Generate LTDE phylogenetic tree

William R. Shoemaker, Jordan Bird, Stuart E. Jones, and Jay T. Lennon

21 February, 2018

1) Setup

A. Retrieve and Set Your Working Directory

```
rm(list = ls())
getwd()
#setwd("~/GitHub/LTDE/")
knitr::opts_knit$set(root.dir = '~/GitHub/LTDE/')
```

B. Load Packages

The `require()` function in `R` returns `TRUE` if the package was successfully loaded or `FALSE` if the package failed to load. This `for` loop loads each package and installs the package when `require()` returns `FALSE`.

```
# package.list <- c('ape', 'seqinr', 'phylobase', 'adephylo',
# 'geiger', 'picante', 'stats', 'RColorBrewer', #'caper',
# 'phylolm', 'pmc', 'ggplot2', 'tidyr', 'dplyr', 'phangorn',
# 'pander', 'phytools', 'psych')
library("ape")
library("seqinr")
```

2) Make the alignment

Run the file `persistence_707_712_721.sh`, which contains the following code.

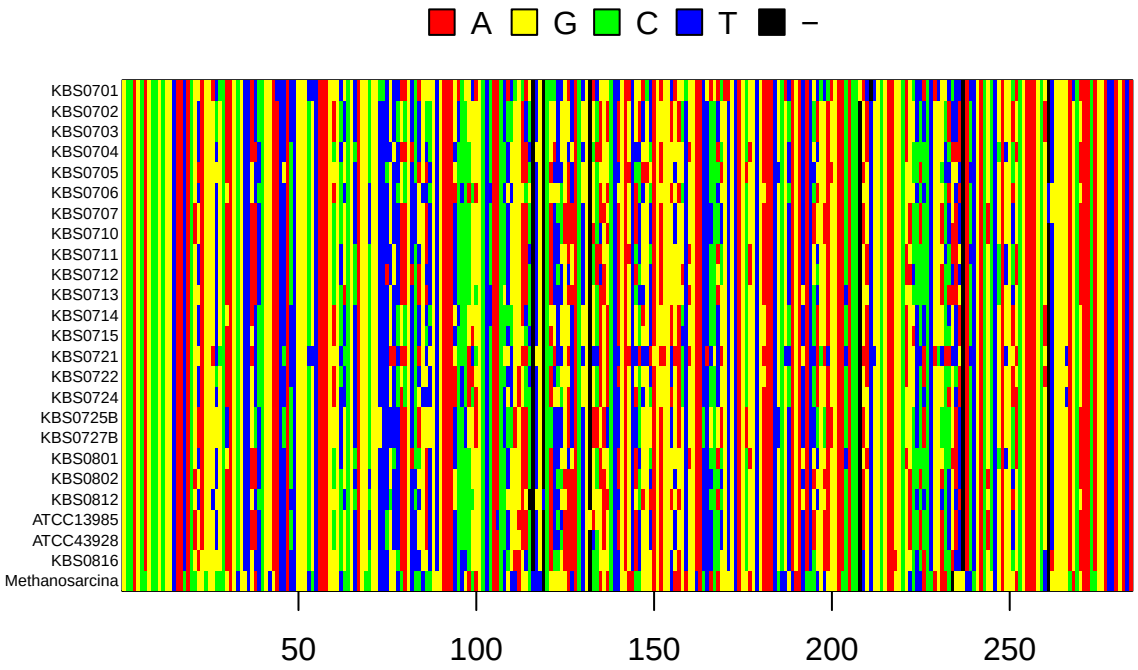
```
#!/bin/bash
#PBS -k o
#PBS -l nodes=1:ppn=8,vmem=100gb,walltime=1:00:00
#PBS -M wrshoema@indiana.edu
#PBS -m abe
#PBS -j oe

module load gcc/4.9.2
module load boost/1.52.0
module load openmpi
module load mothur/1.38.1


cd /N/dc2/projects/Lennon_Sequences/LTDE_Tree

mothur persistence_707_712_721.batch
```

Which calls `persistence_707_712_721.batch`, a batch file containing the following instructions

```
align.seqs(fasta=persistence_707_712_721.fasta, reference=silva.v4.fasta, flip=T, processors=4)
align.seqs(fasta=persistence.fasta, reference=silva.bacteria.rdp.tax, flip=T, processors=4)
summary.seqs(fasta=persistence_707_712_721.align)
screen.seqs(fasta=persistence_707_712_721.align, minlength=200)
filter.seqs(fasta=persistence_707_712_721.good.align, vertical=T, trump=.)
```

3) View alignment

```
read.aln <- read.alignment(file = "data/align/persistence_707_712_721.good.filter.fasta", format = "fas
p.DNAbin <- as.DNAbin(read.aln)
# Identify Base Pair Region of 16S rRNA Gene to Visuzlize (adjust range)
#window.M <- p.DNAbin.M[, 1:200]
# Command to Visusalize Sequence Alignment {ape}
image.DNAbin(p.DNAbin, cex.lab = 0.50)
```



4) Build the tree

We then build the filtered FASTA formatted alignment file using RAxML

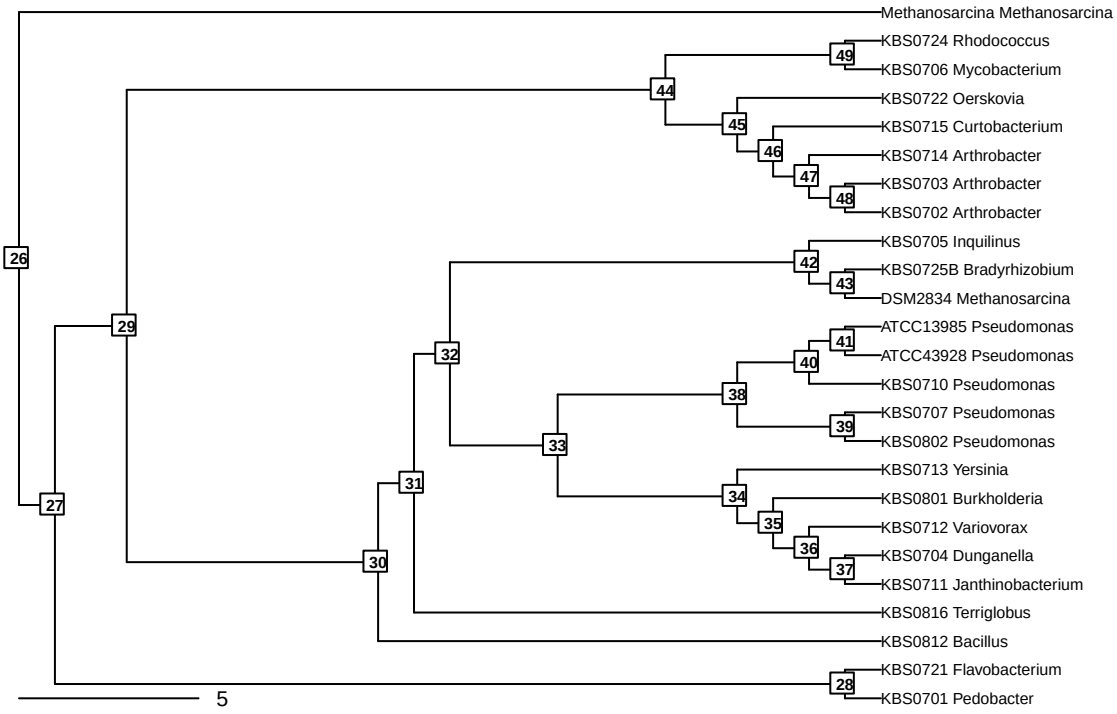
```
raxmlHPC-PTHREADS -T 4 -f a -m GTRGAMMA -p 12345 -x 12345 -o Methanosarcina -# autoMRE \
    -s /Users/WRShoemaker/GitHub/LTDE/phylo/tree/persistence_707_712_721.good.filter.fasta \
    -n T20 -w /Users/WRShoemaker/GitHub/LTDE/phylo/tree

# -T = number of threads
# -f = specifies bootstrapping algorithm with ML generating tree at same time
# -m = substitution model, generalized time reversible gamma
# -p = starts tree randomly
# -x = starts tree randomly
# -o = outgroup (name after fasta entry)
# -#  = determines number of bootstrap replicates
```

```
# -s = aligned fasta file to be analyzed
# -n = name of output file
```

5) Visualize the tree

```
# Read tree
ml.tree <- read.tree("data/tree/RAxML_bipartitionsBranchLabels.T20")

# Identify Outgroup Sequence
outgroup <- match("Methanosarcina", ml.tree$tip.label)

# Root the Tree {ape}
ml.rooted <- root(ml.tree, outgroup, resolve.root = TRUE)

# Load phylo taxonomy data
tax <- read.table("data/traits/persistence.phylo.txt", sep = "\t", header = TRUE)
rownames(tax) <- tax$Code
tax.2 <- tax[ml.tree$tip.label, ]
tax.name <- paste(tax.2$Code, tax.2$Genus)
tax.name[15] <- "DSM2834 Methanosarcina"

# Plot the Rooted Tree{ape}
par(mar = c(1,1,2,1) + 0.1)
plot.phylo(ml.rooted, main = "RAXML Tree",
           "phylogram", use.edge.length = FALSE, direction = "right",
           cex = 0.6, label.offset = 1, show.tip.label = FALSE, x.lim = 30)

tiplabels(tax.name, adj = c(0,0.5), cex = 0.5, frame = "none",
          pch = NULL, thermo = NULL, pie = NULL, piecol = NULL,
          col = NULL, bg = NULL)

nodelabels(ml.rooted$node.label, font = 2, bg = "white", frame = "r", cex = 0.5)
add.scale.bar(cex = 0.7)
```

RAXML Tree


Methanosarcina Methanosarcina
KBS0724 Rhodococcus
KBS0706 Mycobacterium
KBS0722 Oerskovia
KBS0715 Curtobacterium
KBS0714 Arthrobacter
KBS0703 Arthrobacter
KBS0702 Arthrobacter
KBS0705 Inquilinus
KBS0725B Bradyrhizobium
DSM2834 Methanosarcina
ATCC13985 Pseudomonas
ATCC43928 Pseudomonas
KBS0710 Pseudomonas
KBS0707 Pseudomonas
KBS0802 Pseudomonas
KBS0713 Yersinia
KBS0801 Burkholderia
KBS0712 Variovorax
KBS0704 Dunganella
KBS0711 Janthinobacterium
KBS0816 Terriglobus
KBS0812 Bacillus
KBS0721 Flavobacterium
KBS0701 Pedobacter
26
29
44
49
45
46
47
48
42
43
32
41
40
38
39
33
31
34
35
36
37
27
30
28
5