原核微生物群落随机性和确定性装配过程的计算方法

Calculation Method for Stochastic and Deterministic Assembly Processes of Prokaryotic Communities

赵维^{1,2}, 王兴彪^{1,2}, 侍浏洋^{1,2}, 朱婉瑜^{1,2}, 马磊^{1,2}, 王敬敬^{1,2}, 徐松^{1,2}, 杨榕^{1,2}, 张小霞^{1,2}, 韩一凡^{1,2}, 黄志勇^{1,2,*}

¹ 中国科学院天津工业生物技术研究所, 天津市工业生物系统与过程工程重点实验室, 天津; ² 国家合成生物技术创新中心, 天津

*通讯作者邮箱: huangzy@tib.cas.cn

引用格式: 赵维, 王兴彪, 侍浏洋, 朱婉瑜, 马磊, 王敬敬, 徐松, 杨榕, 张小霞, 韩一凡, 黄志勇. (2021). 原核微生物群落随机性和确定性装配过程的计算方法. *Bio-101* e2003400. Doi: 10.21769/BioProtoc.2003400.

How to cite: Zhao, W., Wang, X. B., Shi, L. Y., Zhu, W. Y., Ma, L., Wang, J. J., Xu, S. Yang, R., Zhang, X. X., Han, Y. F. and Huang, Z. Y. (2021). Calculation method for stochastic and deterministic assembly processes of prokaryotic communities. *Bio-101* e2003400. Doi: 10.21769/BioProtoc.2003400. (in Chinese)

摘要: 微生物群落生态学的一个主要目标是了解构成跨时空物种丰度模式的过程 (Stegen *et al.*, 2012; Zhou *et al.*, 2014; Dini-Andreote *et al.*, 2015)。确定性和随机性两种类型的过程会影响群落的聚集。确定性过程 (其中非生物和生物因素决定物种的存在与否和相对丰度) 与生态选择相关 (Vellend, 2010), 随机过程包括不可预测的扰动、概率性的散布和随机的出生-死亡事件等, 这些变化不是由环境决定的适应性结果 (Chase and Myers, 2011)。为了判定群落间变化由确定性过程主导还是随机性过程主导, 本实验使用零模型量化群落的绝对系统发育距离与随机系统发育距离的偏离度, 偏离程度越大, 群落受确定性因素的影响越大, 偏离度越小, 群落受随机性因素的影响越大。结果: 该方法可以使用 NTI/NRI 指示单个群落内共存的分类单元相比偶然预期的关系更为紧密还是分散, 使用 β NTI/ β NRI 指示两两群落间的变化受确定性或随机性因素影响的大小。结论: 该方法可用于评估不同时空尺度下随机性和确定性过程对微生物群落组装的影响。

关键词: 随机性过程, 确定性过程, 群落组装, 群落周转, 时空变化

软件

1. R version 3.4.3

实验步骤

1. 参照《Extraction and 16S rRNA Sequence Analysis of Microbiomes Associated with Rice Roots》(DOI:10.21769/BioProtoc.2884) (Edwards *et al.*, 2018)中的方法对微生物群落进行 16S rRNA 序列分析获得样品的群落组成和系统发育关系树, 对应的两个文件分别为 OTU 表格 `out_table.txt`, 和树文件 `rep_set.tre`。

2. 系统发育群落组成的计算参数

针对单个群落, 使用净关系指数 (NRI) 和最近分类单元指数 (NTI) 来反映物种集聚的谱系结构。二者分别对应平均谱系距离 (MPD) 和最近谱系距离的 (MNTD) 的标准化测度。

针对多个样本间的群落比较, 则是基于 MPD 和 MNTD 使用 β MPD 和 β MNTD 计算群落间谱系 **beta** 多样性, 而对于群落间聚集或分散受确定性或随机性因素影响的判断则是基于 NRI 和 NTI 使用 β NRI 和 β NTI 计算群落间系统发育 **beta** 多样性的显著性。

以上参数均可在 R 软件包“picante”中完成计算 (Webb *et al.*, 2002)。本文将分别基于上述参数介绍具体的含义与计算方法。

2.1 MNTD/MPD

MNTD 是 mean-nearest-taxon-distance (最近种间平均进化距离) 的缩写, 代表任意一个物种与群落中与自己亲缘关系最近的一个物种的平均进化距离; 它们的值越小, 物种间的亲缘关系越近。

计算公式如下:

$$MNTD = \sum_{k=1}^n f_{ik} \min(\Delta_{ikjk})$$

f_{ik} : 群落 k 中 OTU i 的相对丰度;

n_k : 群落 k 中 OTU 的数量;

Δ_{ikjk} : OTU_i 与同样在群落 k 中的所有其他 OTU_j 之间的最小系统发育距离。

MPD 是 mean phylogenetic distance (种间平均进化距离)的缩写, 是群落中所有物种对间平均的进化距离。MPD 分别计算样本内所有种两两之间 (两两之间成对) 的系统发育距离, 再除以两两之间的对数。其计算可在 R 的 `picante` 包中使用 `mpd()` 函数完成。

2.2 NTI/NRI

NRI 和 NTI 这两个系数可以描述实际群落与零群落之间的平均谱系距离, 它们都利用零群落中谱系距离的标准差进行了标准化。

NTI 是 nearest-taxon-index (最近分类单元指数) 的缩写, 是对样本中每个分类单元到最近分类单元的系统发育距离的标准化度量, 量化终端聚类的程度, 对应最近谱系距离的(MNTD)的标准化测度。为了评估系统发育群落结构的非随机程度, 将 OTU 及其相对丰度在系统发育的各个尖端进行随机分配 (null.model='taxa.labels'in'ses.mntd')。NTI 量化观察到的 MNTD 与零分布均值的标准偏差数(999 个随机)。NTI 是 R `picante` 包中“ses.mntd”输出的相反数。计算公式如下:

$$NTI = (\overline{MNTD_{null}} - MNTD_{obs}) / sd(MNTD_{null})$$

NRI 是 net relatedness index (净关系指数)的缩写, 是样本中分类单元相对于系统发育的平均成对系统发生距离的标准化度量, 基于适当的物种库, 量化一棵树上的分类单元的总聚类程度, 对应平均谱系距离(MPD)的标准化测度。NRI 是 R `picante` 包中“ses.mpd”输出的相反数。计算公式如下:

$$NRI = (\overline{MPD_{null}} - MPD_{obs}) / sd(MPD_{null})$$

MPD_{obs} (或 MNTD_{obs}) 为各样地观测值, MPD_{null} (或 MNTD_{null}) 和 sdMPD_{null} (或 sdMNTD_{null}) 则分别对应由零模型得到的平均值和标准偏差。

对于单个群落, NTI (或 NRI) 大于+2 表示共存的分类单元比偶然预期的关系更为紧密 (系统发生聚类); NTI (或 NRI) 小于-2 表示共存的分类单元比偶然预期的亲缘关系更远 (系统发生过度分散)。在所有的群落中获得的平均 NTI (或 NRI) 与零值有显著差异, 表示存在平均聚类 (>0) 或过度分散 (<0) (Kembel, 2009)。

2.3 β MNTD/ β MPPD

系统发育随时间或空间的变化 (系统发育 β 多样性) 使用 beta 最近谱系距离 β MNTD (或 beta 平均谱系距离 β MPPD) 和 beta 最近谱系指数 β NTI (或 beta 平均谱系指数 β NRI) 来量化, 它们分别是 MNTD (MPD) 和 NTI (NRI) 的组间 (between-assembly) 相似物 (Webb *et al.*, 2009; Kembel *et al.*, 2011)。

β MNTD 是发生在两个不同群落中的最近亲属间的丰度加权平均系统发育距离。这种度量量化了群落装配之间的系统发育距离; β MNTD 被使用是因为它强调短的系统发育距离的系统发育周转率, 其中大力支持系统发育信号的假设。其计算公式如下:

$$\beta\text{MNTD}=0.5[\sum_{i_k=1}^n f_{ik}\min(\Delta_{ikjm})+\sum_{i_m=1}^n f_{im}\min(\Delta_{imjk})]$$

$\min(\Delta_{ikjm})$: 群落 k 中的 OTU_i 与群落 m 中所有 OTU_j 之间的最小系统发育距离(其他变量请参见上文)。

β MPPD 是发生在两个不同群落中的任意两个物种间的丰度加平均系统发育距离。其计算公式如下:

$$\beta\text{MPPD}=0.5[\sum_{i_k=1}^n f_{ik}(\Delta_{ikjm})+\sum_{i_m=1}^n f_{im}(\Delta_{imjk})]$$

Δ_{ikjm} : 群落 k 中的 OTU_i 与群落 m 中所有 OTU_j 之间的平均系统发育距离(其他变量请参见上文)。

2.4 β NTI/ β NRI

为了推断出对特定群落装配之间观察到的周转率起主要作用的生态因素, 鉴于随机性生态过程的主导地位, 使用零模型方法生成一个预期水平的 β MNTD (β MPPD)。为了量化观察到的 β MNTD (β MPPD) 值和零 β MNTD (β MPPD) 值分布之间偏差的大小和方向, 使用了 β -最近谱系指数 β NTI (β -平均谱系指数 β NRI)。通过

将整个系统发育过程中的 OTU 随机化并重新计算 β MNTD (β MPD) 999 次，可以得到零 β MNTD (β MPD) 分布。 β NTI (β NRI) 是观察到的 β MNTD (β MPD) 与零 β MNTD (β MPD) 分布均值的标准偏差数。

$$\beta NTI = (\beta MNTD_{Obs} - \overline{\beta MNTD_{null}}) / sd(\beta MNTD_{null})$$

$$\beta NRI = (\beta MPD_{Obs} - \overline{\beta MPD_{null}}) / sd(\beta MPD_{null})$$

对于成对的群落比较， β NTI (β NRI) <-2 或 >+2 分别表示小于或大于预期的系统发育周转。在所有成对比较中获得的平均 β NTI (β NRI) 与零有显著差异，表示大于 (>0) 或小于 (<0) 预期平均周转 (Webb *et al.*, 2009; Kembel *et al.*, 2011)。

结果与分析

1. R 中计算示例——针对一对样本的计算

上述参数的计算可以在 R 中实现，现以两个来自海洋环境的原核微生物群落的数据为例子，计算上述参数。样品分别命名为 S1 和 S2，OTU 文件命名：out_table.txt；树文件命名：rep_set.tre；相应的脚本：betaNTI for one pair sample.R，可在该地址下载：<https://doi.org/10.5281/zenodo.4506365>。

```
library(picante)
library(ape)
comm<-
read.table("out_table.txt",sep='\t',row.names=1,header=T)
#读入 OTU 文件
#读入树文件
phy<-read.tree("rep_set.tre")
#使树文件和 OTU 表文件对齐
prune_tree<-prune.sample(comm,phy)
#计算每个 OTU 之间距离矩阵
phydist<-cophenetic(prune_tree)
```

#计算 MNTD

```
mntd<-
ses.mntd(comm,phydist,null.model="taxa.labels",abundance.weighted=T, runs=999)
```

该步骤计算结果如下:

	ntaxa	mntd.obs	mntd.rand.mean	mntd.rand.sd	mntd.obs.rank	mntd.obs.z	mntd.obs.p	runs
S1	498	0.13913	0.1712452	0.02944938	108	-1.090523	0.108	999
S2	586	0.1432301	0.1591859	0.01302431	90	-1.225076	0.09	999

其中, ntaxa Number of taxa in community, 群落中 OTU 的数目; mntd.obs Observed MNTD in community, 观测到的群落 MNTD 值; mntd.rand.mean Mean MNTD in null communities, 随机群落的 MNTD 均值; mntd.rand.sd Standard deviation of MNTD in null communities, 随机群落 MNTD 值的标准差; mntd.obs.rank Rank of observed MNTD vs. null communities, 观察到的群落的 MNTD 值与随机群落 MNTD 值的排序; mntd.obs.z Standardized effect size of MNTD vs. null communities $(=(\text{mntd.obs}-\text{mntd.rand.mean})/\text{mntd.rand.sd})$, equivalent to -NTI) 观察到的群落的 MNTD 与随机群落的 MNTD 值的标准化数值大小, 相当于-NTI 值; mntd.obs.p P-value (quantile) of observed MNTD vs. null communities $(=\text{mntd.obs.rank}/\text{runs} + 1)$ 观察到的群落的 MNTD 与随机群落的 MNTD 值比较的 p 值; runs Number of randomizations, 随机化的次数。

#计算 β MNTD 值

```
comdist.result<-comdistnt(comm,phydist)
```

该步骤计算结果如下:

	S1
S2	0.06070395

#以下为 β MNTDnull 的计算过程

```
f<-function() {
  g<-
  randomizeMatrix(comm,null.model=c("frequency","richness","independentswap","trialswap"),iterations=1000) #针对两组样方实现随机群
```

落构建

```
f<-comdistnt(g,phydist)#针对随机构建的两个样方进行群落间  $\beta$ MNTD 的
计算
}
mt<-apply(replicate(999,f()),1,mean)#上述 f 计算过程重复进行 999 次,
得到  $\beta$ MNTD 的平均值
mt
mt.sd<-apply(replicate(999,f(comm)),1,sd)#计算上述 999 次计算得到的
 $\beta$ MNTD 值的标准差
mt.sd
 $\beta$ NTI<-(comdist.result-mt)/mt.sd
```

最终，得到的 β NTI 值为 4.93，说明 S1 和 S2 两个样方的群落之间的变化主要由确定性过程影响。

2. R 中计算示例——针对多样本的计算

以上是针对 1 对样本之间计算 β NTI 值的介绍，给出了每一步操作的目的，下面介绍针对多样本间 β NTI 值计算的方法。该方法参考 Stegen *et al.* 2013, ISME (Stegen *et al.*, 2013), 依然在 R 中进行操作，相应的脚本: betaNTI for multiple pairs of samples.R 可在前文地址下载，具体如下：

```
##切换到包含 OTU 表和系统发育树的文件目录
#加载这个数据包
library(picante)
#读取 OTU 表，依然是样品名在行，OTU 种类在列
otu = read.csv("out_table.txt", header=T, row.names=1)
#读取系统发育树
phylo = read.tree("rep_set.tre")
#确保系统发育树上的名称与 otu 表中的名称排序一致
match.phylo.otu = match.phylo.data(phylo, t(otu))
#计算 betaMNTD，这里 abundance.weighted=T 表示考虑物种丰度，如果不设定该
参数，则默认 abundance.weighted=F，不考虑物种丰度
```



```

beta.mntd.weighted =
as.matrix(comdistnt(t(match.phylo.otu$data), cophenetic(match.p
hylo.otu$phy), abundance.weighted=T))
#检查 beta.mntd.weighted 矩阵中的列名是否和系统发育树以及 OTU 表中一致, 结
果应该为 T
identical(colnames(match.phylo.otu$data), colnames(beta.mntd.we
ighted))
#同上, 只是一个检查, 结果应该为 T
identical(colnames(match.phylo.otu$data), rownames(beta.mntd.we
ighted))
##计算 betaMNTD 随机值
#随机化次数
beta.reps = 999
rand.weighted.bMNTD.comp=array(c(-
999), dim=c(ncol(match.phylo.otu$data), ncol(match.phylo.otu$dat
a), beta.reps));
for (rep in 1: beta.reps) {
#这一步执行的随机化是基于对系统发育树中 OTU 标签的随机重排进行的, 如果沿用 1
中的方法对每一个群落进行随机构建后再计算 betaMNTD, 则可以选择使用
rand.weighted.bMNTD.comp[, , rep] =
as.matrix(comdistnt(randomizeMatrix(comun, null.model =
c("frequency", "richness", "independentswap", "trialswap"),
iterations =
1000), cophenetic(match.phylo.comun$phy), abundance.weighted=T, e
xclude.conspecifics = F)) 命令完成该步运算
rand.weighted.bMNTD.comp[, , rep] =
as.matrix(comdistnt(t(match.phylo.otu$data), taxaShuffle(cophen
etic(match.phylo.otu$phy)), abundance.weighted=T, exclude.conspe
cifics = F));
print(c(date(), rep));
}

```



```
##计算  $\beta$ NTI
weighted.bNTI =
matrix(c(NA),nrow=ncol(match.phylo.otu$data),ncol=ncol(match.p
hylo.otu$data));
for (columns in 1:(ncol(match.phylo.otu$data)-1)) {
  for (rows in (columns+1): ncol(match.phylo.otu$data)) {
#按照  $\beta$ NTI 计算公式计算每一对样本间的  $\beta$ NTI 值
    rand.vals = rand.weighted.bMNTD.comp[rows,columns,];
    weighted.bNTI[rows,columns] =
(beta.mntd.weighted[rows,columns] - mean(rand.vals)) /
sd(rand.vals);
    rm("rand.vals");
  };
};
rownames(weighted.bNTI) = colnames(match.phylo.otu$data);
colnames(weighted.bNTI) = colnames(match.phylo.otu$data);
weighted.bNTI;
#得到全部样本对间的  $\beta$ NTI 值并写入文件
write.csv(weighted.bNTI,"weighted_bNTI.csv",quote=F);
```

3. 多组样品计算结果示例

针对多组样品多个样方的 β NTI 计算按照上述示例 2 分别计算两两样方之间的 β NTI 值，并按照设计的样方分组分析 β NTI 在不同分组中的变化范围。本例展示来自渤海湾 6 个不同站位表层海水样品在不同月份 (5、8、10 月) 的 β NTI 统计值。18 个样品的物种组成数据已上传至前文地址。按照示例 2 分别计算了 unweighted β NTI 和 weighted β NTI，并将计算结果在 Origin 中进行 box-plot 制图，结果如下：

图 1. 6 个站位在不同月份的 β -NTI 值分布

如图 1 所示，我们分别计算了 unweighted β NTI 和 weighted β NTI，unweighted β NTI 是不考虑物种丰度的计算结果，weighted β NTI 是考虑物种丰度的计算结果。根据 unweighted β NTI 的统计结果，显示 5 月份、8 月份、10 月份的海水样品 β NTI 值主要集中在 -1~6 之间，说明在各个月份内，不同的站位间的群落变化既有受随机性因素影响的情况，也有受决定性因素影响的情况；而两两月份间的样品分析显示 β NTI 值均大于 2，说明月份间的微生物群落结构变化主要受决定性因素的影响。weighted β NTI 的统计结果与 unweighted β NTI 的统计结果有所不同，显示月份内和月份间的样品对之间的 β NTI 值均有较大的波动，在 -2~9 之间，说明考虑物种丰度变化时，该区域的微生物群落结构受取样点和时间的影响均较大。

结果与分析

1. 软件 Phylocom 也对上述参数提供了标准化的解决方案，本文主要提供了基于 R 中 picante 数据包进行计算的方法 (Webb *et al.*, 2002)。以上我们给出了 MNTD、 β MNTD、NTI、 β NTI 的计算方法，对应的 MPD、 β MPD、NRI、 β NRI 的计算则只需更改相应的函数即可获得。这两组参数在生物学意义上既相似又有不同，它们分别从不同角度描述了群落在系统发育树上的聚集模式，MNTD 侧重指示分支末端的谱系关系，MPD 代表了物种间亲缘关系的平均情况。基于两种谱系测度 (MPD 和 MNTD) 估计群落内部与群落之间成员亲缘关系的强度时，它们对解释变量的灵敏度可能不尽相同。在实际研究中，针对微生物群落的研究显示 MNTD 对肠道微生物

物群落的宿主选择性更敏感 (Youngblut *et al.*, 2019), 而针对草本植物群落的研究则显示 NRI 和 β MNTD 与海拔梯度之间存在显著线性关系, 而 NTI 和 β MPD 并未发现显著趋势 (赵鸣飞等, 2017)。这很可能跟指标本身的算法特性有关。具体来说, 对区域种库整体上具有谱系聚集趋势的情况, 该两指标同时倾向于取得较高值; 而谱系聚集发生于谱系树的多个浅层分支上时, 仅 NTI 易于得到较高值 (Vamosi *et al.*, 2009)。

2. 关于样品采样深度对该方法计算结果的影响, 目前尚无系统性的报道, 根据已有报道基于该方法进行的研究显示 Stegen 等 (Stegen *et al.*, 2012) 研究随机和确定性过程对不同深度的沉积物样品的相对影响时, 对样品的序列取样深度为 500 条; Wang 等 (Wang *et al.*, 2013) 对不同生态系统细菌群落系统发育多样性受随机性和决定性因素影响的研究中, 每个样品的序列取样深度为 1,000 条; Yu 等 (Yu *et al.*, 2018) 研究空间尺度对种植小麦土壤中微生物群落随机性和决定性相对作用的影响时, 对每个样品的序列取样深度为 20,005 条。可见, 随着测序技术越来越成熟, 用于进行群落系统发育分析的取样深度也越来越大。另外, 关于物种数对系统发育多样性的影响, 以 MNTD 和 NTI 为例, 其他系统发育多样性指数类似: MNTDobs 是观测到的 MNTD, MNTDnull 是随机打乱进化树之后得到的 MNTD 值, 这个过程重复多次 (~999 次), 可得到 MNTDnull 的分布。可近似转化为均值为 0, 标准差为 1 的标准正态分布。而这个分布其实是所有物种施加影响结果的集合。NTI 和 β NTI 计算过程已经排除了物种数量对系统发育的影响。
3. 关于使用零模型 (null model) 方法解释原核微生物群落构建: 零模型用于验证基于自然实验的生态学假说, 它以计算机技术为基础, 从已经获得的实验数据出发, 构建受控的生态随机模型, 最后用标准的统计检验方法验证理论假说。在解释原核微生物群落构建机制中, 零模型是群落数据或者进化树数据进行随机化的方式。运行零模型主要是为了解, 在随机状况下, 物种系统发育关系的统计分布, 从而推断真实的群落内物种的系统发育关系是否随机, 以便了解环境过滤或者随机过程的相对重要性。所以在探讨这些生态学过程的相对重要性, 计算 NTI 和 β NTI 的时候, 必须要考虑零模型。目前, 该方法使用基于 randomizeMatrix 函数提供的零模型构建主要包括物种内群落数据矩阵丰度 (维持物种出现频率) 的随机化 (frequency)、样本内群落数据矩阵丰度 (维持样本物种丰富度) 随机化 (richness)、利用独立交

换算法 (Gotelli, 2000) 对群落数据矩阵进行随机化 (保持物种发生频率和样本物种丰富度) (independentswap)、使用试交换算法 (Miklos and Podani, 2004) 随机化群落数据矩阵 (保持物种发生频率和样本物种丰富度) (trialswap)。对于生态零模型, Gotelli 指出“它是基于生态数据的随机化或者基于已知分布、设想分布的随机样本基础上的格局生成模型。与已知数据相比, 一些数据的性质在零模型中保持不变, 另外一些则允许随机变化, 这样就产生了一个新的格局总体。”从中可以看出零模型的构建是从实际所得数据出发的, 并非真正的实验。对于原核微生物群落而言, 环境中的微生物种类只能尽最大的可能通过高通量、高分辨率的分析方法获得, 但是仍然无法完全获知其中的全部微生物组成, 因此, 零模型的构建也只能依据已知样品已获得的物种信息进行预测, 对于未能获得的物种信息则尚不能纳入到零模型的构建中。但是, 针对样方之间的比较, 我们更关注设计样方间的相对变化规律, 因此, 取样深度的一致性和分析对比的前提。

4. 注意, 文中关于 NTI/NRI、 β NTI/ β NRI 的结果解释时其显著性的判断依据是其临界值是否大于|2|, 实际上这是依据标准正态分布的 95%置信区间得到的, 一般认为 NRI 或者 NTI 大于 1.96 或者小于-1.96 的结果, 在 95%的水平是显著的。这是因为计算 NTI/NRI 或者 β NTI/ β NRI 都是基于 Standard effective Size 的变型, 其本质都是基于多次随机化的群落或者进化树计算的, 在随机样本量很大的情况下, MNTD/MPD 或者 β MNTD/ β MPD 的分布会接近正态分布(张金龙, 2019)。所以文献中提及的这几个临界值接近于 2, 实际是接近 1.96。

致谢

感谢国家重点研发计划重点专项项目 (2018YFA0901200); 天津市科技计划项目 (18ZXSZSF00100,19YFZCSF00750); 国家自然科学基金项目 (41977200) 的支持。感谢 Campbell O Webb, Steven W Kembel, James C Stegen 等人先前的研究工作。感谢西北农林科技大学焦硕教授的热心指导!

参考文献

1. Chase, J. M. and Myers, J. A. (2011). [Disentangling the importance of ecological niches from stochastic processes across scales](#). *Philos Trans R Soc Lond B Biol Sci* 366(1576): 2351-2363.

2. Dini-Andreote, F., Stegen, J. C., van Elsas, J. D. and Salles, J. F. (2015). [Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession.](#) *Proc Natl Acad Sci U S A* 112(11): E1326-1332.
3. Edwards, J., Santos-Medellín, C. and Sundaresan, V. (2018). [Extraction and 16S rRNA Sequence Analysis of Microbiomes Associated with Rice Roots.](#) *Bio-protocol* 8(12): e2884.
4. Gotelli, N. J. (2000) [Null Model Analysis of Species Co-Occurrence Patterns.](#) *Ecology* 81: 2606-2621.
5. Kembel, S. W. (2009). [Disentangling niche and neutral influences on community assembly: assessing the performance of community phylogenetic structure tests.](#) *Ecol Lett* 12(9): 949-960.
6. Kembel, S. W., Eisen, J. A., Pollard, K. S. and Green, J. L. (2011). [The phylogenetic diversity of metagenomes.](#) *PLoS One* 6(8): e23214.
7. Miklos, I. and Podani, J.(2004). [Randomization of presence-absence matrices: Comments and new algorithms.](#) *Ecology* 86:92.
8. Stegen, J. C., Lin, X., Konopka, A. E. and Fredrickson, J. K. (2012). [Stochastic and deterministic assembly processes in subsurface microbial communities.](#) *ISME J* 6(9): 1653-1664.
9. Stegen, J. C., Lin, X., Fredrickson, J. K., Chen, X., Kennedy, D. W., Murray, C. J., Rockhold, M. L. and Konopka, A. (2013). [Quantifying community assembly processes and identifying features that impose them.](#) *ISME J* 7(11): 2069-2079.
10. Vamosi, S. M., Heard, S. B., Vamosi, J. C. and Webb, C. O. (2009). [Emerging patterns in the comparative analysis of phylogenetic community structure.](#) *Mol Ecol* 18(4): 572-592.
11. Vellend, M. (2010). [Conceptual synthesis in community ecology.](#) *Q Rev Biol* 85(2): 183-206.
12. Wang, J., Shen, J., Wu, Y., Tu, C., Soininen, J., Stegen, J. C., He, J., Liu, X., Zhang, L. and Zhang, E. (2013). [Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus stochastic processes.](#) *Isme J* 7: 1310-1321.

13. Webb, C. O., Ackerly, D. D. and Kembel, S. W. (2008). [Phylocom: software for the analysis of phylogenetic community structure and trait evolution](#). *Bioinformatics* 24(18): 2098-2100.
14. Webb, C. O., Ackerly, D. D., McPeck, M. A. and Donoghue, M. J. (2002). [Phylogenies and community ecology](#). *Annu Rev Ecol Syst* 33: 475-505.
15. Youngblut, N. D., Reischer, G. H., Walters, W., Stalder, G., Ley, R. E. and Farnleitner, A. H. (2019). [Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades](#). *Nat Commun* 10(1):2200.
16. Shi, Y., Li, Y., Xiang, X., Sun, R., Yang, T., He, D., Zhang, K., Ni, Y., Zhu, Y. G., Adams, J. M. and Chu, H. (2018). [Spatial scale affects the relative role of stochasticity versus determinism in soil bacterial communities in wheat fields across the North China Plain](#). *Microbiome* 6(1): 27.
17. Zhou, J., Deng, Y., Zhang, P., Xue, K., Liang, Y., Van Nostrand, J. D., Yang, Y., He, Z., Wu, L., Stahl, D. A., Hazen, T. C., Tiedje, J. M. and Arkin, A. P. (2014). [Stochasticity, succession, and environmental perturbations in a fluidic ecosystem](#). *Proc Natl Acad Sci U S A* 111(9): E836-845.
18. 赵鸣飞, 薛峰, 王宇航,等.(2017). 山西芦芽山针叶林草本层群落谱系结构与多样性的海拔格局. *植物生态学报*, 41(7): 707-715.
19. 张金龙. (2019). [群落系统发育分析常见问题合集](#). 科学网, 2019-1-6.