

Using microbial composition within sputum transcriptome data to stratify patients by asthma severity

Nneoma Adaku

Daniel J Spakowicz *Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT*

Scott Strobel *Department of Molecular Biophysics and Biochemistry; Yale University, New Haven, CT*

Faye Rogers

Introduction

This is the script used to build a tree for Nneoma's fungus and assign its taxonomy.

I started by rolling through the databases from the recent review <http://jcm.asm.org/content/55/4/1011.full> to check if any would be useful for this project.

- BOLD systems http://v4.boldsystems.org/index.php/IDS_OpenIdEngine only have ITS identification.
- Looks like this is a good place for morphological features <https://aftol.umn.edu/> and can even make a nexus file to include in the tree – but aftol has been lost? (goes to godaddy website...)
- BROAD doesn't have an identification portal
- EZBioCloud doesn't have a fungal id section
- FungiDB is just genomics
- UNITE is just ITS
- IndexFungorum doesn't have an id search (but could be useful for morphology)
- CBS can be searched directly for LSU and there are lots of good hits. However, I'd rather find a paper that has gone through the effort of identifying isolates with multiple loci
- SILVA has an LSU search <https://www.arb-silva.de/>
- Identity: 43.61, LCA tax SILVA: None
- SSU Iden: 99.37, LCA tax. SILVA: None
- RDP <http://rdp.cme.msu.edu/classifier/>
- E14504F-LSU Root(100%) Fungi(100%) Basidiomycota(100%) Agaricomycetes(100%) Cantharellales(100%) Ceratobasidiaceae(100%) Thanatephorus(100%)

The RDP result is strong, with 100% confidence in the genus *Thanatephorus*. The CBS searches also found organisms of either *Thanatephorus* (telomorph) or *Rhizoctonia* (anamorph). This will very likely be the genus to which E14504F belongs. In addition, Nneoma and I found a few papers that deal with isolates of *Rhizoctonia*/*Thanatephorus*:

- [?] has a bunch of *Thanatephorus* isolates with genbank accession numbers for ITS and 28S, but nothing outside the genus (which is necessary to demonstrate the circumscription in this case).
- [?] just have ITS and have several genera that were isolated from orchids. It's more orchid-centric than fungus-centric.
- [?] defines a *Thanatephorus* isolate using ITS alone. The tree shows weak node support separating *Thanatephorus* from *Ceratobasidium*, but clearly their isolate is closest to a *Thana*.
- [?] does a really nice job of creating a multi-locus tree. This should be the model going forward.

Methods

I converted the table of genbank accession numbers from [?] to a google spreadsheet.

Here are the files that Nneoma created using the Staden package (pregap & gap). As soon as these have genbank accession numbers I'll add them to the table so that they can be pulled with the other sequences from the table.