population Genetics in the Human Microbiome](https://doi.org/10.1016/j.tig.2019.10.010). *Trends in Genetics*. **36**(1), pp.53–67.