# 系统发育β多样性分析：微生物群落的决定性和随机性装配过程

# 基本概念

对高通量测序数据进行下游分析时，非限制性排序和限制性排序是群落β多样性最常见的分析方式。在微生物群落生态学中一个主要的目标是推断出形成微生物群落的主导因素既群落装配过程。我们经常使用到的微生物物种丰度就是这一过程的直接结果。所以常见的排序分析不能满足这一条件。为了解决这一问题，群落的分类，β多样性和系统发育就被联系起来建立了一个framework。其中对微生物群落构成的影响主要由四个因素组成：（1）环境过滤（Environment filtering）（2）扩散（Dispersal）（3）生态漂移（Ecological drift）（4）物种形成（Speciation）。基于上述因素，将群落的装配过程划分为决定性过程（deterministic）和随机性过程(stochastic)。这两种过程共同作用于群落，其中一种占主导作用。在分析过程中，需要理解的两个部分：

## 1.系统发育信号

1）系统发育信号（phylogenetic signal）：评估具有紧密关系的OTUs(或ASVs)之间是否具有相似的生境(具备系统发育信号是进行后续分析的前提)。这个部分是利用两个距离矩阵：一个是OTUs之间系统发育距离矩阵，另一个是由OTUs之间的环境最适值构成的矩阵。利用R语言中的Mantel correlogram 进行相关分析，结果中显著的正相关被认为具有系统发育信号。

## 2.平均最近种间距离和最近种间距离

利用测序数据来推断微生物群落装配过程所构建的framework，基于这样一种推断观点：将OTUs和物种丰度进行随机化，这一随机化过程定义为随机化零模型。经过这一随机化模型所得到的值为MNTDnull, 这个值被假定为微生物群落由绝对的随机性过程装配而成。那么作为原样本的观测值βMNTDobs距离βMNTDnull值越远，说明随机性的装配过程影响越弱而决定性的装配过程影响越强。相反距离越近，说明随机性的装配过程影响越强而决定性的装配过程影响越弱。单凭这个距离的远和近是不能直接判断出明确结果的，这个距离需要被量化才能更直接的得出结论，因此引入了βNTI这一指标用于衡量距离远近，也就是确定一个距离，超过这个距离就是由决定性的装配过程主导，而没有超过这个距离就是随机性的装配过程主导。具体判断为：βNTI<-2为决定性过程中的同质化选择（homogeneous selection），

-2<βNTI<2为随机性过程（stochasticity），βNTI>2 为决定性过程中的异质化选择（heterogeneity selection or variable selection）

相关计算公式如下：

1）MNTD（mean-nearest-taxon-distance）：特定时间和空间点上每个样品的系统发育群落组成的量化指标。

计算公式表示为：

表示OTU i 在群落 k 中的相对丰度； 表示群落 k 中的OTUs数量; 表示在群落 k 中OTU i 与k中剩余所有OTU j 之间的最小系统发育距离。

2）NTI（nearest-taxon-index）：量化MNTD观测值与MNTD随机化之后两者距离的指标。

计算公式表示为：

表示随机化一次得到一个, 随机999次之后取999个值的均值； 表示原样本观测值； 表示随机999次之后取999个值的标准差

3）βMNTD（β-mean-nearest-taxon-distance）：群落系统发育组成随时间更替的量化指标。

计算公式表示为：

表示OTU i 在群落 k 中的相对丰度； 表示群落 k 中的OTUs数量; 表示在群落 k 中OTU i 与群落m中所有OTU j 之间的最小系统发育距离,反之亦然。

4）βNTI（β-nearest-taxon-index）：量化βMNTD观测值与βMNTD随机化之后两者距离的指标。

计算公式表示为：

表示随机化一次得到一个, 随机999次之后取999个值的均值； 表示原样本观测值； 表示随机999次之后取999个值的标准差

# 实例解读

**例1. βNTI与pH水平和稀释度的相关性**

本文是2019年南农荀卫兵组发表于Nature Communications的文章，分析了黑壤和红壤中细菌装配过程与稀释水平和pH水平的相关性。

Fig. 3 βNTI values are associated with dilution and pH levels, and bacterial Shannon diversity. **a** The relationships between −Lg(Dil) and βNTI in black soil. **b** The relationships between −Lg(Dil) and βNTI in red soil. **c** The effect of soil pH on βNTI in black soil. **d** The effect of soil pH on βNTI in red soil. **e** Therelationship between |βNTI| and bacterial Shannon diversity indices. Dil indicates the dilution level. Lg(Dil) indicates the Lg-transformed dilution level. Lg(Dil) =0 represents the untreated soil

图表结果：图中直接展示了两种土壤的βNTI随稀释梯度变化以及随pH梯度变化关系和回归线，还有香浓多样性与的相关和回归。

文中用一节来详述结果如下：

To infer alterations in the deterministic/stochastic assembly processes along dilution and pH gradients, we examined the relationships between βNTI and dilution and pH gradients. The βNTI values within each treatment showed various patterns related to pH and dilution level (Fig. 3). In both black and red soils, community assembly shifted from stochastic processes (|βNTI| < 2) to deterministic processes (|βNTI| > 2) with increasing dilution levels and the βNTI values were significantly correlated with dilution gradients for all pH levels.

为了推断在确定性/随机装配过程中随着稀释和pH梯度的改变，我们研究了与稀释和pH梯度之间的关系。每个处理中的βNTI值都表现出与pH值和稀释水平相关的不同模式。在黑土和红土中，随着稀释梯度的增加，群落的装配过程由随机性转变为决定性过程。同时βNTI值与所有稀释梯度上的pH值都显著相关。

However, the trends in βNTI distribution over dilution gradients were not identical among different soil pH levels. Stochastic assembly processes were dominant in the least diluted samples and shifted towards dominance of variable selection (βNTI > 2) at the pH value of 6.5, while there was a shift towards dominance of homogeneous selection (βNTI <- 2) under more acidic (pH 4.5) and alkaline (pH 8.5) conditions in more diluted samples.

然而，各稀释梯度的βNTI分布趋势在pH梯度上不完全相同。在pH为6.5的水平中随机装配过程在稀释最少的样本中占主导地位，并向异质化选择的主导地位转移。同时在酸性pH为4.5和碱性pH为8.5的情况下也出现了同质化选择的转变。

The same βNTI datasets were reorganized to examine βNTI distributions across the pH gradient from 4.5 to 8.5 for each dilution level. For both the black and red soils, the βNTI values exhibited unimodal patterns along the pH gradient.

我们重新整理了相同的βNTI数据集，以检查每一个稀释水平下在pH梯度从4.5到8.5的不同浓度下的不同浓度的径向分布。在黑土和红土中，随着pH梯度的变化，其数值呈现单峰模式。

Peak values were observed for neutral pH soils, whereas valley values were observed under acidic and alkaline conditions. In addition, the magnitude of these patterns increased with higher dilution levels. These results suggested that a bacterial community with lower species richness was more likely to follow deterministic assembly processes.

在中性pH土壤中观察到峰值，而在酸性和碱性条件下观察到谷值。此外，这些模式的大小随着稀释水平的提高而增加。这些结果表明，物种丰富度较低的细菌群落更可能遵循决定性装配过程。

**例2. βNTI与土壤有机质水平的相关性**

本文是格罗宁根大学生命进化科学研究所微生物生态学研究室Francisco Dini-Andreote组在PNAS发表的文章。

Fig. S1. Phylogenetic Mantel correlogram evaluating phylogenetic signal in the soil microbial communities sampled in this study. The plot relates Pearson correlation coefficients to phylogenetic distances classes. Significant correlations (P < 0.05; solid symbols) indicate significant phylogenetic signal in OUT ecological niches but only within the associated phylogenetic distance class. The analysis reveals significant phylogenetic signal but only across relatively short phylogenetic distances.

****图表结果：图表中被填充的黑色点表示显著，且为正相关，结果表明了具备系统发育信号

Fig. 3. βNTI patterns from empirical comparisons and simulated ecological scenarios. (A) βNTI for all pairwise community comparisons—within and between successional stages—as a function of the change in log-transformed soil organic matter (SOM) concentration between communities, presented as the fold change in SOM. The linear regression model is shown as the green line; statistics are provided on the panel. Horizontal dashed lines indicate the upper (+2) and lower (−2) significance thresholds. Boxes laid over the βNTI data represent our conceptual interpretation of how simulation model outputs (summarized in B) align with the empirical relationship between βNTI and changes in SOM. (B) βNTI distributions obtained from simulated ecological scenarios (see Fig. S3 for a detailed description).

图表结果：主要结果是βNTI值与经过log转换的土壤有机质具有显著的相关性。

文章中的结果解读如下：

As the change in SOM increased, there was a continuous transition—on average—from βNTI < −2 at the smallest changes in SOM to −2 < βNTI < +2 at intermediate SOM shifts and to βNTI > +2 for the largest shifts in SOM (Fig. 3A). These patterns suggested that increasing shifts in SOM lead to a transition from homogeneous selection (small to no change in SOM), to weak selection and stochasticity (moderate changes in SOM), to variable selection (large changes in SOM).

随着土壤有机质的增加，βNTI出现了连续提高的趋势。当βNTI小于-2时，土壤有机质有最小的变化，当βNTI值在-2到2之间时，土壤有机质具有适中的变化，而当βNTI值大于2时，土壤有机质变化最大。这样的结果表明了土壤有机质的增加导致的群落装配过程由同质化选择（决定性过程）到随机性过程在到异质化选择（决定性过程）。

# 系统发育β多样性分析实战测试及代码

####系统发育信号（phylogenetic signal）####

###所需的包###

library(dplyr)

library(caret)

library("picante")

###数据导入###

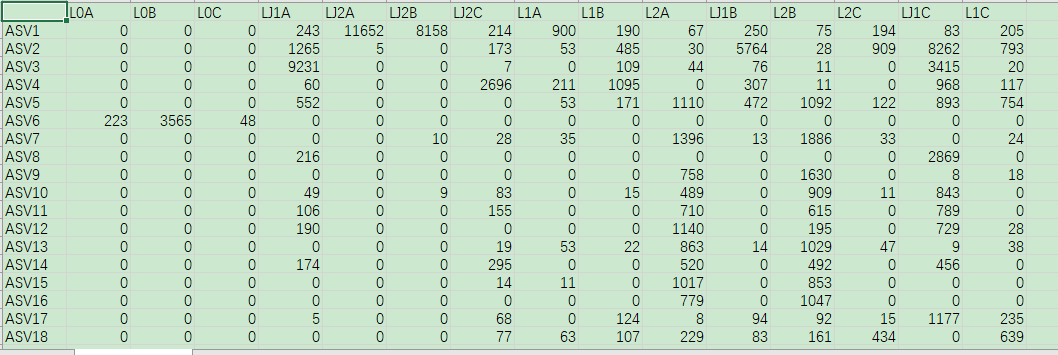
ASV<- read.table("ASVs-table-ZT.txt", row.names = 1, header = T, check.names = F)%>%t()#样本的ASVs表格

env <- read.table("enviromental2.txt", header = T, row.names = 1, check.names = F,sep = '\t')#样本的理化指标

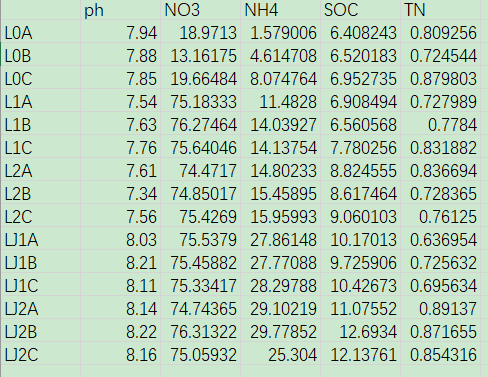
phy <- read.tree("rooted-tip-tree.nwk")#Newick格式的系统发育树文件

####导入的数据样式####

ASVs-table-ZT.txt（为常见的ASV表格，行代表ASV，列代表样本）



enviromental2.txt（样本的理化数据）



###数据的标准化###

ASV <- decostand(ASV, method = "total")

apply(ASV, 1, sum)%>%data.frame()%>%t()

env <- scale(env,center = TRUE, scale = TRUE)

###检验不匹配或缺失的物种###

combined <- match.phylo.comm(phy, ASV)#得到的对象是一个包含$phy和$comm元素的列表

###将原始数据替换为检验后匹配的数据###

phy <- combined$phy

ASV <- combined$comm%>%t()%>%data.frame()

###构建环境最适值矩阵###

result <- (as.matrix(ASV[1,]) %\*% as.matrix(env)) / (sum(ASV[1,]))

for(i in 2:nrow(ASV)){

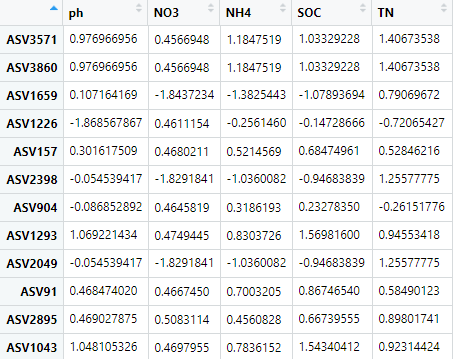
optium <- (as.matrix(ASV[i,]) %\*% as.matrix(env)) / (sum(ASV[i,]))

result <- rbind(result,optium)

}

###得到的环境最适值矩阵样式###

行代表每一个ASV，列代表了每一个理化指标，值为每一个理化指标在每个ASV中的环境最适值



###将环境最适值矩阵转化为欧式距离矩阵###

niche.eud <- vegdist(result, method = "euclidean", upper = FALSE)

###形成系统发育树距离矩阵###

phydist <- cophenetic(phy)

###将形成的系统发育树距离矩阵标准化为0-1之间###

phy.dist <- (phydist - min(phydist))/ (max(phydist) - min(phydist))

###计算系统发育信号###

otu.correlog <- mantel.correlog(niche.eud, phy.dist, nperm = 1000, mult = "bonferroni", n.class=100, cutoff = FALSE)

###查看并出图结果###

otu.correlog

plot(otu.correlog) + title(main = ", n.class=100, n.permutation=1000")

sink('signal-16s.txt')#将结果写出然后自己将图表美化

图中左端具有正相关且显著的相关性，表明具备系统发育信号

####计算系统发育β多样性####

###所需的包###

library(vegan)

library(picante)

library(phylocomr)

library(dplyr)

###数据的导入###

##数据导入的样式与计算系统发育信号时相同##

phy <- read.tree("rooted-tip-tree.nwk")

comm<-read.table("ASVs-table-ZT.txt", header = TRUE, row.names = 1)

###将comm标准化为0-1之间以及转换phy为距离矩阵###

com<-data.frame(t(comm))

com <- decostand(com, method = "total")

apply(com, 1, sum)

###计算MNTD及NTI###

phy1<-phy

phy.dist1<- cophenetic(phy1)

# 计算MNTD和NTI

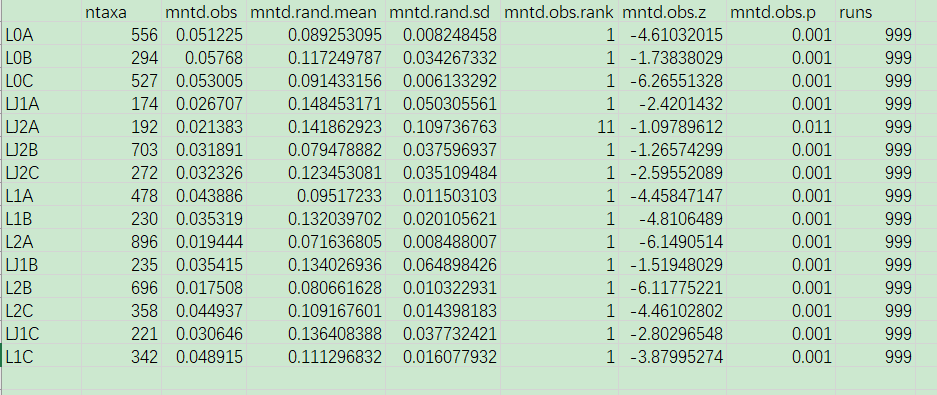
comm.sesmntd <- ses.mntd(com, phy.dist1, null.model = "taxa.labels",abundance.weighted =T,runs = 999)

comm.sesmntd

write.table(comm.sesmntd,"comm.mntd-NTI.txt",row.names = T, sep = '\t', quote = FALSE, na = '')

##结果样式为##

其中mntd.obs为每个样本的MNTD值，mntd.obs.z为每个样本的NTI值

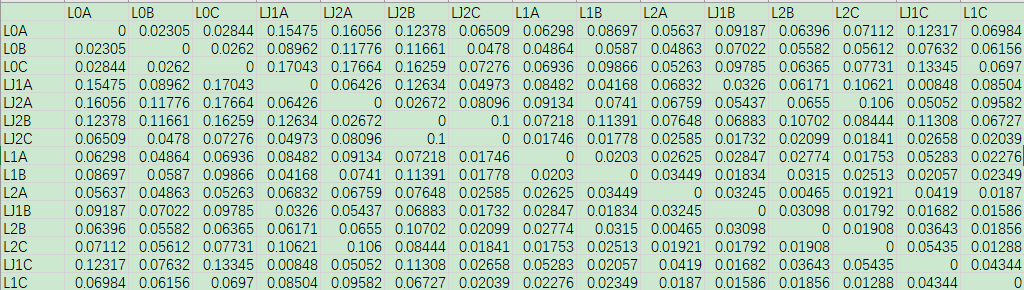


###计算βMNTD及βNTI###

#继续上述步骤#

betamntdobs<-comdistnt(com, phy.dist1, abundance.weighted = TRUE)#此计算即为原样本βMNTDobs值

betamntdobs<-as.matrix(betamntdobs)

#观测值输出样式为#

#写出观测值#

write.table(betamntdobs,"comm.betamntd-obs.txt",row.names = T, sep = '\t', quote = FALSE, na = '')

##计算随机化零模型##

###循环中使用的是tipShuffle代码，是将系统发育树枝端进行随机化，也可以用taxaShuffle，是将系统发育树树枝进行随机化（这个部分有文章中使用软件phylocom 4.2进行计算，但是自己测试了下，发现目前这个软件用不了也可能是自己不会用，所以通过自己理解然后根据公式和参看南土所褚海燕老师文章编辑了这部分代码。附上软件phylocom 4.2的链接http://phylodiversity.net/phylocom/）###

##进入循环，计算时间根据电脑配置，几十分钟到几个小时不等##

std<-list()#用于放置每一次循环输出的output，共放置999个~std[[1:999]]

output<-vector("double", length(mntd))#用于放置每一次循環的結果，每次一個結果然後放入std中，共有999个结果

i=1#设置循环起始值

while(i<1000){

phy3<-phy1

phy1<-tipShuffle(phy3)

phy.dist1<- cophenetic(phy1)

c<-comdistnt(com, phy.dist1, abundance.weighted = TRUE)

c<-as.matrix(c)

k=1-

#for循環用于放置每一次零模型，以便後期計算標準差

for (m in seq\_len(nrow(c))){

for (n in seq\_len(ncol(c))){

output[[k]]<-c[m,n]

k=k+1

}

}

std[[i]]<-output#此結果用於計算標準差

i=i+1

}

#对输出数据进行整理方便后续计算#

name=seq(1:999)#给列上序号

std1<-data.frame(std)

colnames(std1)<-name

std2<-bind\_rows(std1) %>%as.matrix()

###计算零模型均值mean.betaMNTDnull以及其标准差sd.betaMNTDnull###

mean.betaMNTDnull<-vector("double")#设置一个空向量装载均值数据，为后续计算做准备。

sd.betaMNTDnull<-vector("double")#设置一个空向量装载标准差数据，为后续计算做准备。

i=1#设置循环起始值

while(i<nrow(std2)+1) {

mean.betaMNTDnull[[i]]<-mean(std2[i,])

sd.betaMNTDnull[[i]]<-sd(std2[i,])

i=i+1

}

#观察结果#

mean.betaMNTDnull

sd.betaMNTDnull

##调整已经算好的观测值与零模型为同一样式方便后续计算##

betamntdobs<- as.matrix(betamntdobs)

comm.betamntdobs<-vector("double")#设置一个空向量装载样式调整后的数据

k=1#设置循环起始值

for (m in seq\_len(nrow(betamntdobs))){

for (n in seq\_len(ncol(betamntdobs))){

comm.betamntdobs[[k]]<-betamntdobs[m,n]

k=k+1

}

}

##计算观测值与零模型均值之差即 ##

minus.mntd<-comm.betamntdobs-mean.betaMNTDnull

##计算betaNTI##

betaNTI<-vector("double")#设置一个空向量装载betaNTI数据

for (n in seq\_len(nrow(std2))) {

betaNTI[[n]]<-minus.mntd[[n]]/sd.betaMNTDnull[[n]]

}

betaNTI[is.na(betaNTI)] <- 0#计算过程中0/0被认为缺失值，这里手动赋值为0

betaNTI

#最后对数据进行整理，调整为与上面计算βMNTD时一样的输出样式#

TbetaNTI<-data.frame()#设置一个数据框装载调整样式后的betaNTI数据

k=1#设置循环起始值

for (m in seq\_len(nrow(betamntdobs))){

for (n in seq\_len(ncol(betamntdobs))){

TbetaNTI[m,n]<-betaNTI[[k]]

k=k+1

}

}

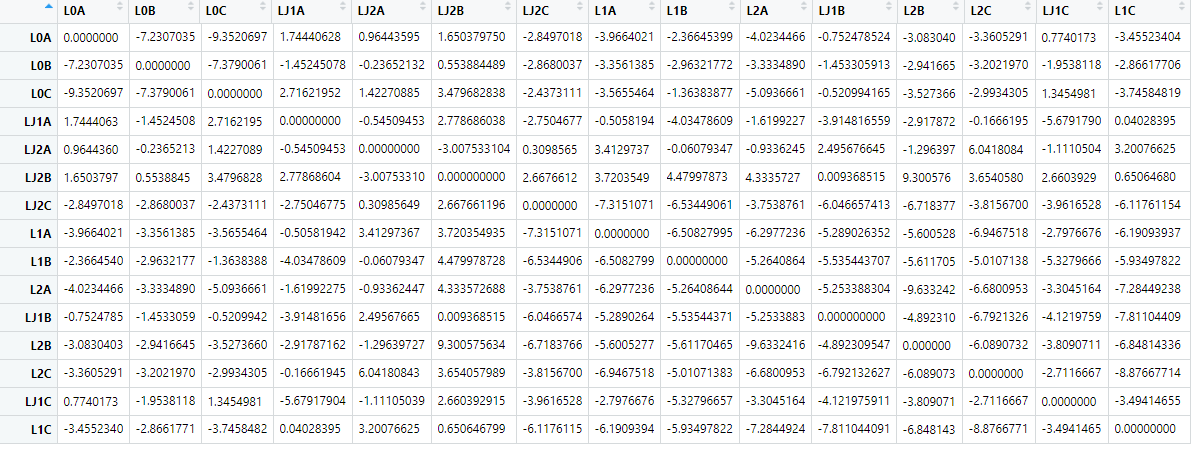
#调整行列名称#

rownames(TbetaNTI)<-rownames(betamntdobs)

colnames(TbetaNTI)<-colnames(betamntdobs)

TbetaNTI

#βNTI的计算结果#



####后续将βNTI的计算结果与样本的环境理化指标进行对应作图然后回归分析就可以做出实例中的图了####

参考文献：

Dini-Andreote F, Stegen JC, van Elsas JD, Salles JF (2015) Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession Proceedings of the National Academy of Sciences of the United States of America 112:E1326-E1332 doi:10.1073/pnas.1414261112

Fan K, Weisenhorn P, Gilbert JA, Shi Y, Bai Y, Chu H (2018) Soil pH correlates with the co-occurrence and assemblage process of diazotrophic communities in rhizosphere and bulk soils of wheat fields Soil Biology and Biochemistry 121:185-192 doi:10.1016/j.soilbio.2018.03.017

Feng M et al. (2018) Long-term fertilization influences community assembly processes of soil diazotrophs Soil Biology & Biochemistry 126:151-158 doi:10.1016/j.soilbio.2018.08.021

Stegen JC, Hurlbert AH (2011) Inferring ecological processes from taxonomic, phylogenetic and functional trait beta-diversity PloS one 6:e20906 doi:10.1371/journal.pone.0020906

Stegen JC et al. (2013) Quantifying community assembly processes and identifying features that impose them The ISME journal 7:2069-2079 doi:10.1038/ismej.2013.93

Stegen JC, Lin X, Konopka AE, Fredrickson JK (2012) Stochastic and deterministic assembly processes in subsurface microbial communities The ISME journal 6:1653-1664 doi:10.1038/ismej.2012.22