Script R8: Plotting Bacterial Taxonomy from MetaPhlan

HANNIGAN GD, GRICE EA, ET AL.

Abstract

This protocol outlines the analysis used to plot MetaPhlAn taxonomic assignments. 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 R8: Plotting Bacterial Taxonomy from MetaPhlan. **protocols.io** dx.doi.org/10.17504/protocols.io.ejbbcin
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.

Load the libraries needed for analysis.

cmd COMMAND

```
library(ggplot2)
packageVersion("ggplot2")

library(reshape2)
packageVersion("reshape2")

library(plyr)
packageVersion("plyr")

library(RColorBrewer)
packageVersion("RColorBrewer")
```

EXPECTED RESULTS

```
## [1] '1.0.1'



## [1] '1.4.1'



## [1] '1.8.2'



## [1] '1.1.2'
```

Step 2.

Read in the metadata and format it so that it matches the samples we are working with.

cmd COMMAND

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

Step 3.

We only want to look at a certain number of taxa to make the data more visually informative.

cmd COMMAND

```
topTaxa<- function(data, numTaxa){
  if(nrow(data)>numTaxa){
    data$RowSum<-rowSums(data)
    data<-data[order(-data$RowSum),]
    tmp<-data[numTaxa:nrow(data),]
```

```
    data<-data[-c(numTaxa:nrow(data)),]
    other<-colSums(tmp)
    data<-rbind(data,other)
    row.names(data)[nrow(data)]<-"Other"
  }
  data$RowSum<-NULL
  return(data)
}
```

Step 4.

Now we are ready to read in the MetaPhlAn merged output at the genus level and format it for plotting.

cmd COMMAND

```
skinmet_data<-
read.delim("../../IntermediateOutput/MetaPhlAn/skinmet_metaphlan_merged_output_genera.txt",
sep="\t",header=TRUE)
```

Step 5.

Format sample IDs.

cmd COMMAND

```
skinmet_data$ID<-gsub(x=skinmet_data$ID,pattern="g__",replacement="")
colnames(skinmet_data)<-
gsub(x=colnames(skinmet_data),pattern="_R1_trimmed_metaphlan_genera",replacement="")
rownames(skinmet_data)<-skinmet_data[,1]
skinmet_data<-skinmet_data[,-1]
head(skinmet_data[,1:4])
```

EXPECTED RESULTS

##	MG100128	MG100129	MG100130	MG100131
## Abiotrophia	0	0	0	0
## Acetobacteraceae_unclassified	0	0	0	0
Achromobacter	0	0	0	0
Acidaminococcaceae_unclassified	0	0	0	0
Acidovorax	0	0	0	0
Acinetobacter	0	0	0	0

Step 6.

Look at top 10 taxa.

cmd COMMAND

```
skinmet_data<-topTaxa(skinmet_data,10)
taxa_order<-as.vector(row.names(skinmet_data))

skinmet_datat<-as.data.frame(t(skinmet_data))
skinmet_data2<-merge(skinmet_datat,skinmet_metadata,by.x="row.names",by.y="SampleID")
colnames(skinmet_data2)[1]<-"SampleID"
skinmet_datam<-melt(skinmet_data2, id.var=colnames(skinmet_metadata))
skinmet_datam$variable<-factor(skinmet_datam$variable, levels=c(taxa_order))
```

Step 7.

Plot by site symbol and site categories.

cmd COMMAND

```
ggplot(skinmet_datam, aes(x=factor(SampleID), y=value, fill=variable, order=variable))+them
e_bw()+geom_bar(stat = "identity") +theme(axis.text.x=element_text(angle=90),legend.positio
n='right')+ggtitle("MetaPhlAn-
 Top 10 Genera Relative Abundance") +facet_wrap(Site_Symbol~TimePoint,scales="free_x",nrow=
1)+guides(fill = guide_legend(reverse = TRUE))+xlab("Sample")+ylab("Relative Abundance")+gu
ides(fill = guide_legend(reverse = TRUE))+scale_fill_manual(values = c("#e41a1c", "#377eb8"
```

```
, "#33CCCC", "#4daf4a", "#984ea3", "#ff7f00", "#ffff33", "#a65628","#f781bf", "#999999"))
```

EXPECTED RESULTS