Script R15: CoNet Analysis - Formatting Relative Abundance Files

HANNIGAN GD, GRICE EA, ET AL.

Abstract

This protocol outlines the analysis used to generate input files for CoNet for the phage-bacteria network. Based on methods from the following publication:

Hannigan, Geoffrey D., et al. "The Human Skin Double-Stranded DNA Virome: Topographical and Temporal Diversity, Genetic Enrichment, and Dynamic Associations with the Host Microbiome." *mBio* 6.5 (2015): e01578-15.

Citation: HANNIGAN GD, GRICE EA, ET AL. Script R15: CoNet Analysis - Formatting Relative Abundance Files. **protocols.io**
dx.doi.org/10.17504/protocols.io.ejpbcmn
Published: 10 Mar 2016

Guidelines

sessionInfo()

```
## R version 3.2.0 (2015-04-16)
## Platform: x86_64-apple-darwin13.4.0 (64-bit)
## Running under: OS X 10.10.4 (Yosemite)
## ## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats graphics grDevices utils datasets methods base
##
## loaded via a namespace (and not attached):
## [1] magrittr_1.5    formatR_1.2    tools_3.2.0    htmltools_0.2.6
## [5] yaml_2.1.13    stringi_0.4-1    rmarkdown_0.7    knitr_1.10.5
## [9] stringr_1.0.0    digest_0.6.8    evaluate_0.7
```

Before start

Supplemental information available at:

https://figshare.com/articles/The_Human_Skin_dsDNA_Virome_Topographical_and_Temporal_Diversity_Genetic_Enrichment_and_Dynamic_Associations_with_the_Host_Microbiome/1281248

Protocol

Step 1.

First, we need to read in the metadata associated with the project and format it to remove samples we aren't interested in:

cmd COMMAND

```
meta<-
read.delim("../../IntermediateOutput/Mapping_files/SkinMet_and_Virome_001_metadata.tsv")
meta<-subset(meta, meta$TimePoint != 1)
meta<-subset(meta, !(meta$SubjectID %in% c(2,3,9,11)))
meta<-subset(meta, !(meta$Site_Symbol %in% c("Neg", "Vf", "Ba", "Ph")))
meta<-subset(meta, meta$NexteraXT_SampleID != "NA")
meta<-subset(meta, meta$NexteraXT_Virome_SampleID != "NA")
```

Step 2.

Let's start with the phages. First, we extract only the metadata for the viral samples. Then, we read in the phage species data file and merge it with the viral metadata. Finally, we order the samples by site, timepoint, and subject.

cmd COMMAND

```
virome.sites<-meta[,c("NexteraXT_Virome_SampleID", "Site_Symbol","SubjectID","TimePoint")]

phage<-
read.delim("../../IntermediateOutput/CoNet/phage_species.txt", header=TRUE, row.names=1, st
ringsAsFactors=FALSE)

phage_species<-
merge(virome.sites, t(phage), by.x="NexteraXT_Virome_SampleID", by.y="row.names")
phage_species<-
phage_species[order(phage_species$Site_Symbol, phage_species$TimePoint, phage_species$Subje
ctID),]
phage_samples<-phage_species[c("Site_Symbol","SubjectID","TimePoint")]
```

Step 3.

fter that, we look at the metagenome samples. Again, we extract only the metadata for the metagenome samples, read in the bacterial genus data file and merge it with the metagenome metadata, making sure to keep only the same samples seen in the phage data. Finally, we order the samples by site, timepoint and subject.

cmd COMMAND

```
skinmet.sites<-meta[,c("NexteraXT_SampleID", "Site_Symbol", "SubjectID", "TimePoint")]
skinmet_samples<-
merge(skinmet.sites, phage_samples, by=c("Site_Symbol","SubjectID","TimePoint"))

met_taxa<-
read.delim("../../IntermediateOutput/CoNet/skinmet_taxa.txt", header=TRUE, row.names=1,stri
ngsAsFactors=FALSE)
row.names(met_taxa)<-gsub(row.names(met_taxa),pattern="g__", replacement="")

skinmet_taxa<-
merge(skinmet_samples, t(met_taxa), by.x="NexteraXT_SampleID",by.y="row.names")
skinmet_taxa<-
skinmet_taxa[order(skinmet_taxa$Site_Symbol, skinmet_taxa$TimePoint, skinmet_taxa$SubjectID
),]
```

Step 4.

We need to remove extraneous information from the data files.

cmd COMMAND

```
phage_species$SubjectID<-NULL
phage_species$TimePoint<-NULL
rownames(phage_species)<-phage_species$NexteraXT_Virome_SampleID
phage_species$NexteraXT_Virome_SampleID<-NULL

skinmet_taxa$SubjectID<-NULL
skinmet_taxa$TimePoint<-NULL
rownames(skinmet_taxa)<-skinmet_taxa$NexteraXT_SampleID
skinmet_taxa$NexteraXT_SampleID<-NULL

phage_species$Site_Symbol<-NULL
skinmet_taxa$Site_Symbol<-NULL
```

Step 5.

We transpose the phage species data so that rows are species and columns are samples. We need to remove duplicate taxa from the phage species file.

cmd COMMAND

```
phage_species.t<-as.data.frame(t(phage_species))

dup<-grep("Mycobacter", row.names(phage_species.t), ignore.case=TRUE)
phage_species.t[nrow(phage_species.t)+1,]<- colSums(phage_species.t[dup,])
phage_species.t<-phage_species.t[-dup,]
row.names(phage_species.t)[nrow(phage_species.t)]<-"Mycobacterium_phage"

dup<-grep("Streptomyce", row.names(phage_species.t), ignore.case=TRUE)
phage_species.t[nrow(phage_species.t)+1,]<- colSums(phage_species.t[dup,])
phage_species.t<-phage_species.t[-dup,]
row.names(phage_species.t)[nrow(phage_species.t)]<-"Streptomyces_phage"

dup<-grep("Environmental_Halophage", row.names(phage_species.t), ignore.case=TRUE)
phage_species.t[nrow(phage_species.t)+1,]<- colSums(phage_species.t[dup,])
phage_species.t<-phage_species.t[-dup,]
row.names(phage_species.t)[nrow(phage_species.t)]<-"Environmental_halophage"

dup<-grep("Enterobacteri", row.names(phage_species.t), ignore.case=TRUE)
phage_species.t[nrow(phage_species.t)+1,]<- colSums(phage_species.t[dup,])
phage_species.t<-phage_species.t[-dup,]
row.names(phage_species.t)[nrow(phage_species.t)]<-"Enterobacteria_phage"

dup<-grep("Coryne", row.names(phage_species.t), ignore.case=TRUE)
phage_species.t[nrow(phage_species.t)+1,]<- colSums(phage_species.t[dup,])
phage_species.t<-phage_species.t[-dup,]
row.names(phage_species.t)[nrow(phage_species.t)]<-"Corynebacterium_phage"
```

Step 6.

Then, we need to convert the phage species table and skinmet genus table from counts to relative abundances (from 0 to 1).

cmd COMMAND

```
y<-colSums(phage_species.t)
phage_species_ra.t<-mapply("/", phage_species.t, y)
phage_species_ra.t<- cbind(row.names(phage_species.t), phage_species_ra.t)
colnames(phage_species_ra.t)[1]<-c("Taxa")

skinmet_taxa.t<-as.data.frame(t(skinmet_taxa))
y<-colSums(skinmet_taxa.t)
skinmet_taxa_ra.t<-mapply("/", skinmet_taxa.t, y)
skinmet_taxa_ra.t<- cbind(row.names(skinmet_taxa.t), skinmet_taxa_ra.t)
colnames(skinmet_taxa_ra.t)[1]<-c("Taxa")
```

Step 7.

Lastly, we need to write our data out to files we can use in CoNet.

cmd COMMAND

```
write.table(phage_species_ra.t,"~/Dropbox/Grice:Hanngian:Meisel/Virome_paper_1/Code/IntermediateOutput/CoNet/org_phage_for_conet_test.txt", quote=FALSE, eol="\r\n", sep="\t", row.names=FALSE)

write.table(skinmet_taxa_ra.t,"~/Dropbox/Grice:Hanngian:Meisel/Virome_paper_1/Code/IntermediateOutput/CoNet/org_skinmet_for_conet_test.txt", quote=FALSE, eol="\r\n", sep="\t", row.names=FALSE)
```

NOTES

Geoffrey Hannigan 16 Feb 2016

Note that we recieved a warning in CoNet to remove redundant row (that summed to 0). This referred to the row for Pectobacterium, which we removed manually.