Rarefaction curve

Haydeé

2023-10-11

Read the biom file for the mines Pmt and Topaz.

```
library("phyloseq")
library("ggplot2")
library("RColorBrewer")
library("patchwork")
library(vegan)
```

```
## Loading required package: permute
```

```
## Loading required package: lattice
```

```
## This is vegan 2.6-2
```

```
setwd("~/metagenomas")


raw_metagenomes <- import_biom("metagenomes_Pmt_S1.biom")
raw_metagenomes@tax_table@.Data <- substring(raw_metagenomes@tax_table@.Data, 4)
colnames(raw_metagenomes@tax_table@.Data)<- c("Kingdom", "Phylum", "Class", "Order", "Family", "Genus",
unique(raw_metagenomes@tax_table@.Data[,"Kingdom"])
```

```
## [1] "Bacteria"  "Archaea"   "Eukaryota" "Viruses"
```

```
raw_metagenomes <- subset_taxa(raw_metagenomes, Kingdom == "Bacteria")
```

Extract the OTUs and obtain the maximun number of counts between the samples.

```
otu <- otu_table(raw_metagenomes)
otu <- as.data.frame(t(otu))
sample_names <- rownames(otu)

S <- specnumber(otu)
S
```

```
## Pmt2_S287   S1_S286
##      7887      7981
```

```
raremax <- min(rowSums(otu))
raremax
```

```
## [1] 2184595
```

We apply rarefy with the following parameters:

- Database: the transpose of the `otu` database.
- sample: a vector from 100000 to maximun number of counts with step 100000.
- Standar error: TRUE
- MARGIN: 2

With the rarefy function, we obtain the species richness in random samples of some specified size. This was calculated for a vector of samples from 100000 to the maximum of the two samples with steps of 100000. In this way we obtained the expected richness for each element of the vector and the standard error.

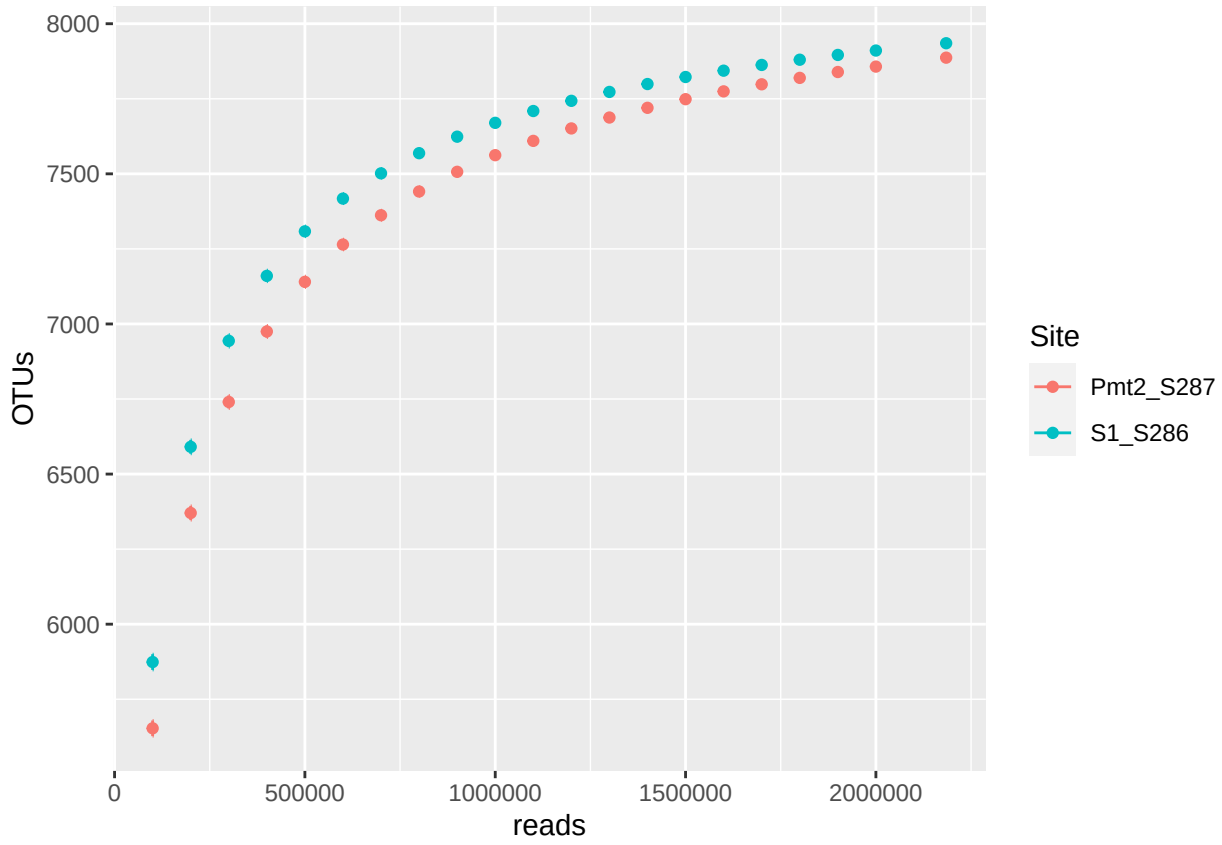
```
rare1 <- rarefy(t(otu),  sample= c(seq(100000,2000000 , by = 100000), raremax)  , se = T, MARG = 2)
rare_t <- t(rare1)
```

Save the data obtained by the rarefy function and compute the standar deviasion for each sample.

```
site <- c(rep("Pmt2_S287",21), rep("S1_S286",21))

reads <- c(c(seq(100000,2000000 , by = 100000), raremax), c(seq(100000,2000000 , by = 100000), raremax))

Pmt2_S <- rare_t[,1]
Pmt2_se <- rare_t[,2]
S1_S <- rare_t[,3]
S1_se <- rare_t[,4]

Svalues <- c(Pmt2_S, S1_S)
sevalues <- c(Pmt2_se, S1_se)

data <- data.frame(Site= site,
                   reads = reads,
                   S = Svalues,
                   se = sevalues)
```

Now we can use the library `ggplot` to obtain the rarefaction curve with the database obtained.

```
rare_plot <- ggplot(data, aes(x= reads, y=S, col=Site)) +
  geom_point()+
  geom_errorbar(aes(ymin=S-se, ymax=S+se), width=.3,
                position=position_dodge(0.001))+
  ylab("OTUs")

rare_plot
```


8000
7500
7000
6500
6000
OTUs
0
500000
1000000
1500000
2000000
reads
Site
Pmt2_S287
S1_S286