Relative Abundance of Bacterial Taxon from Cancer Patient Samples

Hai Liang

12/18/2018

In this project, I will analyze the 16s RNA sequencing result from microbiome samples of cancer patients after immunotherapy treatment. First I need to import the processed sequencing data into R and clean up the column names.

Load data

```
table <- read.table(file="all.good.unique.good.filter.unique.subsample.precluster
.pick.rdp.wang.pick.tax.summary", sep="\t", header=T, stringsAsFactors=F)
```

data table manipulation

```
#cleaning up the names
names(table)<-gsub("X(\\d\\d\\d\\d\\d)", "H\\1", names(table))
names(table)<-gsub("020926", "020916", names(table))
names(table)
```

	taxlevel	rankID	taxon	daughterlevels	total	H80062.BuRUSw.011717	H80062.BuRUSw.102516	H80062.BuRUSw.110816	H80062.BuRUSw.120516	H800
1	0	0	Root	1	1081001	9537	9435	9342	9421	
2	1	0.1	Bacteria	14	1081001	9537	9435	9342	9421	
3	2	0.1.1	Acidobacteria	1	1	0	0	0	0	
4	3	0.1.1.1	Acidobacteria_Gp4	1	1	0	0	0	0	
5	4	0.1.1.1.1	Blastocatella	1	1	0	0	0	0	
6	5	0.1.1.1.1.1	Blastocatella_unclas...	1	1	0	0	0	0	
7	6	0.1.1.1.1.1.1	Blastocatella_unclas...	0	1	0	0	0	0	
8	2	0.1.2	Actinobacteria	1	80714	962	1887	1388	1124	
9	3	0.1.2.1	Actinobacteria	4	80714	962	1887	1388	1124	
10	4	0.1.2.1.1	Actinobacteria_uncl...	1	7	0	0	0	0	
11	5	0.1.2.1.1.1	Actinobacteria_uncl...	1	7	0	0	0	0	
12	6	0.1.2.1.1.1.1	Actinobacteria_uncl...	0	7	0	0	0	0	

In this step I will select the key phyla and genera we are interested in our analysis.

extract the representative taxon

```
attach(table)
#get total counts
sum <- as.numeric(table[taxon=="Bacteria",])

#Key phyla
Actinobacteria <- as.numeric(table[taxon=="Actinobacteria" & taxlevel==2,])
Bacteroidetes <- as.numeric(table[taxon=="Bacteroidetes" & taxlevel==2,])
Firmicutes <- as.numeric(table[taxon=="Firmicutes" & taxlevel==2,])
Proteobacteria <- as.numeric(table[taxon=="Proteobacteria" & taxlevel==2,])
Other_bacteria <- sum - Actinobacteria - Bacteroidetes - Firmicutes - Proteobacteria

#Key genera: Propi, Coryne, Staph, Strep.
Propionibacterium <- as.numeric(table[taxon=="Propionibacterium" & taxlevel==6,])
Corynebacterium <- as.numeric(table[taxon=="Corynebacterium" & taxlevel==6,])
Staphylococcus <- as.numeric(table[taxon=="Staphylococcus" & taxlevel==6,])
Streptococcus <- as.numeric(table[taxon=="Streptococcus" & taxlevel==6,])
```

The sum-up table with selected taxon will be created.

```
#Generate sumup-table
sumup_table <- rbind(Actinobacteria, Propionibacterium, Corynebacterium,
                     Firmicutes, Staphylococcus, Streptococcus,
                     Bacteroidetes, Proteobacteria, Other_bacteria)
colnames(sumup_table) <- colnames(table)
colSums(sumup_table[,6:ncol(sumup_table)])
```

	taxlevel	rankID	taxon	daughterlevels	total	H80062.BuRUSw.011717	H80062.BuRUSw.102516	H80062.BuRUSw.110816	H80062
Actinobacteria	-10	NA	NA	1	75423	957	1878	1385	
Propionibacterium	6	NA	NA	0	745	0	0	0	
Corynebacterium	6	NA	NA	0	4546	5	9	3	
Firmicutes	-10	NA	NA	5	128480	955	798	1555	
Staphylococcus	6	NA	NA	0	4714	3	0	1	
Streptococcus	6	NA	NA	0	667839	4058	4171	3528	
Bacteroidetes	2	NA	NA	4	77184	1588	391	1575	
Proteobacteria	2	NA	NA	6	92911	1405	1862	1008	
Other_bacteria	-7	NA	NA	-2	29159	566	326	287	

I will calculate the percentage of each taxon in total bacteria and create a ratio table.

```
#Make ratio table
ratio_table <- function(df){
  return(data.frame(t(apply(df, 1, function(x){return(unlist(x/colSums(df)))})),
stringsAsFactors=F))
}

newdf <- ratio_table(sumup_table[,6:ncol(sumup_table)])
colSums(newdf)
```

	H80062.BuRUSw.011717	H80062.BuRUSw.102516	H80062.BuRUSw.110816	H80062.BuRUSw.120516	H80063.BuRUSw.112116	H8006
Actinobacteria	0.1003460208	0.1990461049	0.1482551916	0.1183526165	0.0496709173	
Propionibacteri...	0.0000000000	0.0000000000	0.0000000000	0.0002122917	0.0002056767	
Corynebacterium	0.0005242739	0.0009538951	0.0003211304	0.0007430209	0.0011312217	
Firmicutes	0.1001363112	0.0845786963	0.1664525797	0.1093302197	0.0606746195	
Staphylococcus	0.0003145643	0.0000000000	0.0001070435	0.0002122917	0.0004113534	
Streptococcus	0.4255006816	0.4420773715	0.3776493256	0.3421080565	0.8148909914	
Bacteroidetes	0.1665093845	0.0414414414	0.1685934489	0.1048720943	0.0497737557	
Proteobacteria	0.1473209605	0.1973502915	0.1078998073	0.2237554400	0.0208761826	
Other_bacteria	0.0593478033	0.0345521993	0.0307214729	0.1004139688	0.0023652818	

The table will be reshaped with reshape2 package, “Subject” and “Data” columns will be added.

```
library(reshape2)

df.1 <- newdf
df.2 <- cbind(rownames(df.1), rownames(df.1), df.1)
colnames(df.2)[1:2] <- c("Classification", "Tax_order")
df.2$Tax_order <- factor(df.2$Tax_order, levels=c(as.character(rownames(df.1))))
df.2 <- arrange(df.2, Tax_order)

melted <- melt(df.2, id=c("Classification", "Tax_order"))

melted$Subject <- gsub('(H\\d*)\\.\\w*\\.\\d*', '\\1', melted$variable)
melted$Date <- gsub('(H\\d*)\\.\\w*\\.(\\d*)', '\\2', melted$variable)
```

	Classification	Tax_order	variable	value	Subject	Date
1	Actinobacteria	Actinobacteria	H80062.BuRUSw.011717	0.1003460208	H80062	011717
2	Propionibacterium	Propionibacterium	H80062.BuRUSw.011717	0.0000000000	H80062	011717
3	Corynebacterium	Corynebacterium	H80062.BuRUSw.011717	0.0005242739	H80062	011717
4	Firmicutes	Firmicutes	H80062.BuRUSw.011717	0.1001363112	H80062	011717
5	Staphylococcus	Staphylococcus	H80062.BuRUSw.011717	0.0003145643	H80062	011717
6	Streptococcus	Streptococcus	H80062.BuRUSw.011717	0.4255006816	H80062	011717
7	Bacteroidetes	Bacteroidetes	H80062.BuRUSw.011717	0.1665093845	H80062	011717
8	Proteobacteria	Proteobacteria	H80062.BuRUSw.011717	0.1473209605	H80062	011717
9	Other_bacteria	Other_bacteria	H80062.BuRUSw.011717	0.0593478033	H80062	011717
10	Actinobacteria	Actinobacteria	H80062.BuRUSw.102516	0.1990461049	H80062	102516

A subgroup of patients with AntiPDL1 treatment will be selected to build a new table.

```
AntiPDL1_subjects <- c("H80025", "H80011", "H80032", "H80006", "H80010")

AntiPDL1 <- melted[melted$Subject %in% AntiPDL1_subjects,]
AntiPDL1$Subject <- factor(AntiPDL1$Subject, levels=c("H80025", "H80011", "H80032",
"H80006", "H80010"))
```

Filter

	Classification	Tax_order	variable	value	Subject	Date
334	Actinobacteria	Actinobacteria	H80006.BuRUSw.042115	0.0580291971	H80006	042115
335	Propionibacterium	Propionibacterium	H80006.BuRUSw.042115	0.0002433090	H80006	042115
336	Corynebacterium	Corynebacterium	H80006.BuRUSw.042115	0.0004866180	H80006	042115
337	Firmicutes	Firmicutes	H80006.BuRUSw.042115	0.2745742092	H80006	042115
338	Staphylococcus	Staphylococcus	H80006.BuRUSw.042115	0.0000000000	H80006	042115
339	Streptococcus	Streptococcus	H80006.BuRUSw.042115	0.5934306569	H80006	042115
340	Bacteroidetes	Bacteroidetes	H80006.BuRUSw.042115	0.0560827251	H80006	042115
341	Proteobacteria	Proteobacteria	H80006.BuRUSw.042115	0.0034063260	H80006	042115
342	Other_bacteria	Other_bacteria	H80006.BuRUSw.042115	0.0137469586	H80006	042115
343	Actinobacteria	Actinobacteria	H80006.BuRUSw.060215	0.0493255752	H80006	060215
344	Propionibacterium	Propionibacterium	H80006.BuRUSw.060215	0.0003967204	H80006	060215
345	Corynebacterium	Corynebacterium	H80006.BuRUSw.060215	0.0007934409	H80006	060215

Within AntiPDL1 treatment group, there is one patient which showed positive response to the treatment (Responder), and others didn't show response (Non_Responder). Wilcoxon test is used here to study if there is significant difference between Responder and Non_Responder in the "Streptococcus" abundance.

Test result shows there is no difference between these two groups of patients.

statistical analysis

```
#divide subjects into 2 groups
Responder <- AntiPDL1$value[AntiPDL1$Subject=="H80025" & AntiPDL1$Classification==
"Streptococcus"]

Non_Responder <- AntiPDL1$value[AntiPDL1$Subject!="H80025" & AntiPDL1$Classificati
on=="Streptococcus"]

wilcox.test(Responder, Non_Responder)
```

```
##
##  Wilcoxon rank sum test
##
## data:  Responder and Non_Responder
## W = 31, p-value = 0.6235
## alternative hypothesis: true location shift is not equal to 0
```

Drawing a barplot with taxon abundance versus sampling time, grouped by patient IDs.

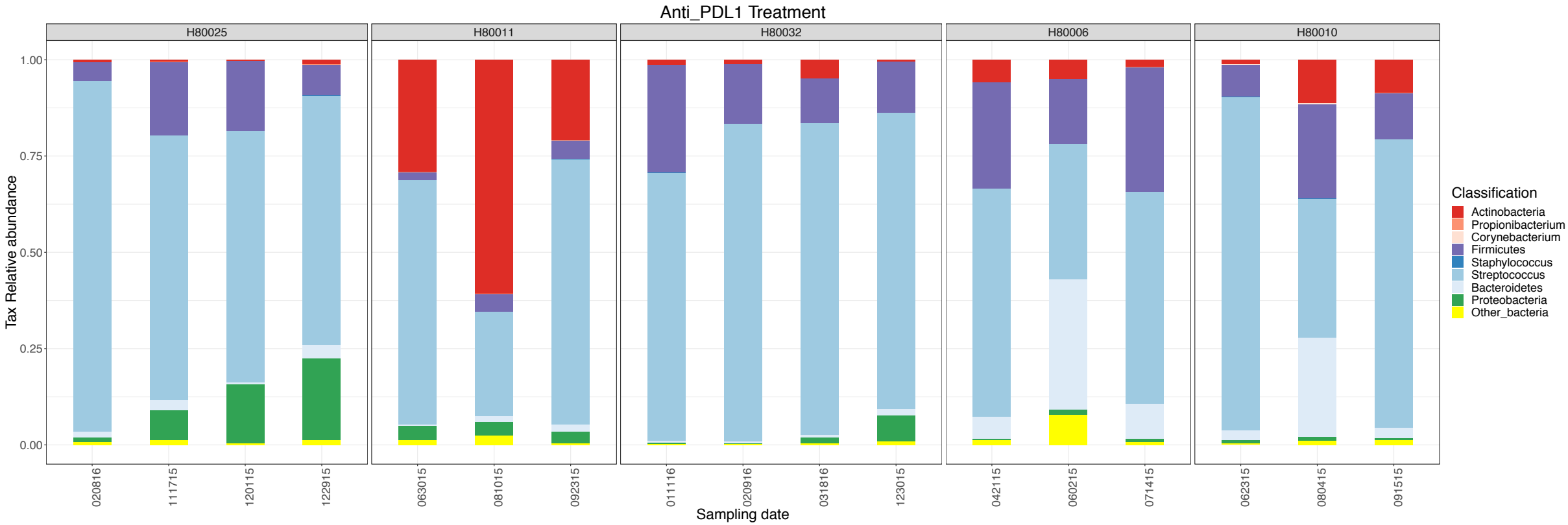
Drawing barplot

```
#library required; RColorBrewer, ggplot2
library(ggplot2)
library(RColorBrewer)

color_pall <- c(rev(brewer.pal(3, "Reds")), brewer.pal(3, "Purples")[3],  rev(brewer.pal
(3, "Blues")), brewer.pal(3, "Greens")[3], "yellow")

pdf(file="AntiPDL1_Tx_Tax.pdf", height=10, width=30)
ggplot(AntiPDL1) +
  aes(x=Date, y=value, fill=Tax_order, order=Tax_order) +
  geom_bar(stat="identity", position="fill", width=0.5) + theme_bw() +
  scale_fill_manual("Classification", labels=df.2$Tax_order, values=color_pall) +
  labs(title="Anti_PDL1 Treatment",
       x="Sampling date",
       y="Tax Relative abundance") +
  theme(plot.title = element_text(hjust = 0.5), text = element_text(size=20), axis.text.
x=element_text(angle=90, hjust=1, vjust=1)) +
  facet_grid(~Subject, scales='free_x', space='free')
dev.off()
```

Output barplot file in PDF format.



Thank you!