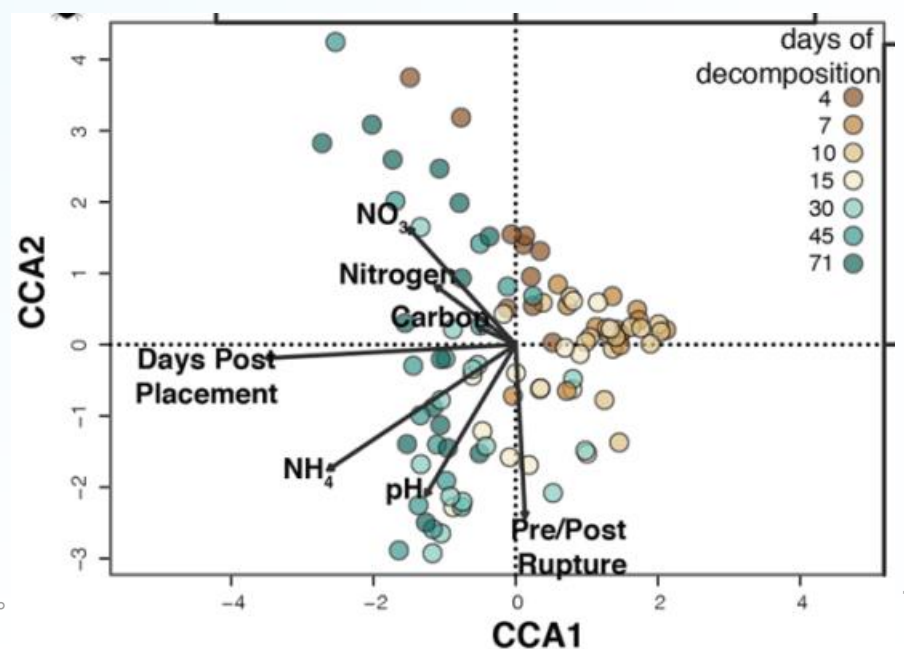




35环境因子分析

周欣 中科院微生物所
2022年1月9日



- 多样性概念介绍
- 相关性分析方法介绍
- 排序分析简介
- Vegan软件包简介
- 实战文献解读
- 实战数据准备与绘图

易生信 生信宝典 宏基因组



多样性的衡量尺度

α scale (within-habitat)

生境内的多样性

β scale (between- habitat)

生境间的多样性

γ scale (regional)

区域多样性

Pictures from Ecology,
Manuel Molles and Interne



- α 多样性：单个样本中的物种多样性；
- β 多样性：是研究不同生境间的群落之间的多样性及其它们之间的关系；
- γ 多样性：在一个地区或陆地生态系统的全部物种多样性的指标。 γ 多样性可以用与 α 多样性相同的单位来测量。

什么是群落

生物圈

Soil carbon storage



生态系统

Soil stability
Nutrient cycling



Conceptual & math.
models of Mycorrhizal
function

群落（不同物种）

Competition among plant species
Feedbacks between fungal and plant species
Interactions with herbivores and other soil biota



个体

Resource acquisition and
allocation by plants and fungi



群体（同一物种）

Competition among individuals
Hierarchies within plant populations



- 生物群落是指在相同时间聚集在同一地段上的各物种的种群集合，包括动物、植物、微生物等各个物种的种群，共同组成生态系统中有生命的部分。
- 组成群落的各种生物种群不是任意地拼凑在一起的，而有规律组合在一起才能形成一个稳定的群落。

群落与环境因子之间关系分析

1. 相关系数分析 (Correlation Coefficient) :

判定哪些环境因子与微生物群落相对丰度变化显著相关，表示两列数据之间的相关性。

2 (偏) 曼特尔检验 ((Partial) Mantel test) :

Mantel test分析主要环境因子对微生物群落的影响，是对两个矩阵相关关系的检验； Partial Mantel test考虑了环境因子之间自相关关系。

3 典范对应分析/冗余分析 (CCA/RDA) : 较直观的反应微生物主要菌群分别受哪些环境因子的影响。

4 方差分解分析 (Variance Partitioning Analysis (VPA)) : 定量分析不同环境因子对微生物群落变化影响的贡献率。

5 地理距离拟合分析 (Correlation with geographic distances)

相关系数分析 (Correlation Coefficient)

- 它反应的是一种线性相关系数，用来反映**某些特定微生物**类群相对丰度与环境因子的统计量。
- 如，在不同施氮肥影响微生物群落的实验中，某些特定微生物类群相对丰度的变化与哪些环境因子存在显著相关性关系？
- 可以用Past (PAleontological STatistics) 软件进行计算。

	pH	TN Proteobacteria
N0	7.66	0.25	25.57
N0	7.34	0.22	26.39
N0	7.30	0.22	24.38
N0	7.00	0.21	23.31
N5	5.73	0.27	31.81
N5	6.43	0.29	31.56
N5	6.62	0.23	28.09
N5	6.16	0.25	28.52
N28	5.73	0.27	43.05
N28	5.06	0.29	44.00

操作步骤：File==》 Import text file (Includes a first row/column of column/row labels, Skip top left cell; Separator: Comma) ==》 Select all data ==》 Statistics ==》 Correlation table (Linear correlation r) ==》 Result

	pH	NH4_N	Proteobacter
pH		0.0030725	0.00097443
NH4_N	-0.77494		0.0012224
Proteobacter	-0.82428	0.81555	

(偏) 曼特尔检验(Partial) Mantel test

曼特尔检验主要是对两个矩阵相关关系的检验，根据所求 **r值** 和显著性水平 **P值** 来对环境因子和微生物群落分布相关性有个大致了解。

➤ **Mantel test** : 常用于实验中探讨哪些环境变量影响了微生物的变化?

➤ **Partial mantel test**: 当这些环境变量间存在自相关关系时，需要控制其他变量

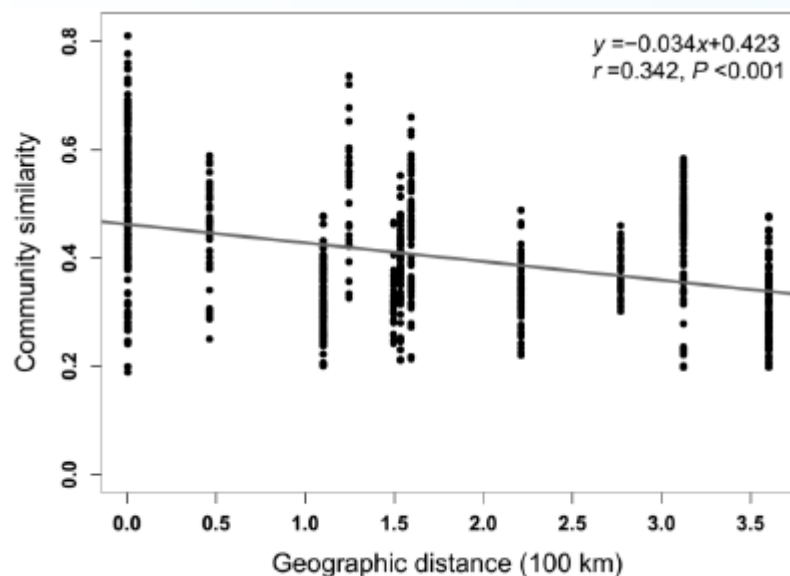
通常我们将微生物群落作为一个距离矩阵（如Bray-curtis），环境变量作为另一个距离矩阵（如pH、温度、土壤化学元素等），再检验两个矩阵之间的相关性。

```
otu_bray<-vegdist(otu, method="bray")
env_Salinity<-vegdist(env$Salinity, method="euclidean")
env_Depth<-vegdist(env$Depth, method="euclidean")
env_DO<-vegdist(env$DO, method="euclidean")
#做999次重复用来计算P值
mantel(otu_bray, env_Salinity, permutations=999)
mantel(otu_bray, env_Depth, permutations=999)
mantel(otu_bray, env_DO, permutations=999)
##当这些环境变量间存在自相关关系时，需要控制其他变量
mantel.partial(otu_bray, env_Salinity, env_Depth, permutations=999)
```



地理距离拟合分析 (Correlation with geographic distances)

- 先把bray相似矩阵算出来，来后把要画图的数据提出来，跟地理距离的值 geographic distance拟合曲线。
- 群落相似度与地理距离之间存在显著的负线性相关关系（如图，斜率=-0.034， $P<0.001$ ），Mantel检验进一步证明了群落结构与地理距离之间的显著关系（ $r=0.342$ ， $P<0.001$ ）



```
library(vegan)

dis<-read.table("meta.env.txt",sep = "\t", head =
T,row.names=1,check.names=FALSE, fill = T)

plot(dis$Altitude,dis$OTU_richness,xlab="Altitude",ylab="OTU_rich
ness") abline(lm(dis$OTU_richness~dis$Altitude))

##用下面的这条命令可以显示R值和P值

summary(lm(dis$OTU_richness~dis$Altitude))
```

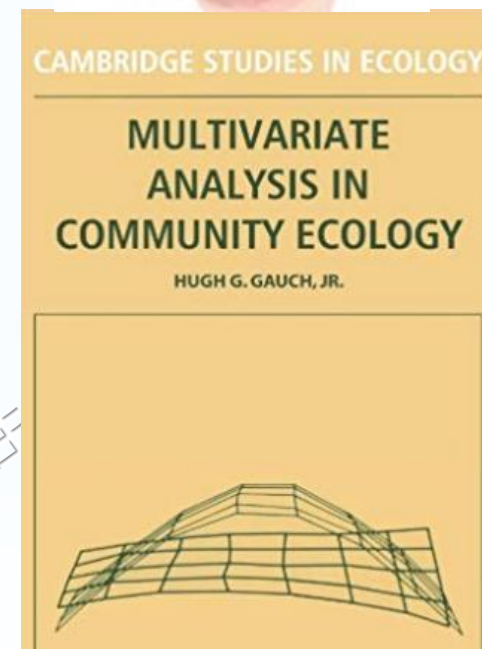

- 排序的过程就是在一个可视化的低维空间或平面重新排列这些样本,使得样本之间的距离最大程度地反映出平面散点图内样本之间的关系信息。
- **间接梯度排序(indirect gradient analysis)**: 又叫**非约束性排序**; 目标就是让尽可能多的变化量能够在尽可能少的轴上展示出来, 并可视化。**其在运算过程中没有受到外部环境因子变量的影响**。因此非约束性排序也只是探索性, 描述性的方法。典型的非约束排序有PCA, PCoA, NMDS, CA分析等。
- **直接梯度排序(direct gradient analysis)**: 又叫**约束性排序或者典范排序**; 约束性排序是明确地探索响应变量和解释变量矩阵的关系, 在特定的环境轴上探讨物种的变化情况; **发现物种在环境因子梯度上的变化情况**, 寻找能最大程度解释响应变量矩阵变差的一系列的解释变量的线性组合。典型的方法有RDA, CCA, db-RDA。
- **RDA就是PCA的约束排序版本, CCA是CA分析对应的约束排序分析方法。**

为什么用排序分析？

康奈尔大学的Gauch教授1982年在《群落生态多元分析》首先提出了提出排序分析的概念：“排序主要是尽力在低维空间中尽可能忠实地代表样本和物种关系”。

但是为什么要做排序分析呢？

- 生成可视化的排序图
- 结合聚类或回归等其它方法解释排序图中的数据结构和趋势；
- 能更好用统计学功能，通过关注重要维度来降噪，实际上排序也是一种很好的降噪手段；
- 可以确定不同梯度的相对重要性，提供了多种beta多样性分析手段，可以从图形结果直接解释物种和环境的关系。



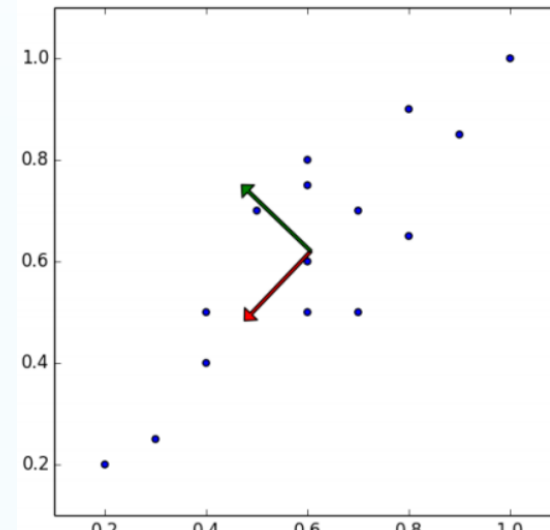
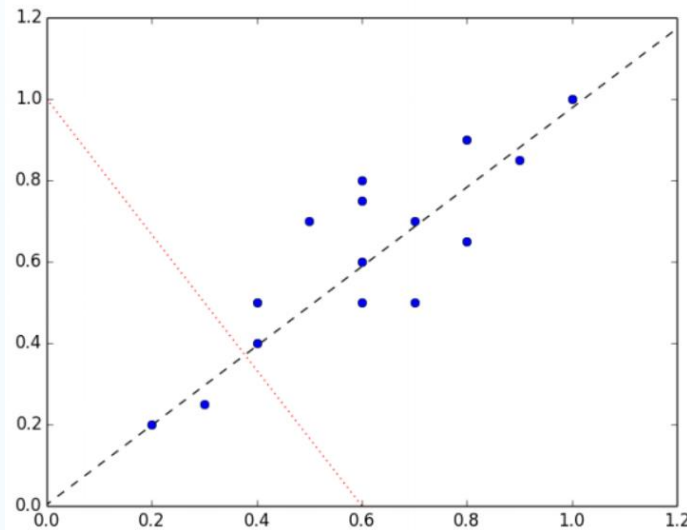
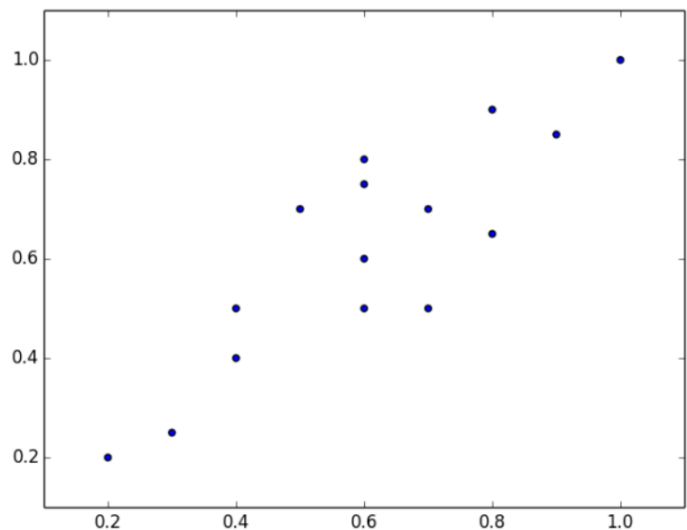
排序分析降维-举个栗子

假如你是一本养花工具宣传册的摄影师，你正在拍摄一个水壶。水壶是三维的，但是照片是二维的，为了更全面的把水壶展示给客户，你需要从不同角度拍几张图片。下图是你从水壶背面，正面，正上方，斜上方的照片：



易生信

假设有下图所示的数据集：



数据集看起来像一个从原点到右上角延伸的细长扁平的椭圆。要降低整个数据集的维度，我们必须把点映射成一条线。下图中的两条线都是数据集可以映射的，映射到哪条线样本变化最大？

显然，样本映射到黑色虚线的变化比映射到红色点线的变化要大的多。实际上，这条黑色虚线就是第一主成分。第二主成分必须与第一主成分正交。后面的每个主成分也会尽量多的保留剩下的变量，唯一的要求就是每一个主成分需和前面的主成分正交。

排序分析的方法

排序的过程是将样品或微生物物种排列在一定的空间，使得排序轴能够反映一定的生态梯度。

分析的方法包括：

- 主成分分析(PCA)
- 主坐标分析(PCoA)
- 非度量多维尺度法(NMDS)
- 典型相关分析(CCA)
- 冗余分析(RDA)

目的：

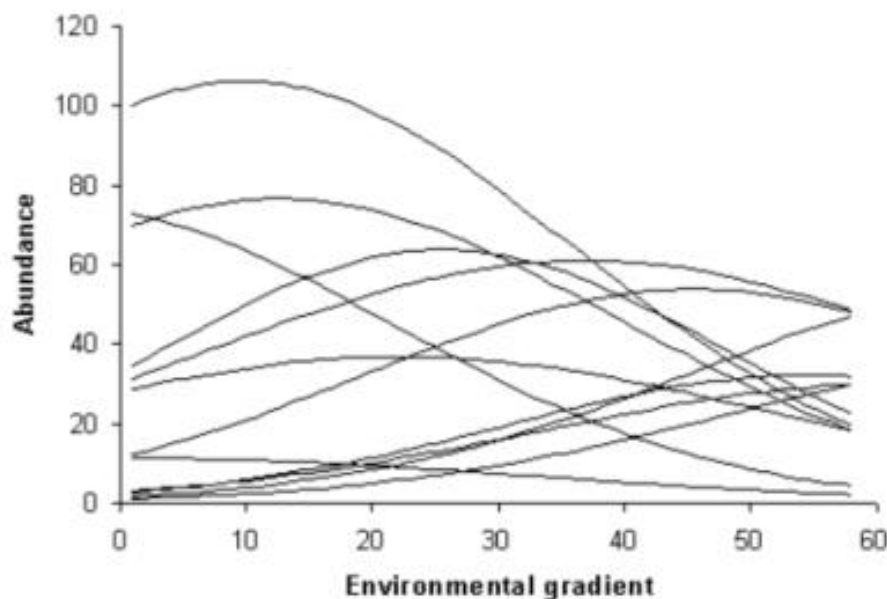
- A. 数据的降维
- B. 多维数据的展示
- C. 多变量间的关系（主成分的解释）
- D. 主成分基础上的回归



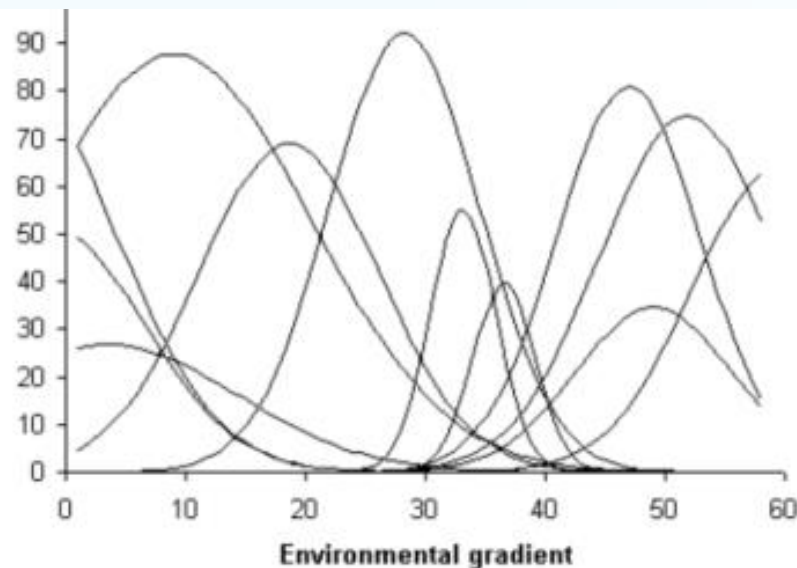
Harold Hotelling

典范排序：RDA/CCA

RDA或者CCA是基于对应分析发展而来的一种排序方法，**RDA基于线性模型，CCA则是基于单峰模型**。将对应分析与多元回归分析相结合，每一步计算均与环境因子进行回归，又称多元直接梯度分析。此分析是主要用来反映菌群与环境因子之间关系。



线性关系假设



单峰曲线假设

宏基因组
三信宝典

典范排序：RDA/CCA

- 并不是所有的环境变量对于微生物群落变化的解释都有显著性贡献，RDA和CCA分析的目的就是通过逐步迭代的方式**筛选出有显著解释性的环境变量**。RDA分析基于线性模型，而CCA分析基于单峰模型。具体还需结合实验结果确定适合的分析模型。
- **那么选择用RDA还是CCA分析？species-sample做DCA分析：decorana()**
 - 根据看分析结果中Axis Lengths的第一轴的大小：
 - 如果大于4.0,就应选CCA（基于单峰模型，典范对应分析）
 - 如果在3.0-4.0之间，选RDA和CCA均可
 - 如果小于3.0, RDA的结果会更合理（基于线性模型，冗余分析）



方差分解分析 (Variance Partitioning Analysis)

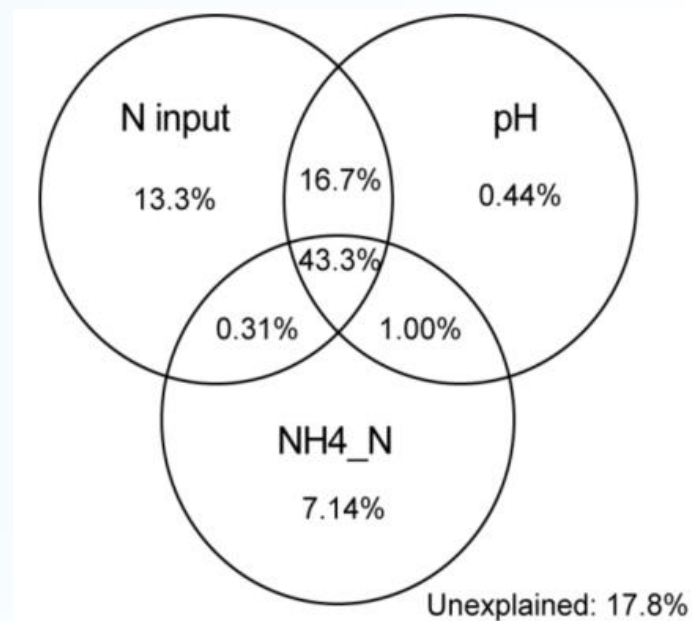
- VPA (Variance Partitioning Analysis), 可以定量地计算变量间的影响关系。也就是可以分析不同环境因子对环境中微生物群落变化的贡献度。

在施肥处理影响土壤微生物群落变化的实验中, 可以评价各环境因子变量对微生物群落变化的贡献度。

VPA分析结果显示: N肥施加量对微生物群落影响单独贡献作用可达13.3%; N肥施加量, 土壤pH和NH₄-N的共同作用最大, 可达43.3%。

方差分解分析 (Variance Partitioning Analysis)

```
VPA <- bioenv(otu,env) ##计算与群落变化最大的环境因子组合
VPA
summary(VPA)
mod <- varpart(otu,~DO,~Salinity,~Depth, data =env)
mod
plot(mod)
```



Vegan程序包介绍

- VEGAN软件包是芬兰Oulu大学Jari Oksanen教授为首的众多一流生态学者合作完成，专门用于群落生态学分析的工具(Oksanen et al., 2007/2010/2013)，多版本合计引用2万多次

[PDF] The **vegan** package

J Oksanen, R Kindt, P Legendre, B O'Hara... - Community ecology ..., 2007 - researchgate.net

Page 1. The **vegan** Package December 19, 2008 Title Community Ecology Package Version 1.15-1 Date December 19, 2008 Author Jari Oksanen, Roeland Kindt, Pierre Legendre, Bob O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Helene Wagner Maintainer ...

★ 77 Cited by 17316 Related articles All 51 versions



Jari Oksanen

University of Oulu · Department of Biology

37.22 · PhD

About

Network

Projects 1

Research 374

About

374

Research items

32,629

Reads

16,981

Citations

https://www.researchgate.net/profile/Jari_Oksanen

- 提供了全部基本排序（ordination）方法，包括(PCA, CA, PCoA和NMDS)
- 支持环境变量拟合和制作排序图，以及方差分解，生成的结果可以直接绘图或者进行其他分析，极为方便。很多人学习R的动力就是为了使用vegan软件！



文献解读1. 尸体降解过程中的微生物变化

Science

Home News Journals Topics Careers

Institution: Institute of Microbiology
Log in | My account | Contact Us
Institute of Microbiology

SHARE

REPORT



0

Microbial community assembly and metabolic function during mammalian corpse decomposition

Jessica L. Metcalf^{1,2,*}, Zhenjiang Zech Xu², Sophie Weiss³, Simon Lax^{4,5}, Will Van Treuren⁶, Embriette R. Hyde², Se Jin Son...

+ See all authors and affiliations

Science 10 Dec 2015;
aad2646

Microbial community assembly and metabolic function during mammalian corpse decomposition

JL Metcalf, ZZ Xu, S Weiss, S Lax... - ..., 2016 - science.sciencemag.org

Vertebrate corpse decomposition provides an important stage in nutrient cycling in most terrestrial habitats, yet microbially mediated processes are poorly understood. Here we combine deep microbial community characterization, community-level metabolic ...

★ 77 Cited by 178 Related articles All 12 versions

ARTICLE 1

Email

Print

Alerts

Citation

RELATED

背景

尸体降解是营养循环中重要的阶段，然而微生物介导的过程仍知之甚少。作者采用扩增子、功能预测和环境因素分析来揭示不同土壤类型下人类和小鼠的尸体分解过程。作者发现了一套细菌和真菌可形成重复的分解者网络和氮循环，并且可预测出现时间。并且分解者主要来自于土壤，这些结果可为法医学研究提供了新的研究手段。

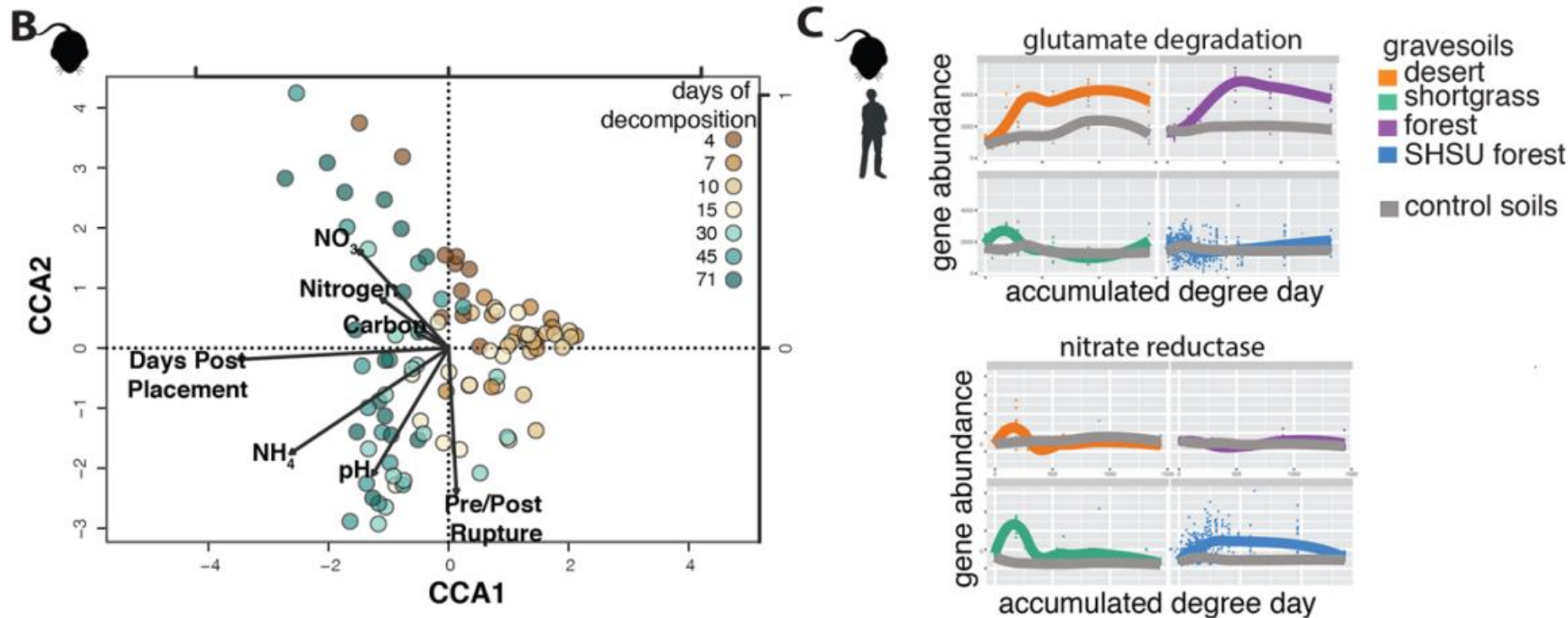
科学问题

降解微生物群落的来源是什么？

- 降解微生物群落的动态变化特征？
- 不同宿主/个体降解微生物群落的共性和个性？



尸体降解过程中的微生物变化



- 作者用了PICRUST预测功能的第三级进行CCA分析结合主要环境因子的分析；首先看到样本点在时间序列上的梯度变化，其次是环境因子与主轴间的关系、贡献大小和相关程度。死亡土中谷氨酸降解酸酶和亚硝酸还原代谢相关基因随时间变化。
- 使用死亡土的微生物群落功能变化来研究与死亡时间的关系。此处的研究为法医学提供了很大帮助，即使尸体被移走，也可以采用土壤的鉴定来判断第一现场，作案时间等信息。

文献解读2. 东海真菌多样性组成样式

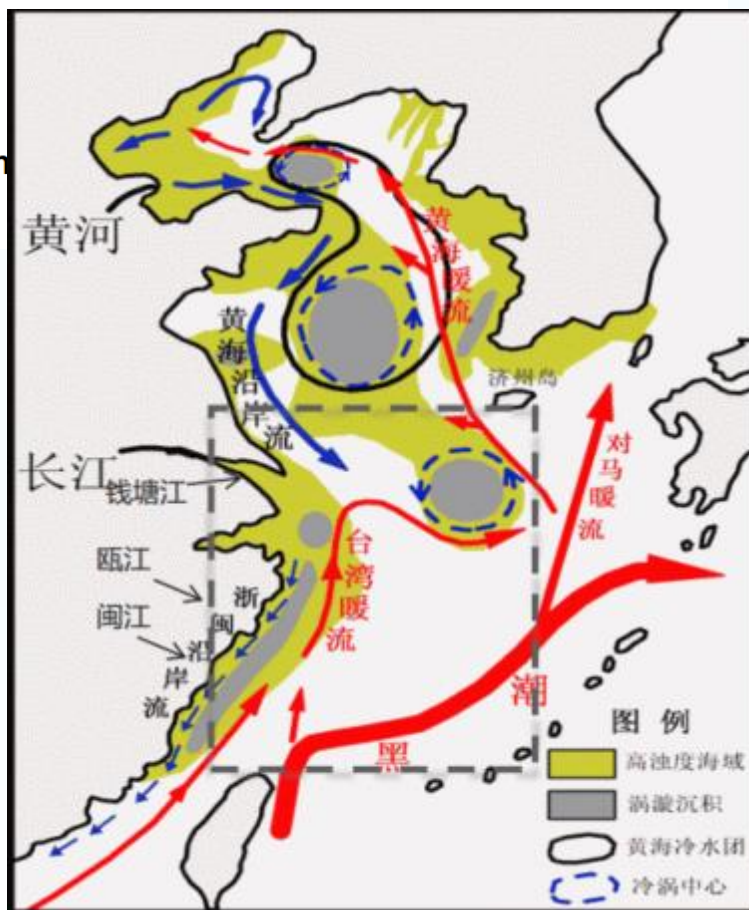
ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Highlighting patterns of fungal diversity and composition shaped by ocean currents using the East China Sea as a model

Wei Li¹ | Mengmeng Wang^{1,2} | Haoqin Pan¹ | Shengkang Liang⁵ | Jiajia Guo¹ | Tian Luo¹ |

Wei Li, Mengmeng Wang, Haoqin Pan, Gaëtan Burgaud, Shengkang Liang, Jiajia Guo, Tian Luo, Zhaoxia Li, Shoumei Zhang & Lei Cai. Highlighting patterns of fungal diversity and composition shaped by ocean currents using the East China Sea as a model. *Molecular Ecology* 27, 564-576, doi:10.1111/mec.14440 (2018).



- 沿岸流（CWM）：秋冬季向南流动、夏季向北流动。向海洋输送了大量陆源营养物质与微生物。

特点：低盐，高营养。

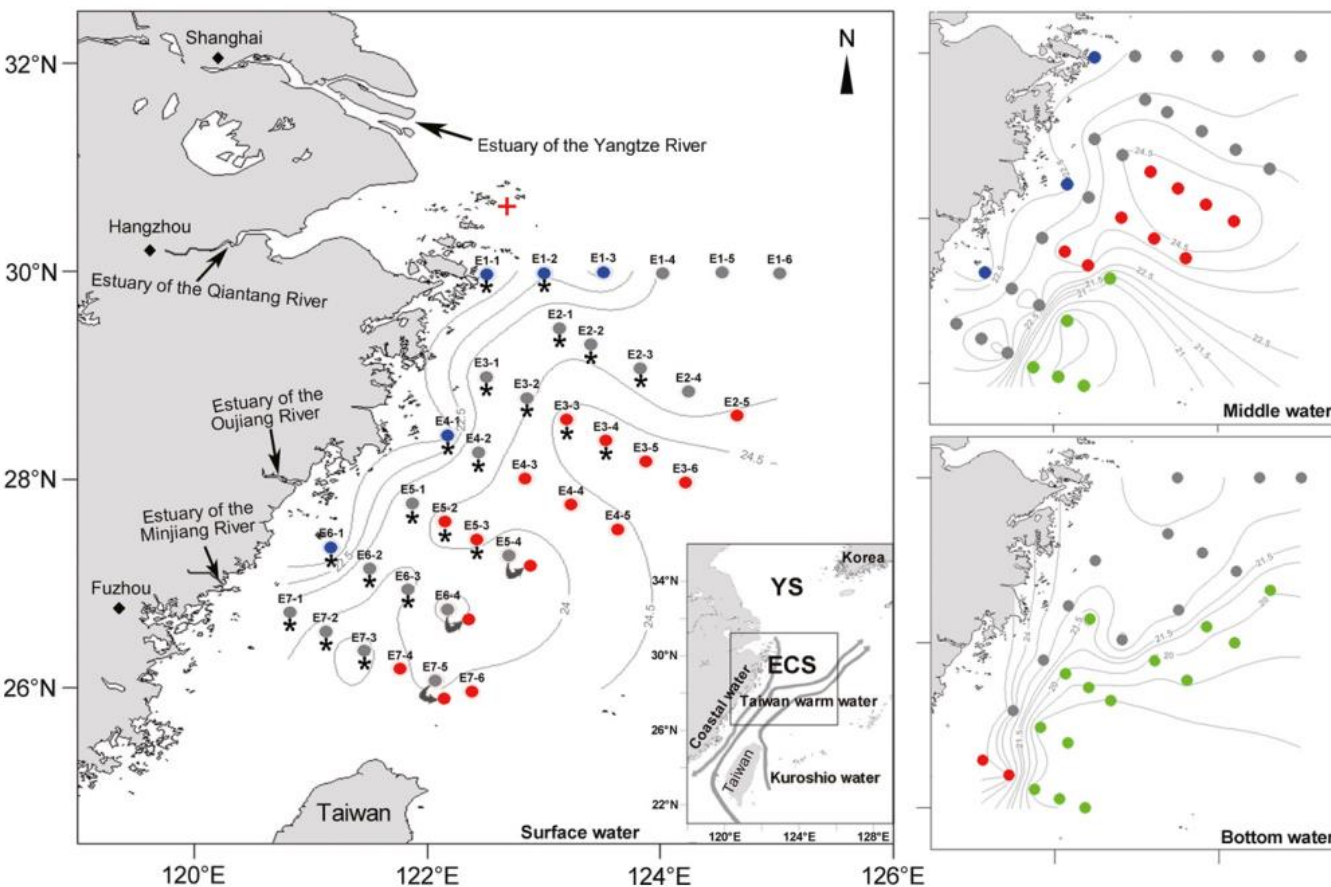
- 黑潮（KSWM）：世界海洋的第二大暖流。起源于菲律宾以东海域。对微生物长距离传播与扩散有重要影响，如原生动物。

特点：低温、高盐。

- 台湾暖流（TWM）：黑潮分支。

特点：高温、高盐。

样本位置与真菌群落结构



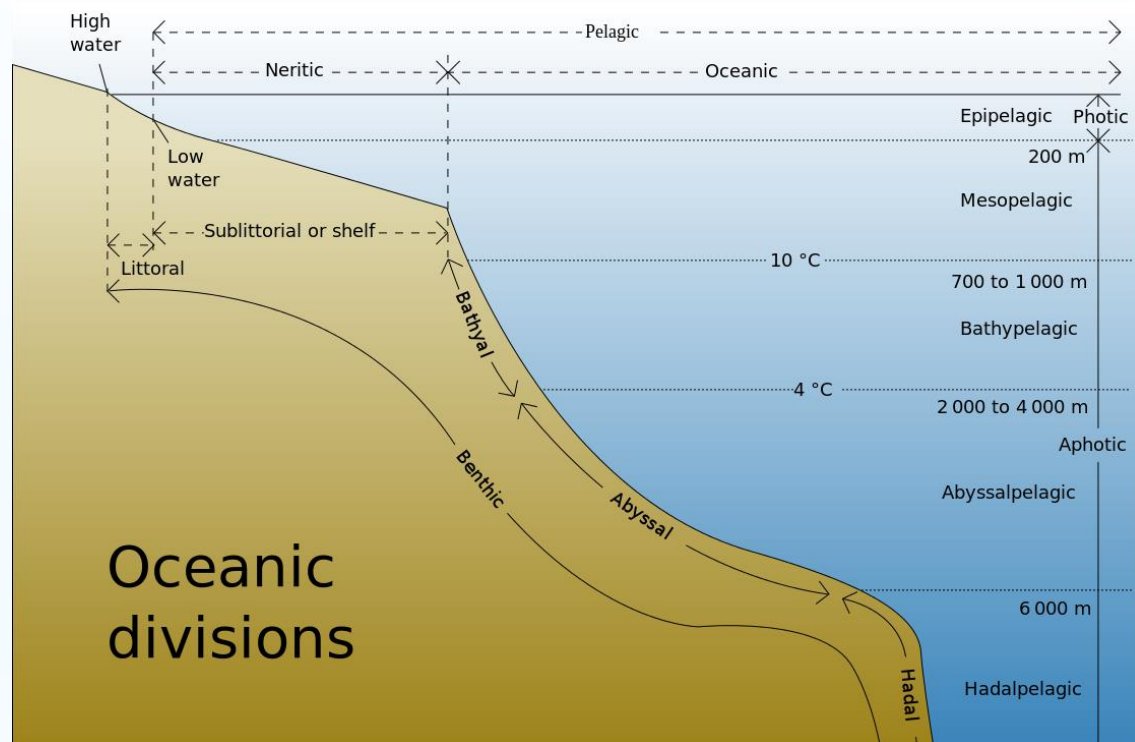
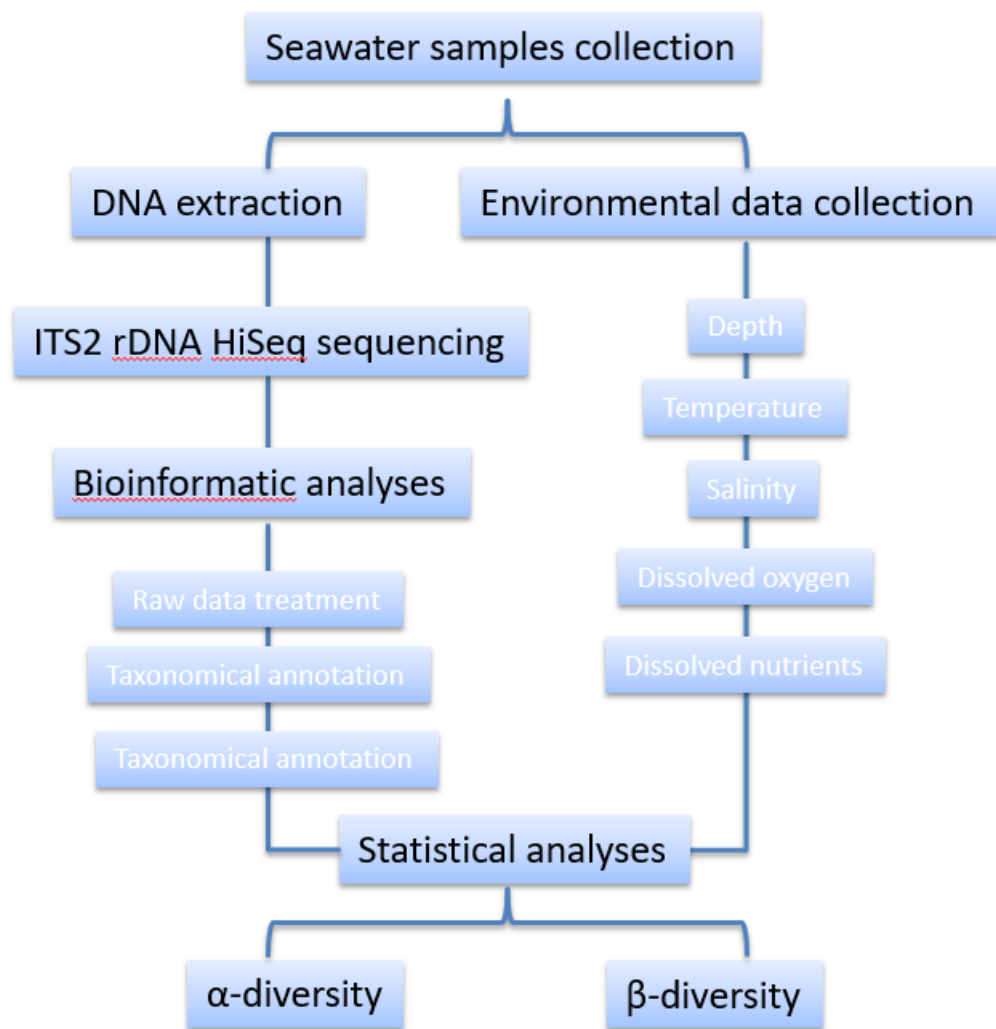
采样站位：36个

水样：100个（表层、中层、底层）

泥样：40个（表层和中层）；

沉积物及海水样品

- 每份水采集1L，采集表层;中层以及底层海水。



海水理化性质:

深度、温度、盐度、密度、溶解氧、浊度数据来自于船载温盐深测量仪。

溶解性营养盐 (NO_3^- 、 NO_2^- 、 NH_4^+ 、 PO_4^{3-} 、 SiO_3^{2-})，营养盐自动分析仪。

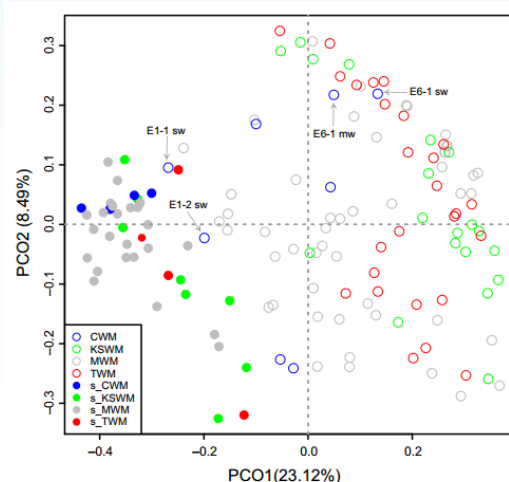
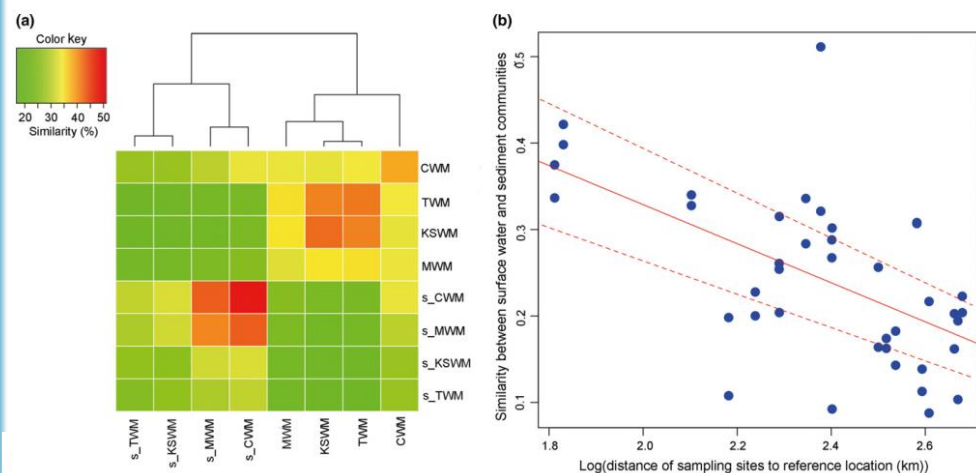
Environmental variable	Total water masses		CWM		MWM		TWM		KSWM	
	VE (%)	p	VE (%)	p	VE (%)	p	VE (%)	p	VE (%)	p
Depth	2.52	.001*	14.57	.175	3.31	.015*	5.47	.060	13.16	.006*
Temperature	—	—	—	—	5.78	.001*	—	—	—	—
Density	3.35	.001*	—	—	—	—	13.14	.001*	—	—
Salinity	—	—	—	—	—	—	—	—	—	—
Turbidity	2.19	.005*	—	—	6.74	.001*	5.90	.032*	3.51	.744
DO	3.39	.001*	23.23	.022*	6.56	.001*	3.34	.425	10.00	.033*
NO ₂ ⁻	1.10	.216	24.29	.017*	2.11	.231	6.30	.030*	—	—
NO ₃ ⁻	—	—	—	—	3.80	.006*	1.70	.989	—	—
PO ₄ ³⁻	—	—	—	—	1.96	.331	3.58	.307	—	—
SiO ₃ ²⁻	1.29	.117	—	—	2.55	.089	3.83	.253	—	—
NH ₄ ⁺	1.03	.313	—	—	2.13	0.227	3.49	.332	4.16	.588
Total explanation (%)	14.87		62.09		34.93		46.75		30.83	
ANOVA test of model	F = 2.30	p = .001*	F = 2.18	p = .001*	F = 2.21	p = .001*	F = 1.56	p = .002*	F = 1.56	p = .027*

海水及沉积物真菌群落存在极显著差异 ($R^2=0.191, p=0.001$) 不同水团对海水样品真菌群落组成影响显著 ($R^2=0.094, p=0.001$)

群落结构上, 沿岸流与大陆架混合水团和台湾暖流存在显著差异 ($p<0.05$); 与黑潮洋流海水样品间无显著差异 ($p > 0.05$)

从沿海到开放蓝水区域的水体真菌多样性区别明显, 来自长江, 闽江等河流的淡水真菌在沿岸河流尤其是长江冲淡水的输入, 可能会提高真菌的多样性。

可溶性氧以及深度是影响真菌群落的主要环境因素, 并且显著地被洋流影响。



- ## OTU物种矩阵表

环境因子矩阵表

022.8923 7.22

RDA/CCA实战：数据均一化

- 将环境因子进行 $\log_1p()$ 转化：（多度+1的自然对数，保证0转化后还是0）

Vegan中的`decostand()`函数可以提供多种生态学数据常用的标准化。

- 将OTU表进行hellinger转化：`decostand(otu.t, "hellinger")`



```
> otu.hell <- decostand(otu.t, "hellinger")
> sel <- decorana(otu.hell)
Warning message:
In decorana(otu.hell) :
  some species were removed because they were missing in the data
> sel
```

```
Call:
decorana(veg = otu.hell)
```

```
Detrended correspondence analysis with 26 segments.
Rescaling of axes with 4 iterations.
```

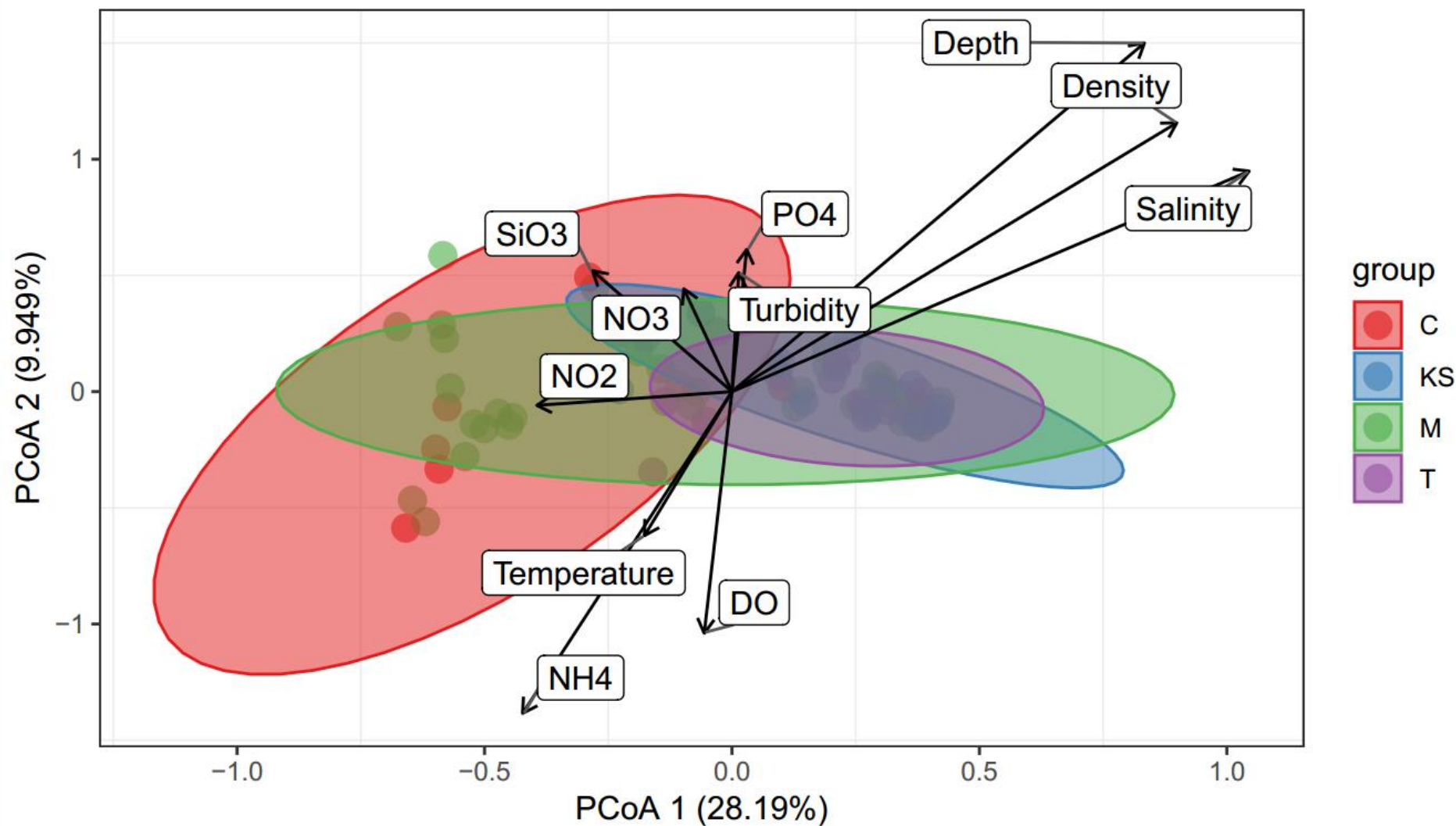
	DCA1	DCA2	DCA3	DCA4
Eigenvalues	0.2720	0.2630	0.1545	0.1586
Decorana values	0.5036	0.2666	0.2029	0.1474
Axis lengths	3.1896	2.8502	3.5283	2.1834

为什么要把原始数据进行转换？

因为我们的微生物多样性数据是一个随机分布的数据类型，而非正态分布，因而不能用均值标准差描述。这些公式都是基于在数据正态分布的条件下推导出来的，所以使用时，你的数据必须符合这些规则。

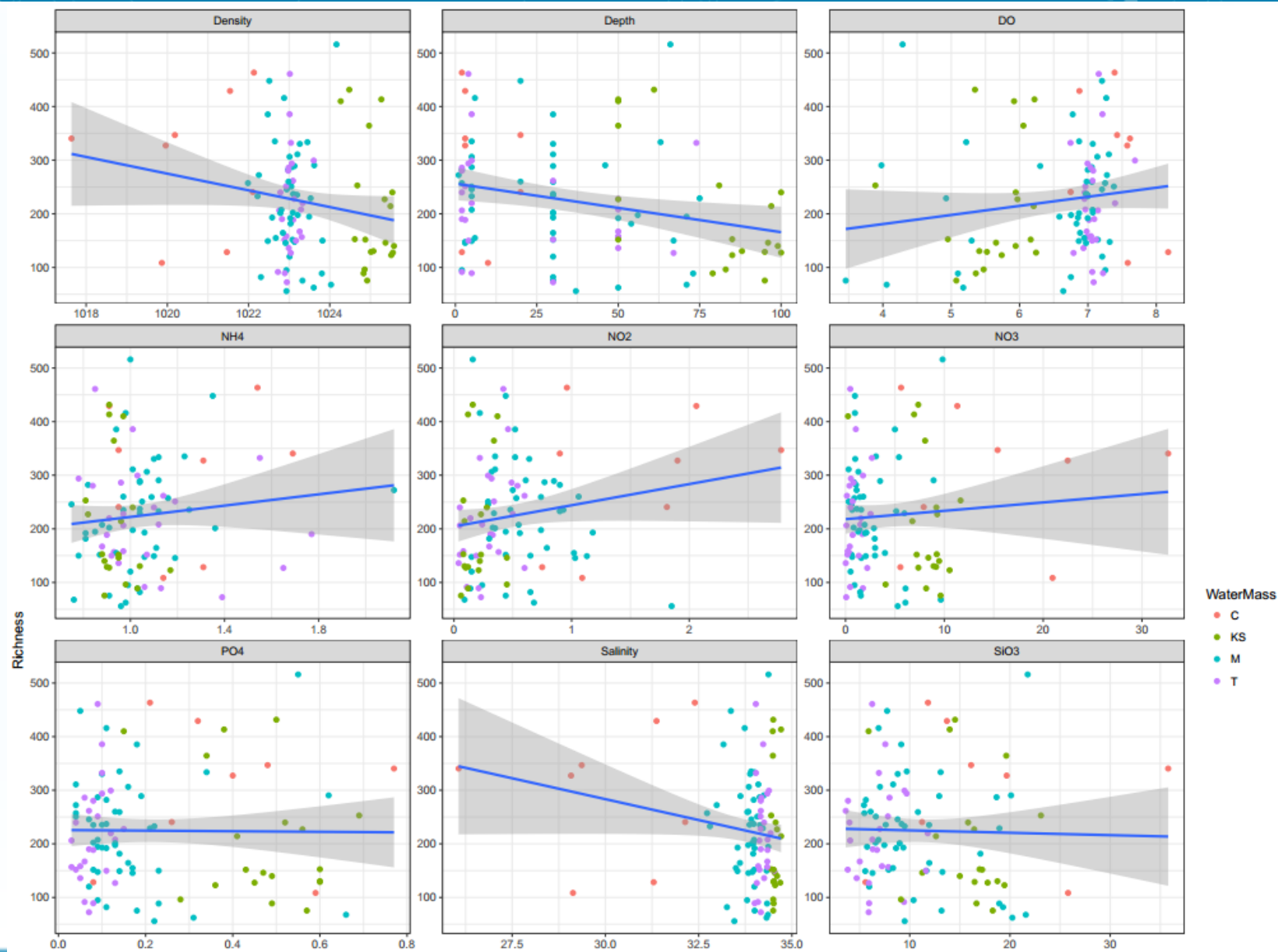


实战绘图展示. 环境因子与群落结构关系



信富興
宏基因组

实战绘图展示. 丰富度与各环境因子拟合



宏基因组



- ◆ 介绍了不同多样性指数的尺度，不同排序方法的比较。
- ◆ 环境因子分析的方法非常多，本课程介绍了几种比较常用的环境因子分析方法及差异统计方法。
- ◆ 具体关于环境因子分析的目的和描述还需要大家多结合自己的实验设计思路和所要讲述的故事多参考同行的文献，针对不同的项目和样品使用不同的分析方法。
- ◆ 最后大家一定要注意环境因子中的值一定不要有非数值型的值或者负数值，如果有可以考虑转换成相应的非负数数值。





扫码关注生信宝典，学习更多生信知识



扫码关注宏基因组，获取专业学习资料

易生信，没有难学的生信知识

