

Processing Stool Sample Whole Genome Sequencing Data

Ankan Choudhury

2025-02-19

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(biomformat)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.3      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(data.table)
```

```
##
## Attaching package: 'data.table'
##
## The following objects are masked from 'package:lubridate':
##
##   hour, isoweek, mday, minute, month, quarter, second, wday, week,
##   yday, year
##
## The following objects are masked from 'package:dplyr':
##
##   between, first, last
##
## The following object is masked from 'package:purrr':
##
##   transpose
```

```
library(fs)
library(phyloseq)
library(writexl)
library(phylostratr)
library(ape)
```

```
##
## Attaching package: 'ape'
##
## The following object is masked from 'package:dplyr':
##
##     where
```

```
library(taxize)
```

```
##
## Attaching package: 'taxize'
##
## The following object is masked from 'package:phylostratr':
##
##     children
```

```
Kaijubioms<-dir("Kaiju Taxonomy Abundance",pattern="*biom")
Kaijutaxa<-dir("Kaiju Taxonomy Abundance",pattern="kaiju_taxonomy*")
for (biom in Kaijubioms){
  for(taxa in Kaijutaxa){
    if(substring(biom,21,nchar(biom)-5)==substring(taxa,16,nchar(taxa)-4))
      name=substring(biom,21,nchar(biom)-5)
    OTUtable<-as.data.frame(as.matrix(biom_data(read_biom(paste("Kaiju Taxonomy Abundance",biom,sep="/"),
      rownames_to_column("Feature.ID")
    taxatable<-read.table(paste("Kaiju Taxonomy Abundance",taxa,sep="/"),sep='\t',header = TRUE)
    Abundance<-inner_join(OTUtable,taxatable)%>%
      select(-"Feature.ID")%>%
      relocate(Taxon)%>%
      mutate(Taxon=str_replace_all(Taxon,".__(.*)","."))%>%
      separate(Taxon,c("Domain","Phylum","Class","Order","Family","Genus","Species"),sep=";")%>%
      mutate(Phylum=if_else(is.na(Phylum),Domain,Phylum),
        Class=if_else(is.na(Class),Phylum,Class),
        Order=if_else(is.na(Order),Class,Order),
        Family=if_else(is.na(Family),Order,Family),
        Genus=if_else(is.na(Genus),Family,Genus),
        Species=if_else(is.na(Species),Genus,Species))
    write_xlsx(Abundance,paste("Kaiju Taxonomy Abundance/kaiju_abun-",name,".xlsx",sep=""))
  }
}
```

```
## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'
```

```
## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'
```

```
head(select(as.data.frame(OTUtable),c(1:10)))
```

```
## Feature.ID 1-MDcol002-5-2 11-MDcol002-label-12-on-bottle
## 1 0 107567 128946
## 2 100886 1063 1096
## 3 102148 0 0
## 4 10239 297 344
## 5 1042156 754 1061
## 6 105841 837 1093
## 14-MDcol003-Stool-1B-Baseline-BSB-2-4-1-4 15-MDcol004-3-1-pretreat
## 1 92867 164303
## 2 568 856
## 3 0 734
## 4 2207 84670
## 5 2083 2220
## 6 1655 938
## 17-MDcol006-3-1A 19-MDcol006-3-2A 21-MDcol006-2-pre-chemo-A 23-MDcol033-1A
## 1 66642 70074 166933 74818
## 2 0 0 1157 0
## 3 0 0 0 0
## 4 204 383 520 34634
## 5 0 0 0 622
## 6 455 861 3895 3628
## 25-MDcol036-1A
## 1 104008
## 2 872
## 3 0
## 4 515
## 5 1147
## 6 6639
```

```
head(taxatable)
```

```
## Feature.ID
## 1 0
## 2 100886
## 3 102148
## 4 10239
## 5 1042156
## 6 105841
##
## 1
```

d__belong to a (non-viral) speci

```
## 2 d__Bacteria;p__Bacillota;c__Erysipelotrichia;o__Erysipelotrichales;f__Coprobacillaceae;g__Cateniba
## 3     d__Bacteria;p__Bacillota;c__Erysipelotrichia;o__Erysipelotrichales;f__Erysipelotrichaceae;g__S
## 4
## 5         d__Bacteria;p__Bacillota;c__Clostridia;o__Eubacteriales;f__Clostridiaceae;g__C
## 6         d__Bacteria;p__Bacillota;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__
```

```
head(select(as.data.frame(Abundance),c(1:10)))
```

```
##                                     Domain
## 1 belong to a (non-viral) species with less than 0.01% of all reads
## 2                                     Bacteria
## 3                                     Bacteria
## 4                                     Viruses
## 5                                     Bacteria
## 6                                     Bacteria
##                                     Phylum
## 1 belong to a (non-viral) species with less than 0.01% of all reads
## 2                                     Bacillota
## 3                                     Bacillota
## 4                                     Viruses
## 5                                     Bacillota
## 6                                     Bacillota
##                                     Class
## 1 belong to a (non-viral) species with less than 0.01% of all reads
## 2                                     Erysipelotrichia
## 3                                     Erysipelotrichia
## 4                                     Viruses
## 5                                     Clostridia
## 6                                     Clostridia
##                                     Order
## 1 belong to a (non-viral) species with less than 0.01% of all reads
## 2                                     Erysipelotrichales
## 3                                     Erysipelotrichales
## 4                                     Viruses
## 5                                     Eubacteriales
## 6                                     Lachnospirales
##                                     Family
## 1 belong to a (non-viral) species with less than 0.01% of all reads
## 2                                     Coprobacillaceae
## 3                                     Erysipelotrichaceae
## 4                                     Viruses
## 5                                     Clostridiaceae
## 6                                     Lachnospiraceae
##                                     Genus
## 1 belong to a (non-viral) species with less than 0.01% of all reads
## 2                                     Catenibacterium
## 3                                     Solobacterium
## 4                                     Viruses
## 5                                     Clostridium
## 6                                     Anaerostipes
##                                     Species
## 1 belong to a (non-viral) species with less than 0.01% of all reads
## 2                                     Catenibacterium mitsuokai
## 3                                     Solobacterium moorei
```

```
## 4                                     Viruses
## 5                                Clostridium sp. SY8519
## 6                                Anaerostipes caccae
## 1-MDcol002-5-2 11-MDcol002-label-12-on-bottle
## 1                107567                128946
## 2                1063                1096
## 3                 0                0
## 4                297                344
## 5                754                1061
## 6                837                1093
## 14-MDCOL003-Stool-1B-Baseline-BSB-2-4-1-4
## 1                92867
## 2                568
## 3                 0
## 4                2207
## 5                2083
## 6                1655
```

```
Kaijubioms<-dir("Kaiju Taxonomy Abundance",pattern="*biom")
Kaijutaxa<-dir("Kaiju Taxonomy Abundance",pattern="kaiju_taxonomy*")
dir.create("PhyloseqOutputs/KaijuPhyloseq")
for (biom in Kaijubioms){
  for(taxa in Kaijutaxa){
    if(substring(biom,21,nchar(biom)-5)==substring(taxa,16,nchar(taxa)-4))
      name=substring(biom,21,nchar(biom)-5)
    OTUtable<-as.matrix(biom_data(read_biom(paste("Kaiju Taxonomy Abundance",biom,sep="/"))))
    taxatable<-read.table(paste("Kaiju Taxonomy Abundance",taxa,sep="/"),sep='\t',header = TRUE)%>%
      mutate(Taxon=str_replace_all(Taxon,"._",""))%>%
      separate(Taxon,c("Domain","Phylum","Class","Order","Family","Genus","Species"),sep=";")%>%
      mutate(Phylum=if_else(is.na(Phylum),Domain,Phylum),
             Class=if_else(is.na(Class),Phylum,Class),
             Order=if_else(is.na(Order),Class,Order),
             Family=if_else(is.na(Family),Order,Family),
             Genus=if_else(is.na(Genus),Family,Genus),
             Species=if_else(is.na(Species),Genus,Species))%>%
      column_to_rownames("Feature.ID")%>%
      as.matrix()
    physeq <- phyloseq(otu_table(OTUtable, taxa_are_rows = TRUE), tax_table(taxatable))
    saveRDS(physeq,paste("PhyloseqOutputs/KaijuPhyloseq/phyloseq_kaiju-",name,".rds",sep=""))
  }
}

summary(physeq)
```

```
##   Length   Class   Mode
##      1 phyloseq    S4
```

```
physeq
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 621 taxa and 37 samples ]
## tax_table()   Taxonomy Table:    [ 621 taxa by 7 taxonomic ranks ]
```

```

brackentables<-dir(path="BrackenOutputs/Do",pattern="table-bracken*")
brackentaxonomy<-dir(path="BrackenOutputs/Do",pattern="taxonomy-bracken*")

dir.create("Kraken-Bracken Taxonomy Abundance")
for(each in brackentables){
  name<-substring(each,15)
  for(each2 in brackentaxonomy){
    name2<-substring(each2,18)
    if(name==name2){
      table<-as.data.frame(as.matrix(biom_data(read_biom(paste("BrackenOutputs/Do",each,"feature-table.b
      rownames_to_column("Feature.ID")
      taxonomy<-read.table(paste("BrackenOutputs/Do",each2,"taxonomy.tsv",sep='/'),sep="\t",header = TRUE)
      mutate(Feature.ID=as.character(Feature.ID))
      taxatable<-inner_join(table,taxonomy)%>%
        select(-"Feature.ID")%>%
        relocate(Taxon)%>%
        mutate(Taxon=str_replace_all(Taxon,".__(.*)","."))%>%
        separate(Taxon,c("Domain","Kingdom","Phylum","Class","Order","Family","Genus","Species"),sep=";")
      write_xlsx(taxatable,paste("Kraken-Bracken Taxonomy Abundance/Kraken-Bracken_abun-",name,".xlsx",sep=""
    }
  }
}

```

```

## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'
## Joining with 'by = join_by(Feature.ID)'

```

```

head(select(as.data.frame(table),c(1:10)))

```

```

##   Feature.ID 1-MDcol002-5-2 11-MDcol002-label-12-on-bottle
## 1      100886              5                             8
## 2      102148              0                             0
## 3      102684              0                             0
## 4      1042156             0                             0
## 5      104609              0                             0
## 6      1051631             0                             0
##   14-MDCOL003-Stool-1B-Baseline-BSB-2-4-1-4 15-MDcol004-3-1-pretreat
## 1                      0                      1
## 2                      0                      2
## 3                      0                      0
## 4                      0                      3
## 5                      0                      0
## 6                      1                      0
##   17-MDCOL006-3-1A 19-MDCOL006-3-2A 21-MDCOL006-2-pre-chemo-A 23-MDcol033-1A
## 1                9                5                12                5
## 2                0                0                0                0
## 3                0                0                0                0
## 4                0                0                1                1
## 5                0                0                0                0
## 6                0                0                0                0
##   25-MDcol036-1A

```

```
## 1      1
## 2      0
## 3      0
## 4      0
## 5      0
## 6      0
```

```
head(as.data.frame(taxonomy))
```

```
##   Feature.ID
## 1      40520
## 2      418240
## 3      2479767
## 4      89014
## 5      33035
## 6      1532
##
## 1      d__Bacteria;k__Bacteria;p__Bacillota;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__B
## 2      d__Bacteria;k__Bacteria;p__Bacillota;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Blau
## 3 d__Bacteria;k__Bacteria;p__Bacillota;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Blautia
## 4      d__Bacteria;k__Bacteria;p__Bacillota;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__L
## 5      d__Bacteria;k__Bacteria;p__Bacillota;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Blau
## 6      d__Bacteria;k__Bacteria;p__Bacillota;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Blaut
```

```
head(select(as.data.frame(taxatable),c(1:10)))
```

```
##   Domain      Kingdom      Phylum      Class
## 1 Bacteria      Bacteria      Bacillota      Erysipelotrichia
## 2 Bacteria      Bacteria      Bacillota      Erysipelotrichia
## 3 Bacteria      Bacteria      Bacillota      Bacilli
## 4 Bacteria      Bacteria      Bacillota      Clostridia
## 5 Bacteria      Bacteria Pseudomonadota Gammaproteobacteria
## 6 Viruses Heunggongvirae      Uroviricota      Caudoviricetes
##
##           Order      Family
## 1      Erysipelotrichales      Coprobacillaceae
## 2      Erysipelotrichales      Erysipelotrichaceae
## 3      Lactobacillales      Streptococcaceae
## 4      Eubacteriales      Clostridiaceae
## 5      Vibrionales      Vibrionaceae
## 6 containing Aliceevansviridae      Aliceevansviridae
##
##           Genus      Species
## 1      Catenibacterium      Catenibacterium mitsuokai
## 2      Solobacterium      Solobacterium moorei
## 3      Streptococcus      Streptococcus infantarius
## 4      Clostridium      Clostridium sp. SY8519
## 5      Vibrio      Vibrio penaeicida
## 6 containing Streptococcus phage YMC-2011 Streptococcus phage YMC-2011
## 1-MDcol002-5-2 11-MDcol002-label-12-on-bottle
## 1      5      8
## 2      0      0
## 3      0      0
## 4      0      0
## 5      0      0
## 6      0      0
```

```

brackentables<-dir(path="BrackenOutputs/Do",pattern="table-bracken*")
brackentaxonomy<-dir(path="BrackenOutputs/Do",pattern="taxonomy-bracken*")
brackentree<-dir(path="BrackenOutputs/Do",pattern="bracken-reports*")

dir.create("PhyloseqOutputs/BrackenPhyloseq")
counter=0
for(each in brackentables){
  name<-substring(each,15)
  table<-as.matrix(biom_data(read_biom(paste("BrackenOutputs/Do",each,"feature-table.biom",sep='/'))))
  taxonomy<-read.table(paste("BrackenOutputs/Do/taxonomy-bracken-",name,"/taxonomy.tsv",sep=''),sep="\t")
  mutate(Taxon=str_replace_all(Taxon,"_",""))%>%
  separate(Taxon,c("Domain","Phylum","Class","Order","Family","Genus","Species"),sep=";")%>%
  mutate(Phylum=if_else(is.na(Phylum),Domain,Phylum),
         Class=if_else(is.na(Class),Phylum,Class),
         Order=if_else(is.na(Order),Class,Order),
         Family=if_else(is.na(Family),Order,Family),
         Genus=if_else(is.na(Genus),Family,Genus),
         Species=if_else(is.na(Species),Genus,Species))%>%
  column_to_rownames("Feature.ID")%>%
  as.matrix()
  tree<-read_tree(paste("BrackenOutputs/Do/bracken-reports-",name,"/output.tree",sep=''))
  counter=counter+1
  print(counter)
  physeqBracken<-phyloseq(otu_table(table,taxa_are_rows = TRUE),tax_table(taxonomy))
  physeqBracken<-merge_phyloseq(physeqBracken,tree)
  saveRDS(physeqBracken,paste("PhyloseqOutputs/BrackenPhyloseq/phyloseq_bracken-",name,".rds",sep=""))
  print(name)
}

```

```
## [1] 1
```

```
## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'
```

```
## Also defined by 'phylostratr'
```

```
## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'
```

```
## Also defined by 'phylostratr'
```

```
## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'
```

```
## Also defined by 'phylostratr'
```

```
## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'
```

```
## Also defined by 'phylostratr'
```

```
## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'
```

```
## Also defined by 'phylostratr'
```



```

## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'

## Also defined by 'phylostratr'

## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'

## Also defined by 'phylostratr'

## [1] "demux-paired"
## [1] 2

## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'
## Also defined by 'phylostratr'

## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'

## Also defined by 'phylostratr'

## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'

## Also defined by 'phylostratr'

## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'

## Also defined by 'phylostratr'

## [1] "filtered"
## [1] 3

## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'
## Also defined by 'phylostratr'

## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'

## Also defined by 'phylostratr'

## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'

## Also defined by 'phylostratr'

## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'

## Also defined by 'phylostratr'

## [1] "trimmed"
## [1] 4

## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'
## Also defined by 'phylostratr'

```

```
## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'

## Also defined by 'phylostratr'

## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'

## Also defined by 'phylostratr'

## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'

## Also defined by 'phylostratr'

## [1] "trimmed-filtered"
## [1] 5

## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'
## Also defined by 'phylostratr'

## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'

## Also defined by 'phylostratr'

## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'

## Also defined by 'phylostratr'

## Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'

## Also defined by 'phylostratr'

## [1] "trimmed-filtered-contigs"
```

```
head(select(as.data.frame(table),c(1:10)))
```

```
##          1-MDcol002-5-2 11-MDcol002-label-12-on-bottle
## 100886                5                             8
## 102148                0                             0
## 102684                0                             0
## 1042156               0                             0
## 104609                0                             0
## 1051631               0                             0
##          14-MDCOL003-Stool-1B-Baseline-BSB-2-4-1-4 15-MDcol004-3-1-pretreat
## 100886                0                             1
## 102148                0                             2
## 102684                0                             0
## 1042156               0                             3
## 104609                0                             0
## 1051631               1                             0
##          17-MDCOL006-3-1A 19-MDCOL006-3-2A 21-MDCOL006-2-pre-chemo-A
## 100886                9                 5                 12
```

```
## 102148      0      0      0
## 102684      0      0      0
## 1042156     0      0      1
## 104609      0      0      0
## 1051631     0      0      0
##      23-MDcol033-1A 25-MDcol036-1A 27-MDcol025-1B
## 100886      5      1      2
## 102148      0      0      0
## 102684      0      0      0
## 1042156     1      0      3
## 104609      0      0      0
## 1051631     0      0      0
```

```
head(as.data.frame(taxonomy))
```

```
##      Domain  Phylum    Class    Order      Family      Genus
## 40520  Bacteria Bacteria Bacillota Clostridia Lachnospirales Lachnospiraceae
## 418240 Bacteria Bacteria Bacillota Clostridia Lachnospirales Lachnospiraceae
## 2479767 Bacteria Bacteria Bacillota Clostridia Lachnospirales Lachnospiraceae
## 89014  Bacteria Bacteria Bacillota Clostridia Lachnospirales Lachnospiraceae
## 33035  Bacteria Bacteria Bacillota Clostridia Lachnospirales Lachnospiraceae
## 1532   Bacteria Bacteria Bacillota Clostridia Lachnospirales Lachnospiraceae
##      Species
## 40520  Blautia
## 418240 Blautia
## 2479767 Blautia
## 89014  Blautia
## 33035  Blautia
## 1532   Blautia
```

```
summary(physeqBracken)
```

```
##      Length      Class      Mode
##           1 phyloseq          S4
```

```
physeqBracken
```

```
## phyloseq-class experiment-level object
## otu_table() OTU Table: [ 812 taxa and 37 samples ]
## tax_table() Taxonomy Table: [ 812 taxa by 7 taxonomic ranks ]
## phy_tree() Phylogenetic Tree: [ 812 tips and 234 internal nodes ]
```

```
MetaphlanOutputs<-dir("MetaphlanOutput")
dir.create("PhyloseqOutputs/MetaphlanPhyloseq")
MetaphlanTree<-read_tree("20201013_mpav30_speciesMod.nwk")
new_names <- MetaphlanTree$tip.label
new_names<-new_names%>%
  str_replace_all("s_","")
MetaphlanTree$tip.label<-new_names
for(folder in MetaphlanOutputs){
  Alltaxonomy<-NULL
```

```

AllOTU<-NULL
for(file in list.files(paste("MetaphlanOutput",folder,sep="/"))){
  content<-readLines(paste("MetaphlanOutput",folder,file,sep="/"))[-c(1:4)]
  id<-substring(basename(file),1,nchar(basename(file))-11)
  totalreads<-readLines(paste("MetaphlanOutput",folder,file,sep="/"))[3]
  totalreads<-as.numeric(gsub("^#(\\d+).*", "\\1", totalreads))
  table<-read.table(text=content,header=FALSE,sep="\t")%>%
    select(-4)%>%
    mutate(V1=str_replace_all(V1,"._",""),
           V1 = str_replace_all(V1, "\\|", ";"),
           reads=round(V3/100*totalreads)) %>%
    separate(V1,c("Kingdom","Phylum","Class","Order","Family","Genus","Species"),sep=";")%>%
    separate(V2,c("KID","PID","CID","OID","FID","GID","SID"),sep="\\|")%>%
    unique()%>%
    filter(!(is.na(Species)))%>%
    mutate(PID=if_else(PID==" " | is.na(PID),KID,PID),
           CID=if_else(CID==" " | is.na(CID),PID,CID),
           OID=if_else(OID==" " | is.na(OID),CID,OID),
           FID=if_else(FID==" " | is.na(FID),OID,FID),
           GID=if_else(GID==" " | is.na(GID),FID,GID),
           SID=if_else(SID==" " | is.na(SID),GID,SID))
  taxonomy<-table[,c(1:7)]
  OTU<-table[,c(7,16)]%>%
    unique()%>%
    mutate(Sample=id)
  Alltaxonomy<-rbind(Alltaxonomy,taxonomy)
  AllOTU<-rbind(AllOTU,OTU)
  print(paste("Processed: MetaphlanOutput",folder,file,sep="/"))
}

taxa_table<-Alltaxonomy%>%
  arrange("Species")%>%
  unique()
rownames(taxa_table)<-taxa_table$Species
otu_table<-AllOTU%>%
  group_by(Sample,Species)%>%
  summarize(reads=sum(reads))%>%
  pivot_wider(id_cols = "Species",values_from = "reads",names_from = "Sample",values_fill = 0)%>%
  column_to_rownames("Species")
physeqMetaphlan<-phyloseq(otu_table(as.matrix(otu_table),taxa_are_rows = TRUE),tax_table(as.matrix(taxa_table)))
physeqMetaphlan<-merge_phyloseq(physeqMetaphlan,phy_tree(MetaphlanTree))
saveRDS(physeqMetaphlan,paste("PhyloseqOutputs/MetaphlanPhyloseq/phyloseq_Metaphlan-",substring(folder,1,15),".rds"))
}

## [1] "Processed: MetaphlanOutput/filtered-paired/1-MDcol002-5-2.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/11-MDcol002-label-12-on-bottle.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/14-MDCOL003-Stool-1B-Baseline-BSB-2-4-1-4.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/15-MDcol004-3-1-pretreat.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/17-MDCOL006-3-1A.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/19-MDCOL006-3-2A.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/21-MDCOL006-2-pre-chemo-A.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/23-MDcol033-1A.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/25-MDcol036-1A.output.txt"

```

```

## [1] "Processed: MetaphlanOutput/filtered-paired/27-MDcol025-1B.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/28-7A-1-1.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/29-7A.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/3-MDcol002-5-1.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/30-8B.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/31-8B.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/33-6-2.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/36-40-1B.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/38-S-Swanson.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/40-MDcol002-7-43.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/42-51-1B.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/45-MDCOL005-1B.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/46-MDCOL015-1a.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/48-MDCOL039-1a.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/5-MDCOL055-Stool-1A-Baseline-BSB-2-4-1-2.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/50-MDCOL027-1a.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/52-MDCOL021-1b.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/54-MDCOL028-1b.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/60-MDcol006-4-1A.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/63-MDcol003-3-2.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/65-MDCOL003-Post-Chemo-Collection-Cycle-1-A.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/66-MDCOL004-1A.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/68-MDCOL003-3-1-Before-Chemo-2-3A.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/7-MDCOL058-Stool-1A-Baseline-BSB-2-4-1-3.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/70-MDCOL004-3-2-Post-Chemo-A.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/72-MDCOL003-4-1-4A-Before-Chemo-6.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/75-MDCOL053-1-1-1-Unknown.output.txt"
## [1] "Processed: MetaphlanOutput/filtered-paired/9-MDCOL062-Stool-1A-Baseline-BSB-2-4-1-3.output.txt"

## 'summarise()' has grouped output by 'Sample'. You can override using the
## '.groups' argument.
## Found more than one class "phylo" in cache; using the first, from namespace
## 'phyloseq'
## Also defined by 'phylostratr'
## Found more than one class "phylo" in cache; using the first, from namespace
## 'phyloseq'
## Also defined by 'phylostratr'
## Found more than one class "phylo" in cache; using the first, from namespace
## 'phyloseq'
## Also defined by 'phylostratr'
## Found more than one class "phylo" in cache; using the first, from namespace
## 'phyloseq'
## Also defined by 'phylostratr'

## [1] "Processed: MetaphlanOutput/raw-paired/1-MDcol002-5-2.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/11-MDcol002-label-12-on-bottle.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/14-MDCOL003-Stool-1B-Baseline-BSB-2-4-1-4.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/15-MDcol004-3-1-pretreat.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/17-MDCOL006-3-1A.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/19-MDCOL006-3-2A.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/21-MDCOL006-2-pre-chemo-A.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/23-MDcol033-1A.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/25-MDcol036-1A.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/27-MDcol025-1B.output.txt"

```

```

## [1] "Processed: MetaphlanOutput/raw-paired/28-7A-1-1.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/29-7A.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/3-MDcol002-5-1.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/30-8B.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/31-8B.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/33-6-2.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/36-40-1B.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/38-S-Swanson.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/40-MDcol002-7-43.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/42-51-1B.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/45-MDCOL005-1B.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/46-MDCOL015-1a.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/48-MDCOL039-1a.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/5-MDCOL055-Stool-1A-Baseline-BSB-2-4-1-2.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/50-MDCOL027-1a.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/52-MDCOL021-1b.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/54-MDCOL028-1b.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/60-MDcol006-4-1A.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/63-MDcol003-3-2.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/65-MDCOL003-Post-Chemo-Collection-Cycle-1-A.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/66-MDCOL004-1A.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/68-MDCOL003-3-1-Before-Chemo-2-3A.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/7-MDCOL058-Stool-1A-Baseline-BSB-2-4-1-3.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/70-MDCOL004-3-2-Post-Chemo-A.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/72-MDCOL003-4-1-4A-Before-Chemo-6.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/75-MDCOL053-1-1-1-Unknown.output.txt"
## [1] "Processed: MetaphlanOutput/raw-paired/9-MDCOL062-Stool-1A-Baseline-BSB-2-4-1-3.output.txt"

## 'summarise()' has grouped output by 'Sample'. You can override using the
## '.groups' argument.
## Found more than one class "phylo" in cache; using the first, from namespace
## 'phyloseq'
## Also defined by 'phylostratr'
## Found more than one class "phylo" in cache; using the first, from namespace
## 'phyloseq'
## Also defined by 'phylostratr'
## Found more than one class "phylo" in cache; using the first, from namespace
## 'phyloseq'
## Also defined by 'phylostratr'
## Found more than one class "phylo" in cache; using the first, from namespace
## 'phyloseq'
## Also defined by 'phylostratr'

## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/1-MDcol002-5-2.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/11-MDcol002-label-12-on-bottle.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/14-MDCOL003-Stool-1B-Baseline-BSB-2-4-1-4.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/15-MDcol004-3-1-pretreat.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/17-MDCOL006-3-1A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/19-MDCOL006-3-2A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/21-MDCOL006-2-pre-chemo-A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/23-MDcol033-1A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/25-MDcol036-1A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/27-MDcol025-1B.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/28-7A-1-1.output.txt"

```

```

## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/29-7A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/3-MDcol002-5-1.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/30-8B.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/31-8B.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/33-6-2.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/36-40-1B.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/38-S-Swanson.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/40-MDCol002-7-43.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/42-51-1B.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/45-MDCOL005-1B.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/46-MDCOL015-1a.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/48-MDCOL039-1a.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/5-MDCOL055-Stool-1A-Baseline-BSB-2-4-1-2.outp
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/50-MDCOL027-1a.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/52-MDCOL021-1b.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/54-MDCOL028-1b.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/60-MDcol006-4-1A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/63-MDcol003-3-2.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/65-MDCOL003-Post-Chemo-Collection-Cycle-1-A.
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/66-MDCOL004-1A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/68-MDCOL003-3-1-Before-Chemo-2-3A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/7-MDCOL058-Stool-1A-Baseline-BSB-2-4-1-3.outp
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/70-MDCOL004-3-2-Post-Chemo-A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/72-MDCOL003-4-1-4A-Before-Chemo-6.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/75-MDCOL053-1-1-1-Unknown.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-filtered-paired/9-MDCOL062-Stool-1A-Baseline-BSB-2-4-1-3.outp

## 'summarise()' has grouped output by 'Sample'. You can override using the
## '.groups' argument.
## Found more than one class "phylo" in cache; using the first, from namespace
## 'phyloseq'
## Also defined by 'phylostratr'
## Found more than one class "phylo" in cache; using the first, from namespace
## 'phyloseq'
## Also defined by 'phylostratr'
## Found more than one class "phylo" in cache; using the first, from namespace
## 'phyloseq'
## Also defined by 'phylostratr'
## Found more than one class "phylo" in cache; using the first, from namespace
## 'phyloseq'
## Also defined by 'phylostratr'

## [1] "Processed: MetaphlanOutput/trimmed-paired/1-MDcol002-5-2.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/11-MDcol002-label-12-on-bottle.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/14-MDCOL003-Stool-1B-Baseline-BSB-2-4-1-4.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/15-MDcol004-3-1-pretreat.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/17-MDCOL006-3-1A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/19-MDCOL006-3-2A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/21-MDCOL006-2-pre-chemo-A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/23-MDcol033-1A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/25-MDcol036-1A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/27-MDcol025-1B.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/28-7A-1-1.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/29-7A.output.txt"

```

```
## [1] "Processed: MetaphlanOutput/trimmed-paired/3-MDcol002-5-1.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/30-8B.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/31-8B.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/33-6-2.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/36-40-1B.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/38-S-Swanson.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/40-MDCol002-7-43.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/42-51-1B.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/45-MDCOL005-1B.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/46-MDCOL015-1a.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/48-MDCOL039-1a.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/5-MDCOL055-Stool-1A-Baseline-BSB-2-4-1-2.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/50-MDCOL027-1a.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/52-MDCOL021-1b.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/54-MDCOL028-1b.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/60-MDcol006-4-1A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/63-MDcol003-3-2.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/65-MDCOL003-Post-Chemo-Collection-Cycle-1-A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/66-MDCOL004-1A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/68-MDCOL003-3-1-Before-Chemo-2-3A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/7-MDCOL058-Stool-1A-Baseline-BSB-2-4-1-3.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/70-MDCOL004-3-2-Post-Chemo-A.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/72-MDCOL003-4-1-4A-Before-Chemo-6.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/75-MDCOL053-1-1-1-Unknown.output.txt"
## [1] "Processed: MetaphlanOutput/trimmed-paired/9-MDCOL062-Stool-1A-Baseline-BSB-2-4-1-3.output.txt"
```

```
## 'summarise()' has grouped output by 'Sample'. You can override using the
## '.groups' argument.
## Found more than one class "phylo" in cache; using the first, from namespace
## 'phyloseq'
## Also defined by 'phylostratr'
## Found more than one class "phylo" in cache; using the first, from namespace
## 'phyloseq'
## Also defined by 'phylostratr'
## Found more than one class "phylo" in cache; using the first, from namespace
## 'phyloseq'
## Also defined by 'phylostratr'
## Found more than one class "phylo" in cache; using the first, from namespace
## 'phyloseq'
## Also defined by 'phylostratr'
```

```
head(taxa_table)
```

| | Kingdom | Phylum | Class | Order |
|--------------------------|----------------|---------------|-------------|----------------------|
| ## Phocaeicola_vulgatus | Bacteria | Bacteroidetes | Bacteroidia | Bacteroidales |
| ## GGB9464_SGB14857 | Bacteria | Firmicutes | CFGB3002 | OFGB3002 |
| ## Bacteroides_uniformis | Bacteria | Bacteroidetes | Bacteroidia | Bacteroidales |
| ## Blautia_wexlerae | Bacteria | Firmicutes | Clostridia | Eubacteriales |
| ## Bacteroides_ovatus | Bacteria | Bacteroidetes | Bacteroidia | Bacteroidales |
| ## Vescimonas_coprocola | Bacteria | Firmicutes | Clostridia | Eubacteriales |
| ## | | Family | Genus | Species |
| ## Phocaeicola_vulgatus | Bacteroidaceae | Phocaeicola | Phocaeicola | Phocaeicola_vulgatus |
| ## GGB9464_SGB14857 | | FGB3002 | GGB9464 | GGB9464_SGB14857 |


```
## Bacteroides_uniformis    Bacteroidaceae Bacteroides Bacteroides_uniformis
## Blautia_wexlerae        Lachnospiraceae Blautia Blautia_wexlerae
## Bacteroides_ovatus      Bacteroidaceae Bacteroides Bacteroides_ovatus
## Vescimonas_coprocola    Oscillospiraceae Vescimonas Vescimonas_coprocola
```

```
head(select(as.data.frame(otu_table),c(1:10)))
```

```
##                                1-MDcol002-5-2 11-MDcol002-label-12-on-bottle
## Adlercreutzia_equolifaciens          74820          146653
## Agathobaculum_butyriciproducens       814            0
## Alistipes_communis                    24064            0
## Amedibacillus_dolichus                11375          12634
## Anaerobutyricum_soehngenii           68201          85794
## Anaerostipes_hadrus                   31247          13956
##                                14-MDCOL003-Stool-1B-Baseline-BSB-2-4-1-4
## Adlercreutzia_equolifaciens              0
## Agathobaculum_butyriciproducens          0
## Alistipes_communis                      0
## Amedibacillus_dolichus                  0
## Anaerobutyricum_soehngenii              0
## Anaerostipes_hadrus                     0
##                                15-MDcol004-3-1-pretreat 17-MDCOL006-3-1A
## Adlercreutzia_equolifaciens             51300          288837
## Agathobaculum_butyriciproducens         3473            0
## Alistipes_communis                     36052            0
## Amedibacillus_dolichus                  0            0
## Anaerobutyricum_soehngenii             37622            0
## Anaerostipes_hadrus                     65840            0
##                                19-MDCOL006-3-2A 21-MDCOL006-2-pre-chemo-A
## Adlercreutzia_equolifaciens             668954          493108
## Agathobaculum_butyriciproducens          0            0
## Alistipes_communis                      0            0
## Amedibacillus_dolichus                  0            0
## Anaerobutyricum_soehngenii              0            0
## Anaerostipes_hadrus                     3079            0
##                                23-MDcol033-1A 25-MDcol036-1A 27-MDcol025-1B
## Adlercreutzia_equolifaciens              0            0          33002
## Agathobaculum_butyriciproducens          0            0          58308
## Alistipes_communis                      0            0            0
## Amedibacillus_dolichus                  0            0            0
## Anaerobutyricum_soehngenii              0            0          2980
## Anaerostipes_hadrus                     0            0          323549
```

```
HumannOutput<-dir("HumannOutput")
dir.create("PhyloseqOutputs/HumannPhyloseq")
for (folder in HumannOutput){
  AllCoverage<-data.frame()
  AllAbundance<-data.frame()
  counter=1
  for(file in list.files(paste("HumannOutput",folder,sep="/"),pattern="*_pathabundance.tsv")){
    id<-substring(basename(file),1,nchar(basename(file))-18)
    abundance<-read.delim(paste("HumannOutput",folder,file,sep="/"),sep='\t',header = TRUE,comment.char
    df<-colnames(abundance)%>%
```

```

    str_replace_all("X\\.\\.\\.","")%>%
    str_replace_all("X","")%>%
    str_replace_all("\\\\.","_")%>%
    str_replace_all("_Abundance","")
  colnames(abundance)<-df
  coverage<-read.delim(paste("HumannOutput/",folder,"/",id,"_pathcoverage.tsv",sep=""),sep='\t',header=
  df<-colnames(coverage)%>%
  str_replace_all("X\\.\\.\\.","")%>%
  str_replace_all("X","")%>%
  str_replace_all("\\\\.","-")%>%
  str_replace_all("_Coverage","")
  colnames(coverage)<-df
  if (counter==1){
    AllAbundance<-abundance}
  else{AllAbundance<-left_join(AllAbundance,abundance)}
  if(counter==1){
    AllCoverage<-coverage}
  else{AllCoverage<-left_join(AllCoverage,coverage)}
  counter=counter+1
  print(paste("Processed",folder,":",id))
}

AllAbundance<-AllAbundance%>%
  column_to_rownames("Pathway")%>%
  as.matrix()
AllCoverage<-AllCoverage%>%
  column_to_rownames("Pathway")%>%
  as.matrix()
physeqcoverage<-phyloseq(otu_table(AllCoverage,taxa_are_rows=TRUE))
physeqabundance<-phyloseq(otu_table(AllAbundance,taxa_are_rows=TRUE))
saveRDS(physeqabundance,paste("PhyloseqOutputs/HumannPhyloseq/phyloseq_humann-",folder,"_pathwayabund",
saveRDS(physeqcoverage,paste("PhyloseqOutputs/HumannPhyloseq/phyloseq_humann-",folder,"_pathwaycovera
}

```

```
## [1] "Processed raw : 1-MDcol002-5-2"
```

```
## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'
```

```
## [1] "Processed raw : 11-MDCol002-label-12-on-bottle"
```

```
## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'
```

```
## [1] "Processed raw : 14-MDCOL003-Stool-1B-Baseline-BSB-2-4-1-4"
```

```
## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'
```

```
## [1] "Processed raw : 15-MDcol004-3-1-pretreat"
```

```
## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'
```

```

## [1] "Processed raw : 17-MDCOL006-3-1A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 19-MDCOL006-3-2A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 21-MDCOL006-2-pre-chemo-A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 23-MDcol033-1A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 25-MDcol036-1A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 27-MDcol025-1B"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 28-7A-1-1"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 29-7A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 3-MDcol002-5-1"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 30-8B"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

```

```

## [1] "Processed raw : 31-8B"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 33-6-2"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 36-40-1B"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 38-S-Swanson"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 40-MDCol002-7-43"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 42-51-1B"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 45-MDCOL005-1B"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 46-MDCOL015-1a"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 48-MDCOL039-1a"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 5-MDCOL055-Stool-1A-Baseline-BSB-2-4-1-2"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

```

```

## [1] "Processed raw : 50-MDCOL027-1a"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 52-MDCOL021-1b"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 54-MDCOL028-1b"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 60-MDcol006-4-1A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 63-MDcol003-3-2"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 65-MDCOL003-Post-Chemo-Collection-Cycle-1-A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 66-MDCOL004-1A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 68-MDCOL003-3-1-Before-Chemo-2-3A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 7-MDCOL058-Stool-1A-Baseline-BSB-2-4-1-3"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 70-MDCOL004-3-2-Post-Chemo-A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

```

```

## [1] "Processed raw : 72-MDCOL003-4-1-4A-Before-Chemo-6"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 75-MDCOL053-1-1-1-Unknown"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed raw : 9-MDCOL062-Stool-1A-Baseline-BSB-2-4-1-3"
## [1] "Processed trimmed-filtered : 1-MDcol002-5-2"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 11-MDcol002-label-12-on-bottle"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 14-MDCOL003-Stool-1B-Baseline-BSB-2-4-1-4"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 15-MDcol004-3-1-pretreat"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 17-MDCOL006-3-1A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 19-MDCOL006-3-2A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 21-MDCOL006-2-pre-chemo-A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 23-MDcol033-1A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

```

```

## [1] "Processed trimmed-filtered : 25-MDcol036-1A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 27-MDcol025-1B"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 28-7A-1-1"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 29-7A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 3-MDcol002-5-1"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 30-8B"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 31-8B"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 33-6-2"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 36-40-1B"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 38-S-Swanson"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

```

```

## [1] "Processed trimmed-filtered : 40-MDCol002-7-43"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 42-51-1B"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 45-MDCOL005-1B"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 46-MDCOL015-1a"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 48-MDCOL039-1a"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 5-MDCOL055-Stool-1A-Baseline-BSB-2-4-1-2"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 50-MDCOL027-1a"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 52-MDCOL021-1b"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 54-MDCOL028-1b"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 60-MDcol006-4-1A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

```



```

## [1] "Processed trimmed-filtered : 63-MDcol003-3-2"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 65-MDCOL003-Post-Chemo-Collection-Cycle-1-A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 66-MDCOL004-1A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 68-MDCOL003-3-1-Before-Chemo-2-3A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 7-MDCOL058-Stool-1A-Baseline-BSB-2-4-1-3"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 70-MDCOL004-3-2-Post-Chemo-A"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 72-MDCOL003-4-1-4A-Before-Chemo-6"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 75-MDCOL053-1-1-1-Unknown"

## Joining with 'by = join_by(Pathway)'
## Joining with 'by = join_by(Pathway)'

## [1] "Processed trimmed-filtered : 9-MDCOL062-Stool-1A-Baseline-BSB-2-4-1-3"

for (folder in HumannOutput){
  AllRPK<-data.frame()
  counter=1
  for(file in list.files(paste("HumannOutput",folder,sep="/"),pattern="*_genefamilies.tsv")){
    id<-substring(basename(file),1,nchar(basename(file))-17)
    RPK<-read.delim(paste("HumannOutput",folder,file,sep="/"),sep='\t',header = TRUE,comment.char = "")
    df<-colnames(RPK)%>%

```

```

    str_replace_all("X\\.\\.\\.","")%>%
    str_replace_all("X","")%>%
    str_replace_all("\\\\.\\.","-")%>%
    str_replace_all("_Abundance-RPKs","")
  colnames(RPK)<-df
  if (counter==1){
    AllRPK<-RPK}
  else{AllRPK<-left_join(AllRPK,RPK)}
  counter=counter+1
  print(paste("Processed",folder,":",id))
}

AllRPK<-AllRPK%>%
  column_to_rownames("Gene-Family")%>%
  as.matrix()
physeqRPK<-phyloseq(otu_table(AllRPK,taxa_are_rows=TRUE))
saveRDS(physeqRPK,paste("PhyloseqOutputs/HumannPhyloseq/phyloseq_humann-",folder,"_genefamilies.rds",
})

```

```

## [1] "Processed raw : 1-MDcol002-5-2"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 11-MDcol002-label-12-on-bottle"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 14-MDCOL003-Stool-1B-Baseline-BSB-2-4-1-4"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 15-MDcol004-3-1-pretreat"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 17-MDCOL006-3-1A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 19-MDCOL006-3-2A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 21-MDCOL006-2-pre-chemo-A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 23-MDcol033-1A"

```

```

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 25-MDcol036-1A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 27-MDcol025-1B"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 28-7A-1-1"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 29-7A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 3-MDcol002-5-1"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 30-8B"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 31-8B"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 33-6-2"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 36-40-1B"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 38-S-Swanson"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 40-MDcol002-7-43"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 42-51-1B"

```

```

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 45-MDCOL005-1B"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 46-MDCOL015-1a"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 48-MDCOL039-1a"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 5-MDCOL055-Stool-1A-Baseline-BSB-2-4-1-2"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 50-MDCOL027-1a"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 52-MDCOL021-1b"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 54-MDCOL028-1b"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 60-MDcol006-4-1A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 63-MDcol003-3-2"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 65-MDCOL003-Post-Chemo-Collection-Cycle-1-A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 66-MDCOL004-1A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 68-MDCOL003-3-1-Before-Chemo-2-3A"

```

```

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 7-MDCOL058-Stool-1A-Baseline-BSB-2-4-1-3"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 70-MDCOL004-3-2-Post-Chemo-A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 72-MDCOL003-4-1-4A-Before-Chemo-6"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 75-MDCOL053-1-1-1-Unknown"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed raw : 9-MDCOL062-Stool-1A-Baseline-BSB-2-4-1-3"
## [1] "Processed trimmed-filtered : 1-MDcol002-5-2"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 11-MDCol002-label-12-on-bottle"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 14-MDCOL003-Stool-1B-Baseline-BSB-2-4-1-4"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 15-MDcol004-3-1-pretreat"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 17-MDCOL006-3-1A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 19-MDCOL006-3-2A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 21-MDCOL006-2-pre-chemo-A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 23-MDcol033-1A"

```

```

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 25-MDcol036-1A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 27-MDcol025-1B"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 28-7A-1-1"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 29-7A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 3-MDcol002-5-1"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 30-8B"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 31-8B"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 33-6-2"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 36-40-1B"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 38-S-Swanson"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 40-MDcol002-7-43"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 42-51-1B"

```

```

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 45-MDCOL005-1B"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 46-MDCOL015-1a"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 48-MDCOL039-1a"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 5-MDCOL055-Stool-1A-Baseline-BSB-2-4-1-2"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 50-MDCOL027-1a"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 52-MDCOL021-1b"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 54-MDCOL028-1b"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 60-MDcol006-4-1A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 63-MDcol003-3-2"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 65-MDCOL003-Post-Chemo-Collection-Cycle-1-A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 66-MDCOL004-1A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 68-MDCOL003-3-1-Before-Chemo-2-3A"

```

```
## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 7-MDCOL058-Stool-1A-Baseline-BSB-2-4-1-3"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 70-MDCOL004-3-2-Post-Chemo-A"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 72-MDCOL003-4-1-4A-Before-Chemo-6"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 75-MDCOL053-1-1-1-Unknown"

## Joining with 'by = join_by('Gene-Family')'

## [1] "Processed trimmed-filtered : 9-MDCOL062-Stool-1A-Baseline-BSB-2-4-1-3"
```

```
head(abundance)
```

```
##                               Pathway
## 1                               UNMAPPED
## 2                               UNINTEGRATED
## 3 UNINTEGRATED|g__Prevotella.s__Prevotella_copri
## 4                               UNINTEGRATED|unclassified
## 5 UNINTEGRATED|g__Alistipes.s__Alistipes_finegoldii
## 6 UNINTEGRATED|g__Blautia.s__Blautia_obeum
## 9_MDCOL062_Stool_1A_Baseline_BSB_2_4_1_3
## 1                               706374.82
## 2                               1990573.07
## 3                               280605.01
## 4                               211052.96
## 5                               161814.73
## 6                               77940.21
```

```
head(select(as.data.frame(AllAbundance),c(1:10)))
```

```
##                               1_MDcol002_5_2
## UNMAPPED                               778748.2
## UNINTEGRATED                          2481983.1
## UNINTEGRATED|g__Bacteroides.s__Bacteroides_vulgatus 917463.5
## UNINTEGRATED|g__Bacteroides.s__Bacteroides_uniformis 464620.0
## UNINTEGRATED|g__Bacteroides.s__Bacteroides_ovatus   354114.6
## UNINTEGRATED|unclassified              147854.1
##                               11_MDCol002_label_12_on_bottle
## UNMAPPED                               812515.6
## UNINTEGRATED                          2338562.7
## UNINTEGRATED|g__Bacteroides.s__Bacteroides_vulgatus 740989.9
```


| | |
|---|---|
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_uniformis | 205282.2 |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_ovatus | 248232.9 |
| ## UNINTEGRATED unclassified | 229298.6 |
| ## | 14_MDCOL003_Stool_1B_Baseline_BSB_2_4_1_4 |
| ## UNMAPPED | 721273.4 |
| ## UNINTEGRATED | 3179209.0 |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_vulgatus | NA |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_uniformis | NA |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_ovatus | NA |
| ## UNINTEGRATED unclassified | 304960.8 |
| ## | 15_MDcol004_3_1_pretreat |
| ## UNMAPPED | 853278.4 |
| ## UNINTEGRATED | 2607555.3 |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_vulgatus | 187785.6 |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_uniformis | 193568.8 |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_ovatus | 27733.6 |
| ## UNINTEGRATED unclassified | 406532.7 |
| ## | 17_MDCOL006_3_1A |
| ## UNMAPPED | 640092.1 |
| ## UNINTEGRATED | 2365706.0 |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_vulgatus | NA |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_uniformis | 255956.6 |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_ovatus | 486901.1 |
| ## UNINTEGRATED unclassified | 176732.7 |
| ## | 19_MDCOL006_3_2A |
| ## UNMAPPED | 728047.75 |
| ## UNINTEGRATED | 2412350.76 |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_vulgatus | NA |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_uniformis | 84115.85 |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_ovatus | 78634.97 |
| ## UNINTEGRATED unclassified | 178803.49 |
| ## | 21_MDCOL006_2_pre_chemo_A |
| ## UNMAPPED | 1170228.6 |
| ## UNINTEGRATED | 4808488.1 |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_vulgatus | NA |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_uniformis | 240857.1 |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_ovatus | 828564.4 |
| ## UNINTEGRATED unclassified | 532576.9 |
| ## | 23_MDcol033_1A |
| ## UNMAPPED | 691496.1 |
| ## UNINTEGRATED | 2425535.6 |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_vulgatus | NA |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_uniformis | 159044.4 |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_ovatus | 178257.6 |
| ## UNINTEGRATED unclassified | 141054.7 |
| ## | 25_MDcol036_1A |
| ## UNMAPPED | 1068527.81 |
| ## UNINTEGRATED | 4715990.30 |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_vulgatus | 665352.47 |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_uniformis | NA |
| ## UNINTEGRATED g__Bacteroides.s__Bacteroides_ovatus | 85896.49 |
| ## UNINTEGRATED unclassified | 245014.75 |
| ## | 27_MDcol025_1B |
| ## UNMAPPED | 945989.87 |

```
## UNINTEGRATED 3843549.24
## UNINTEGRATED|g__Bacteroides.s__Bacteroides_vulgatus 313605.33
## UNINTEGRATED|g__Bacteroides.s__Bacteroides_uniformis 271299.80
## UNINTEGRATED|g__Bacteroides.s__Bacteroides_ovatus 59771.61
## UNINTEGRATED|unclassified 191873.16
```

```
head(coverage)
```

```
## Pathway
## 1 UNMAPPED
## 2 UNINTEGRATED
## 3 UNINTEGRATED|g__Agathobaculum.s__Agathobaculum_butyriciproducens
## 4 UNINTEGRATED|g__Akkermansia.s__Akkermansia_muciniphila
## 5 UNINTEGRATED|g__Alistipes.s__Alistipes_finegoldii
## 6 UNINTEGRATED|g__Alistipes.s__Alistipes_onderdonkii
## 9-MDCOL062-Stool-1A-Baseline-BSB-2-4-1-3
## 1 1
## 2 1
## 3 1
## 4 1
## 5 1
## 6 1
```

```
head(select(as.data.frame(AllCoverage),c(1:10)))
```

```
## 1-MDcol002-5-2
## UNMAPPED 1
## UNINTEGRATED 1
## UNINTEGRATED|g__Adlercreutzia.s__Adlercreutzia_equolifaciens 1
## UNINTEGRATED|g__Anaerostipes.s__Anaerostipes_hadrus 1
## UNINTEGRATED|g__Asaccharobacter.s__Asaccharobacter_celatus 1
## UNINTEGRATED|g__Bacteroides.s__Bacteroides_ovatus 1
## 11-MDCol002-label-12-on-bottle
## UNMAPPED 1
## UNINTEGRATED 1
## UNINTEGRATED|g__Adlercreutzia.s__Adlercreutzia_equolifaciens 1
## UNINTEGRATED|g__Anaerostipes.s__Anaerostipes_hadrus 1
## UNINTEGRATED|g__Asaccharobacter.s__Asaccharobacter_celatus 1
## UNINTEGRATED|g__Bacteroides.s__Bacteroides_ovatus 1
## 14-MDCOL003-Stool-1B-Baseline-BSB-2-4-1
## UNMAPPED
## UNINTEGRATED
## UNINTEGRATED|g__Adlercreutzia.s__Adlercreutzia_equolifaciens
## UNINTEGRATED|g__Anaerostipes.s__Anaerostipes_hadrus
## UNINTEGRATED|g__Asaccharobacter.s__Asaccharobacter_celatus
## UNINTEGRATED|g__Bacteroides.s__Bacteroides_ovatus
## 15-MDcol004-3-1-pretreat
## UNMAPPED 1
## UNINTEGRATED 1
## UNINTEGRATED|g__Adlercreutzia.s__Adlercreutzia_equolifaciens 1
## UNINTEGRATED|g__Anaerostipes.s__Anaerostipes_hadrus 1
## UNINTEGRATED|g__Asaccharobacter.s__Asaccharobacter_celatus 1
## UNINTEGRATED|g__Bacteroides.s__Bacteroides_ovatus 1
```

```

## 17-MDCOL006-3-1A
## UNMAPPED 1
## UNINTEGRATED 1
## UNINTEGRATED|g__Adlercreutzia.s__Adlercreutzia_equolifaciens 1
## UNINTEGRATED|g__Anaerostipes.s__Anaerostipes_hadrus NA
## UNINTEGRATED|g__Asaccharobacter.s__Asaccharobacter_celatus 1
## UNINTEGRATED|g__Bacteroides.s__Bacteroides_ovatus 1
## 19-MDCOL006-3-2A
## UNMAPPED 1
## UNINTEGRATED 1
## UNINTEGRATED|g__Adlercreutzia.s__Adlercreutzia_equolifaciens 1
## UNINTEGRATED|g__Anaerostipes.s__Anaerostipes_hadrus NA
## UNINTEGRATED|g__Asaccharobacter.s__Asaccharobacter_celatus 1
## UNINTEGRATED|g__Bacteroides.s__Bacteroides_ovatus 1
## 21-MDCOL006-2-pre-chemo-A
## UNMAPPED 1
## UNINTEGRATED 1
## UNINTEGRATED|g__Adlercreutzia.s__Adlercreutzia_equolifaciens 1
## UNINTEGRATED|g__Anaerostipes.s__Anaerostipes_hadrus NA
## UNINTEGRATED|g__Asaccharobacter.s__Asaccharobacter_celatus 1
## UNINTEGRATED|g__Bacteroides.s__Bacteroides_ovatus 1
## 23-MDcol033-1A
## UNMAPPED 1
## UNINTEGRATED 1
## UNINTEGRATED|g__Adlercreutzia.s__Adlercreutzia_equolifaciens NA
## UNINTEGRATED|g__Anaerostipes.s__Anaerostipes_hadrus NA
## UNINTEGRATED|g__Asaccharobacter.s__Asaccharobacter_celatus NA
## UNINTEGRATED|g__Bacteroides.s__Bacteroides_ovatus 1
## 25-MDcol036-1A
## UNMAPPED 1
## UNINTEGRATED 1
## UNINTEGRATED|g__Adlercreutzia.s__Adlercreutzia_equolifaciens NA
## UNINTEGRATED|g__Anaerostipes.s__Anaerostipes_hadrus NA
## UNINTEGRATED|g__Asaccharobacter.s__Asaccharobacter_celatus NA
## UNINTEGRATED|g__Bacteroides.s__Bacteroides_ovatus 1
## 27-MDcol025-1B
## UNMAPPED 1
## UNINTEGRATED 1
## UNINTEGRATED|g__Adlercreutzia.s__Adlercreutzia_equolifaciens 1
## UNINTEGRATED|g__Anaerostipes.s__Anaerostipes_hadrus 1
## UNINTEGRATED|g__Asaccharobacter.s__Asaccharobacter_celatus 1
## UNINTEGRATED|g__Bacteroides.s__Bacteroides_ovatus 1

```

head(RPK)

```

## Gene-Family
## 1 UNMAPPED
## 2 UniRef90_A5ZYV3
## 3 UniRef90_A5ZYV3|g__Blautia.s__Blautia_obeum
## 4 UniRef90_U3JBQ2
## 5 UniRef90_U3JBQ2|g__Eggerthella.s__Eggerthella_lenta
## 6 UniRef90_A0A174PA28
## 9-MDCOL062-Stool-1A-Baseline-BSB-2-4-1-3
## 1 1987609.00

```

```
## 2 19139.59
## 3 19139.59
## 4 15087.70
## 5 15087.70
## 6 13552.05
```

```
head(select(as.data.frame(AllRPK),c(1:10)))
```

```
## 1-MDcol002-5-2
## UNMAPPED 2364890.000
## UniRef90_A0A174NA57 55247.258
## UniRef90_A0A174NA57|g__Bacteroides.s__Bacteroides_uniformis 55247.258
## UniRef90_A0A395VRV3 49394.382
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_ovatus 25032.358
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_vulgatus 8720.539
## 11-MDcol002-label-12-on-bottle
## UNMAPPED 2374486.000
## UniRef90_A0A174NA57 43582.089
## UniRef90_A0A174NA57|g__Bacteroides.s__Bacteroides_uniformis 43582.089
## UniRef90_A0A395VRV3 31559.493
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_ovatus 7825.126
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_vulgatus 5414.094
## 14-MDCOL003-Stool-1B-Baseline-BSB-2-4-1-
## UNMAPPED 1.685473e+0
## UniRef90_A0A174NA57 NA
## UniRef90_A0A174NA57|g__Bacteroides.s__Bacteroides_uniformis NA
## UniRef90_A0A395VRV3 1.546558e+0
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_ovatus NA
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_vulgatus NA
## 15-MDcol004-3-1-pretreat
## UNMAPPED 2462301.0000
## UniRef90_A0A174NA57 12391.8543
## UniRef90_A0A174NA57|g__Bacteroides.s__Bacteroides_uniformis 12391.8543
## UniRef90_A0A395VRV3 6052.5065
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_ovatus 2456.2799
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_vulgatus 823.7582
## 17-MDCOL006-3-1A
## UNMAPPED 1570111.000
## UniRef90_A0A174NA57 NA
## UniRef90_A0A174NA57|g__Bacteroides.s__Bacteroides_uniformis NA
## UniRef90_A0A395VRV3 5697.658
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_ovatus NA
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_vulgatus NA
## 19-MDCOL006-3-2A
## UNMAPPED 1791726.000
## UniRef90_A0A174NA57 NA
## UniRef90_A0A174NA57|g__Bacteroides.s__Bacteroides_uniformis NA
## UniRef90_A0A395VRV3 629.729
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_ovatus NA
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_vulgatus NA
## 21-MDCOL006-2-pre-chemo-A
## UNMAPPED 2.798869e+06
## UniRef90_A0A174NA57 NA
## UniRef90_A0A174NA57|g__Bacteroides.s__Bacteroides_uniformis NA
```

```

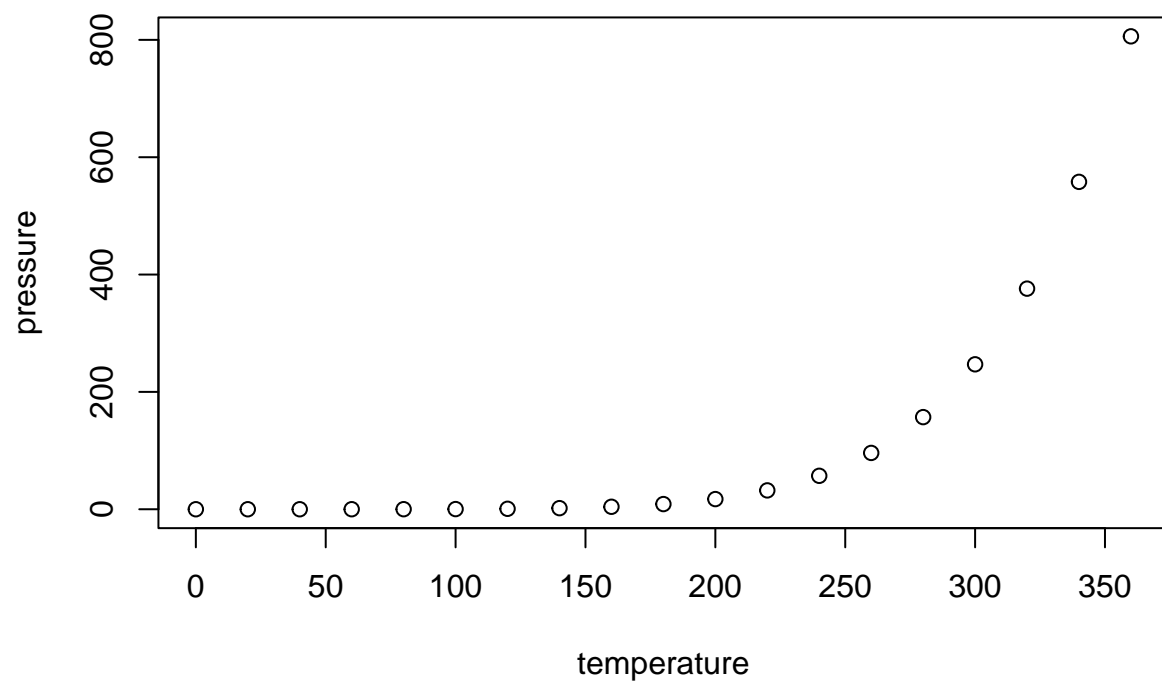
## UniRef90_A0A395VRV3 1.730327e+04
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_ovatus 7.697973e+00
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_vulgatus NA
## 23-MDcol033-1A
## UNMAPPED 1.816065e+06
## UniRef90_A0A174NA57 1.139728e+01
## UniRef90_A0A174NA57|g__Bacteroides.s__Bacteroides_uniformis NA
## UniRef90_A0A395VRV3 1.754432e+04
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_ovatus 8.162051e+03
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_vulgatus NA
## 25-MDcol036-1A
## UNMAPPED 2.876898e+06
## UniRef90_A0A174NA57 3.411955e+02
## UniRef90_A0A174NA57|g__Bacteroides.s__Bacteroides_uniformis NA
## UniRef90_A0A395VRV3 1.683431e+04
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_ovatus 1.683046e+04
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_vulgatus 4.744231e-01
## 27-MDcol025-1B
## UNMAPPED 2.535351e+06
## UniRef90_A0A174NA57 9.803922e+00
## UniRef90_A0A174NA57|g__Bacteroides.s__Bacteroides_uniformis NA
## UniRef90_A0A395VRV3 6.165756e+03
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_ovatus 6.164478e+03
## UniRef90_A0A395VRV3|g__Bacteroides.s__Bacteroides_vulgatus 6.428122e-01

```

```
rm(AllRPK)
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.