



H3ABioNet

Pan African Bioinformatics Network for H3Africa

16SrRNA Intermediate Bioinformatics Online Course: Int_BT_2019

16S rRNA analysis pipeline Taxonomic classification and alignment using the dada2 pipeline



16SrRNA Intermediate Bioinformatics Online Course:
Int_BT_2019
Imane Allali

Track Reads through DADA2 pipeline

```
getN <- function(x) sum(getUniques(x))
track.nbr.reads <- cbind(out, sapply(dadaF, getN), sapply(dadaR, getN), sapply(merge.reads, getN), rowSums(seqtab
.nochim))

colnames(track.nbr.reads) <- c("input", "filtered", "denoisedF", "denoisedR", "merged", "nonchim")
rownames(track.nbr.reads) <- list.sample.names
head(track.nbr.reads)
```

| ## | | input | filtered | denoisedF | denoisedR | merged | nonchim |
|----|-------|--------|----------|-----------|-----------|--------|---------|
| ## | Dog1 | 118343 | 84815 | 82888 | 83710 | 76306 | 35440 |
| ## | Dog10 | 79342 | 61952 | 60537 | 61130 | 54948 | 24103 |
| ## | Dog15 | 131483 | 91235 | 89781 | 90372 | 83013 | 48592 |
| ## | Dog16 | 114424 | 83564 | 81753 | 82442 | 74533 | 46006 |
| ## | Dog17 | 99610 | 72919 | 71427 | 72229 | 67364 | 42960 |
| ## | Dog2 | 108679 | 79668 | 77366 | 78301 | 69262 | 30129 |

DADA2 workflow

Assign Taxonomy

```
taxa <- assignTaxonomy(seqtab.nochim, paste(MY_HOME, "/dada2_tutorial_dog/RefSeq-RDP16S_v3_May2018.fa.gz", sep='')  
, multithread=TRUE) # change the path
```

RefSeq-RDP16S_v3_May2018.fa.gz

Three most common 16S databases: [Silva](#), [RDP](#) and [GreenGenes](#)

DADA2 workflow

Assign Taxonomy

Maintained:

- Silva version 132, Silva version 128, Silva version 123 (Silva dual-license)
- RDP trainset 16, RDP trainset 14
- GreenGenes version 13.8
- UNITE (use the General Fasta releases)

Contributed:

- RefSeq + RDP (NCBI RefSeq 16S rRNA database supplemented by RDP)
 - Reference files formatted for `assignTaxonomy`
 - Reference files formatted for `assignSpecies`
- GTDB: Genome Taxonomy Database (More info: <http://gtdb.ecogenomic.org/>)
 - Reference files formatted for `assignTaxonomy`
 - Reference files formatted for `assignSpecies`
- HitDB version 1 (Human Intestinal 16S rRNA)
- RDP fungi LSU trainset 11
- Silva Eukaryotic 18S, v132 & v128
- PR2 version 4.7.2+. SEE NOTE BELOW.

Assign Taxonomy

```
taxa.print <- taxa # Removing sequence rownames for display only
rownames(taxa.print) <- NULL
head(taxa.print)
```

```
##      Kingdom      Phylum      Class      Order
## [1,] "Bacteria" "Firmicutes" "Clostridia" "Clostridiales"
## [2,] "Bacteria" "Firmicutes" "Clostridia" "Clostridiales"
## [3,] "Bacteria" "Bacteroidetes" "Bacteroidia" "Bacteroidales"
## [4,] "Bacteria" "Bacteroidetes" "Bacteroidia" "Bacteroidales"
## [5,] "Bacteria" "Firmicutes" "Clostridia" "Clostridiales"
## [6,] "Bacteria" "Firmicutes" "Clostridia" "Clostridiales"
##      Family      Genus
## [1,] "Peptostreptococcaceae" "Clostridium_XI"
## [2,] "Peptostreptococcaceae" "Clostridium_XI"
## [3,] "Prevotellaceae" "Alloprevotella"
## [4,] "Prevotellaceae" "Alloprevotella"
## [5,] "Peptostreptococcaceae" "Clostridium_XI"
## [6,] "Peptostreptococcaceae" "Clostridium_XI"
##      Species
## [1,] "Clostridium_hiranonis(AB023970)"
## [2,] "Clostridium_hiranonis(AB023970)"
## [3,] "Prevotellamassilia_timonensis(NR_144750.1)"
## [4,] "Prevotellamassilia_timonensis(NR_144750.1)"
## [5,] "Clostridium_hiranonis(AB023970)"
## [6,] "Clostridium_hiranonis(AB023970)"
```

DADA2 workflow

Assign Taxonomy

```
write.csv(taxa, file="ASVs_taxonomy.csv")
saveRDS(taxa, "ASVs_taxonomy.rds")
```

| | A | B | C | D | E | F | G | H | I | J | K |
|----|--------|----------|---------------|-------------|---------------|-------------------|----------------|--|---|---|---|
| 1 | | Kingdom | Phylum | Class | Order | Family | Genus | Species | | | |
| 2 | ASV_1 | Bacteria | Firmicutes | Clostridia | Clostridiales | Peptostreptococci | Clostridium | Clostridium_hiranonis(AB023970) | | | |
| 3 | ASV_2 | Bacteria | Firmicutes | Clostridia | Clostridiales | Peptostreptococci | Clostridium | Clostridium_hiranonis(AB023970) | | | |
| 4 | ASV_3 | Bacteria | Bacteroidetes | Bacteroidia | Bacteroidales | Prevotellaceae | Alloprevotella | Prevotellamassilia_timonensis(NR_144750.1) | | | |
| 5 | ASV_4 | Bacteria | Bacteroidetes | Bacteroidia | Bacteroidales | Prevotellaceae | Alloprevotella | Prevotellamassilia_timonensis(NR_144750.1) | | | |
| 6 | ASV_5 | Bacteria | Firmicutes | Clostridia | Clostridiales | Peptostreptococci | Clostridium | Clostridium_hiranonis(AB023970) | | | |
| 7 | ASV_6 | Bacteria | Firmicutes | Clostridia | Clostridiales | Peptostreptococci | Clostridium | Clostridium_hiranonis(AB023970) | | | |
| 8 | ASV_7 | Bacteria | Bacteroidetes | Bacteroidia | Bacteroidales | Prevotellaceae | Alloprevotella | Prevotellamassilia_timonensis(NR_144750.1) | | | |
| 9 | ASV_8 | Bacteria | Bacteroidetes | Bacteroidia | Bacteroidales | Prevotellaceae | Alloprevotella | Prevotellamassilia_timonensis(NR_144750.1) | | | |
| 10 | ASV_9 | Bacteria | Firmicutes | Clostridia | Clostridiales | Peptostreptococci | Clostridium | Clostridium_hiranonis(AB023970) | | | |
| 11 | ASV_10 | Bacteria | Firmicutes | Clostridia | Clostridiales | Peptostreptococci | Clostridium | Clostridium_hiranonis(AB023970) | | | |
| 12 | ASV_11 | Bacteria | Bacteroidetes | Bacteroidia | Bacteroidales | Prevotellaceae | Alloprevotella | Prevotellamassilia_timonensis(NR_144750.1) | | | |
| 13 | ASV_12 | Bacteria | Firmicutes | Clostridia | Clostridiales | Peptostreptococci | Clostridium | Clostridium_hiranonis(AB023970) | | | |
| 14 | ASV_13 | Bacteria | Bacteroidetes | Bacteroidia | Bacteroidales | Prevotellaceae | Alloprevotella | Prevotellamassilia_timonensis(NR_144750.1) | | | |
| 15 | ASV_14 | Bacteria | Firmicutes | Clostridia | Clostridiales | Peptostreptococci | Clostridium | Clostridium_hiranonis(AB023970) | | | |
| 16 | ASV_15 | Bacteria | Bacteroidetes | Bacteroidia | Bacteroidales | Prevotellaceae | Alloprevotella | Prevotellamassilia_timonensis(NR_144750.1) | | | |
| 17 | ASV_16 | Bacteria | Bacteroidetes | Bacteroidia | Bacteroidales | Prevotellaceae | Alloprevotella | Prevotellamassilia_timonensis(NR_144750.1) | | | |
| 18 | ASV_17 | Bacteria | Firmicutes | Clostridia | Clostridiales | Peptostreptococci | Clostridium | Clostridium_hiranonis(AB023970) | | | |
| 19 | ASV_18 | Bacteria | Bacteroidetes | Bacteroidia | Bacteroidales | Prevotellaceae | Alloprevotella | Prevotellamassilia_timonensis(NR_144750.1) | | | |
| 20 | ASV_19 | Bacteria | Firmicutes | Clostridia | Clostridiales | Peptostreptococci | Clostridium | Clostridium_hiranonis(AB023970) | | | |
| 21 | ASV_20 | Bacteria | Bacteroidetes | Bacteroidia | Bacteroidales | Prevotellaceae | Alloprevotella | Prevotellamassilia_timonensis(NR_144750.1) | | | |
| 22 | ASV_21 | Bacteria | Bacteroidetes | Bacteroidia | Bacteroidales | Prevotellaceae | Alloprevotella | Prevotellamassilia_timonensis(NR_144750.1) | | | |
| 23 | ASV_22 | Bacteria | Firmicutes | Clostridia | Clostridiales | Peptostreptococci | Clostridium | Clostridium_hiranonis(AB023970) | | | |
| 24 | ASV_23 | Bacteria | Bacteroidetes | Bacteroidia | Bacteroidales | Prevotellaceae | Alloprevotella | Prevotellamassilia_timonensis(NR_144750.1) | | | |
| 25 | ASV_24 | Bacteria | Bacteroidetes | Bacteroidia | Bacteroidales | Prevotellaceae | Alloprevotella | Prevotellamassilia_timonensis(NR_144750.1) | | | |
| 26 | ASV_25 | Bacteria | Bacteroidetes | Bacteroidia | Bacteroidales | Prevotellaceae | Alloprevotella | Prevotellamassilia_timonensis(NR_144750.1) | | | |

Assign Taxonomy

```
asv_headers <- vector(dim(seqtab.nochim)[2], mode="character")
count.asv.tab <- t(seqtab.nochim)
row.names(count.asv.tab) <- sub(">", "", asv_headers)
write.csv(count.asv.tab, file="ASVs_counts.csv")
saveRDS(count.asv.tab, file="ASVs_counts.rds")
```

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
|----|--------|-------|-------|-------|-------|------|-------|-------|-------|-------|------|-------|-------|------|------|---|
| 1 | Dog1 | Dog10 | Dog15 | Dog16 | Dog17 | Dog2 | Dog22 | Dog23 | Dog24 | Dog29 | Dog3 | Dog30 | Dog31 | Dog8 | Dog9 | |
| 2 | ASV_1 | 0 | 0 | 0 | 0 | 0 | 373 | 1926 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | ASV_2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 955 | 0 | 0 | 921 | 0 | 0 | 0 | 0 |
| 4 | ASV_3 | 242 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1596 | 0 | 0 |
| 5 | ASV_4 | 205 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1625 | 0 | 0 |
| 6 | ASV_5 | 0 | 0 | 0 | 0 | 0 | 276 | 1516 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | ASV_6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 805 | 0 | 0 | 944 | 0 | 0 | 0 | 0 |
| 8 | ASV_7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1747 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | ASV_8 | 223 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1523 | 0 | 0 |
| 10 | ASV_9 | 0 | 0 | 0 | 0 | 0 | 283 | 1453 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 | ASV_10 | 0 | 0 | 0 | 0 | 0 | 277 | 1459 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | ASV_11 | 195 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1536 | 0 | 0 |
| 13 | ASV_12 | 0 | 0 | 0 | 0 | 0 | 322 | 1366 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | ASV_13 | 184 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1479 | 0 | 0 |
| 15 | ASV_14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 806 | 0 | 0 | 857 | 0 | 0 | 0 | 0 |
| 16 | ASV_15 | 170 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1480 | 0 | 0 |
| 17 | ASV_16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1645 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | ASV_17 | 0 | 0 | 0 | 0 | 0 | 270 | 1369 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 19 | ASV_18 | 220 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1373 | 0 | 0 |
| 20 | ASV_19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 772 | 0 | 0 | 798 | 0 | 0 | 0 | 0 |
| 21 | ASV_20 | 174 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1376 | 0 | 0 |
| 22 | ASV_21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1524 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | ASV_22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 743 | 0 | 0 | 720 | 0 | 0 | 0 | 0 |
| 24 | ASV_23 | 188 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1254 | 0 | 0 |
| 25 | ASV_24 | 179 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1260 | 0 | 0 |
| 26 | ASV_25 | 154 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1263 | 0 | 0 |
| 27 | ASV_26 | 167 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1238 | 0 | 0 |
| 28 | ASV_27 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 689 | 0 | 0 | 716 | 0 | 0 | 0 | 0 |

DADA2 workflow

Alignment

```
library(DECIPHER)
```

Tools for curating, analyzing, and manipulating biological sequences.

```
seqs <- getSequences(seqtab.nochim)
names(seqs) <- seqs # This propagates to the tip labels of the tree
alignment <- AlignSeqs(DNAStringSet(seqs), anchor=NA)
```


DADA2 workflow

Construct Phylogenetic Tree

```
library(phangorn)
```

```
phang.align <- phyDat(as(alignment, "matrix"), type="DNA")  
dm <- dist.ml(phang.align)  
treeNJ <- NJ(dm) # Note, tip order != sequence order  
fit = pml(treeNJ, data=phang.align)
```

It constructs a neighbor-joining tree.

1. Change sequence alignment output into **phyDat** structure.
2. Create distance matrix using **dist.ml**.
3. Perform neighbor joining.
4. Perform internal maximum likelihood.

DADA2 workflow

Construct Phylogenetic Tree

```
fitGTR <- update(fit, k=4, inv=0.2)
fitGTR <- optim.pml(fitGTR, model="GTR", optInv=TRUE, optGamma=TRUE,
rearrangement = "stochastic", control = pml.control(trace = 0))
```

It fits a GTR+G+I (Generalized time-reversible with Gamma rate variation) maximum likelihood tree using the neighbor-joining tree as a starting point.

```
saveRDS(fitGTR, "phangorn.tree.RDS")
```