Taxonomic assignment of endophytic isolate E14504F

Dan Spakowicz

May 10, 2017

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:

- [Gonzalez et al.] 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).
- [Tupac Otero et al.] just have ITS and have several genera that were isolated from orchids. It's more orchid-centric than fungus-centric.
- [López-Chávez et al.] 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.
- [Gónzalez et al.] 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 [Gónzalez et al.] to a google spreadsheet.

```
# Load table into dataframe
sheet <- gs_title("E14504F")
```

```
## Sheet successfully identified: "E14504F"
```

```
x <- gs_read(sheet)
```

```
## Accessing worksheet titled 'Sheet1'.
```

```
## Parsed with column specification:
## cols(
##   Name = col_character(),
##   `Voucher ID` = col_character(),
##   Host = col_character(),
##   `Geographic Origin` = col_character(),
##   ITS = col_character(),
##   SSU = col_character(),
##   LSU = col_character(),
##   RPB1 = col_character(),
##   RPB2 = col_character(),
##   TEF1 = col_character(),
##   ATP6 = col_character(),
##   Reference = col_character(),
##   Notes = col_character()
## )
```

```
# Convert hyphen-only columns to NA
x <- data.frame(apply(x, 2, function(x) gsub("^-$", NA, x)), as.is = TRUE)

# Take in a character vector of genbank accession numbers and return a fasta file in ape forma
RetrieveSequences <- function(charvec){
  try({
    string <- entrez_fetch(db = "nucleotide", id = charvec, rettype = "fasta")
    lsu <- unlist(strsplit(string, split = "\n"))
    temp <- tempfile()
    write(lsu, temp)
    lsu <- ape::read.dna(temp, format = "fasta")
    return(lsu)
  }, silent = TRUE)
}

# Retrieve all sequence into a list
loci <- colnames(x[,5:11])
fastas <- list()
for (i in loci) {
  fastas[[i]] <- RetrieveSequences(x[,grep(i, colnames(x))])
```

```
}
# Remove those without any sequences
fastas <- fastas[-(which(sapply(fastas, class) == "try-error"))]
```

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.

```
# Read in Nneoma's files
lsu <- read.dna(file = "~/Dropbox/Rainforest project/E14504F sequences/E14504F LSU FULL.fasta"
# rpb1 <- read.dna(file = "~/Dropbox/Rainforest project/E14504F sequences/E14504 RPB1.fasta")
rpb2 <- read.dna(file = "~/Dropbox/Rainforest project/E14504F sequences/E14504F RPB2.fasta", f
# ssu <- read.dna(file = "~/Dropbox/Rainforest project/E14504F sequences/E14504F SSU ver2.fast
tef1 <- read.dna(file = "~/Dropbox/Rainforest project/E14504F sequences/E14504F TEF1.fasta", f
```

```
# Align sequences
alns <- lapply(fastas, ape::muscle)
```

Results and Discussion

References

Dolorez Gonzalez, Donald E. Carling, Shiro Kuninaga, Rytas Vilgalys, and Marc A. Cubeta. Ribosomal DNA systematics of ceratobasidium and thanatephorus with rhizoctonia anamorphs. 93(6): 1138–1150. ISSN 0027-5514. doi: 10.2307/3761674. URL http://www.jstor.org/stable/3761674.

Dolores Gónzalez, Marianela Rodriguez-Carres, Teun Boekhout, Joost Stalpers, Eiko E. Kuramae, Andreia K. Nakatani, Rytas Vilgalys, and Marc A. Cubeta. Phylogenetic relationships of rhizoctonia fungi within the cantharellales. 120(4):603–619. ISSN 1878-6146. doi: 10.1016/j.funbio.2016.01.012. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5013834/.

Mariana Yadira López-Chávez, Karina Guillén-Navarro, Vincenzo Bertolini, Sergio Encarnación, Magdalena Hernández-Ortiz, Irene Sánchez-Moreno, and Anne Damon. Proteomic and morphometric study of the in vitro interaction between oncidium sphacelatum lindl. (orchidaceae) and thanatephorus sp. RG26 (ceratobasidiaceae). 26(5):353–365. ISSN 1432-1890. doi: 10.1007/s00572-015-0676-x.

J. Tupac Otero, James D. Ackerman, and Paul Bayman. Diversity and host specificity of endophytic rhizoctonia-like fungi from tropical orchids. 89(11):1852–1858. ISSN 0002-9122. doi: 10.3732/ajb.89.11.1852.