

罗宁生物云-操作指南

分析项目：ALPHA 多样性指数计算

版本号：V1.0

1. 分析项目简述：

通过计算样本的 Alpha-多样性指数可以反映微生物群落的丰度 (Richness)、均匀性 (Evenness) 及多样性等。常用来计算物种丰度或多样性的指数包括：

Chao1: Chao1 指数常用来估算物种总数；

Simpson: 用来估算微生物多样性指数之一，既考虑了物种的丰度 (Richness) 也考虑了均匀性 (Evenness)。这里的结果为 1-D (D 为 Simpson's Index)，值越高代表多样性越高；

Shannon: 香农-维纳指数(Shannon-Wiener index)，最常用来评估样本中微生物多样性。Shannon 是一种基于信息理论的测量指数，结果包含物种数及均匀性两种成分，值越大说明群落多样性越高；

Faith's PD: Faith's Phylogenetic Diversity²⁶是最常用的系统发育多样性度量方法，其计算方法是对进化树的所有枝长求和，因此需要用到进化树；

指标说明：其它指标及公式可参考 <http://scikit-bio.org/docs/latest/generated/skbio.diversity.alpha.html#module-skbio.diversity.alpha>。

2. 工具用途及使用场景：

本工具用于从原始 OUT 表统计计算群落在各分类水平上的相对丰度。

本工具计算输出结果文件可用于以下云工具的输入文件：

Alpha 多样性盒型图

Alpha 多样性 Duncan 检验

Alpha 多样性散点图

稀释曲线

等级丰度曲线 (Rank Abundance Curve)

3. 您需要准备以下文件：

3.1 如果测序样本无分组，请准备以下文件：

文件 1: otu_table_v1.csv

路径: \02.OTUAnalysis\Core_files

3.1.1 如果测序样本有分组，请再准备以下文件：

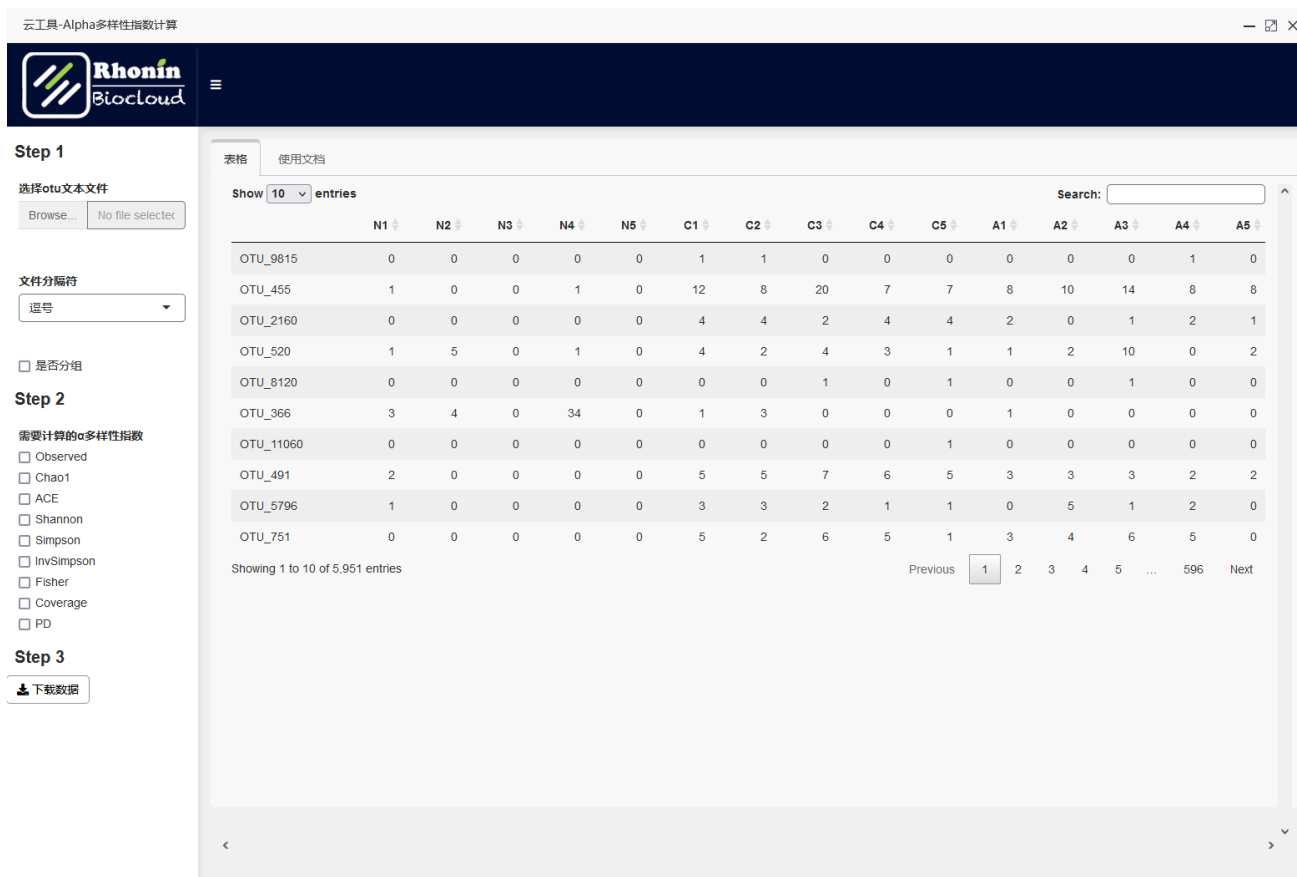
必须文件: sample_info_v1.csv

路径: \02.OTUAnalysis\Core_files

3.2 其他可选文件

无。

4. 操作界面：



群落组成计算界面左侧为文件输入及参数调整区域，右侧为输入文件实时展示区域。

5. 操作步骤：

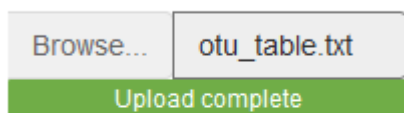
本项操作由数据整理上传、alpha 多样性指数选择选择和数据输出 3 部分组成。

第 1 步：数据整理上传

在 Step 1 区域的“上传 out 文本文件”入口点击浏览，选择 otu_table_v1.csv 文件（路径：\02.OTUAnalysis\Core_files），文件分隔符选择默认“逗号”，如下图所示：

Step 1

选择otu文本文件



文件分隔符



是否分组

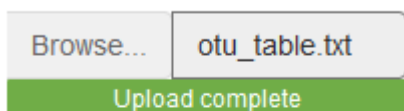
如果样品无分组，请接下来进行第 2 步操作。

注意：如果样品进行了分组分析，则需上传样品信息表，详细操作如下：

在完成了“上传 out 文本文件”导入 OTU 数据后，勾选“是否分组”复选框，在下方“选择分组文件”入口出选择导入 sample_info_v1.csv 文件（路径：\02.OTUAnalysis\Core_files），如下图所示：

Step 1

选择otu文本文件

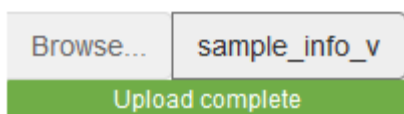


文件分隔符



是否分组

选择分组文件



完成以上操作后，请接下来进行第 2 步操作。

第 2 步：数据整理参数选择

在 Step 2 区域对导入数据进行 alpha 多样性指数进行选择，在此步骤可选择 1 个或多个指数。该步骤提供 9 项 alpha 多样性指数进行选择，分别是 Observed、Chao1、ACE、Shannon、Simpson、InvSimpson、Fisher、Coverage、PD，请根据后续展示需求调整所需参数。

以下对该步骤参数进行详细说明：

Alpha 多样性是指通过单样本的多样性分析反映样品内的微生物群落的丰富度和多样性，通常有测序深度指数（Observed species 和 Good's coverage）、菌群丰度（Chao1 和 ACE）和菌群多样性指数（shannon 和 simpson）。

a. Observed

Observed species，代表 OTUs 的直观统计。

b. Chao1

用 chao1 算法估计群落中含 OTU 数目的指数，chao1 在生态学中常用来估计物种总数，由 Chao (1984) 最早提出。Chao1 值越大代表物种总数越多。 $Schao1 = Sobs + n1(n1-1)/2(n2+1)$ ，其中 Schao1 为估计的 OTU 数，Sobs 为观测到的 OTU 数，n1 为只有一条序列的 OTU 数目，n2 为只有两条序列的 OTU 数目。Chao1 指数越大，表明群落的丰富度越高。

c. ACE

是用来估计群落中含有 OTU 数目的指数，同样由 Chao 提出(Chao and Yang, 1993)，是生态学中估计物种总数的常用指数之一。默认将序列量 10 以下的 OTU 都计算在内，从而估计群落中实际存在的物种数。ACE 指数越大，表明群落的丰富度越高。

d. Shannon

香农-维纳指数(Shannon-Wiener index)，最常用来评估样本中微生物多样性。Shannon 是一种基于信息理论的测量指数，结果包含物种数及均匀性两种成分，值越大说明群落多样性越高。

e. Simpson

用来估算微生物多样性指数之一，既考虑了物种的丰度 (Richness) 也考虑了均匀性 (Evenness)。这里的结果为 1-D (D 为 Simpson's Index)，值越高代表多样性越高。

f. InvSimpson

g. Fisher

h. Coverage

Good's coverage 指加入丰度为 1 的 OTUs 数目，加入低丰度影响

i. PD

Faith's PD: Faith's Phylogenetic Diversity, 是最常用的系统发育多样性度量方法，其计算方法是对进化树的所有枝长求和，因此需要用到进化树。如勾选 PD 复选框，则会出现进化树文件输入框，请在结题报告以下路径寻找并输入进化树文件： 02.OTUAnalysis\Core_files，文件名：rep_phylo_v1.tre。示例如下图所示：

Step 2

需要计算的 α 多样性指数

- Observed
- Chao1
- ACE
- Shannon
- Simpson
- InvSimpson
- Fisher
- Coverage
- PD

选择树文件

Browse... rep_phylo_v1.tre

Upload complete

第 3 步：数据输出

通过以上 2 个步骤进行数据导入及参数设置后，点击“下载数据”按钮导出计算结果数据，用于后续作图分析。

默认导出文件文件名：Alpha_diversity_data.csv。

未分组结果：

| | A | B | C | D | E | F | G | H | I | J | K | L |
|----|----|----------|----------|----------|----------|----------|----------|----------|-----------|----------|----------|----------|
| 1 | | Observed | Chao1 | se_Chao1 | ACE | se_ACE | Shannon | Simpson | InvSimpso | Fisher | Coverage | PD |
| 2 | N1 | 1235 | 2007.75 | 89.12498 | 2088.243 | 27.5572 | 3.580845 | 0.915404 | 11.82082 | 243.4032 | 0.98546 | 89.71251 |
| 3 | N2 | 1530 | 2245.624 | 80.76772 | 2214.401 | 25.34959 | 4.105052 | 0.905755 | 10.61064 | 309.2376 | 0.98675 | 104.3546 |
| 4 | N3 | 1019 | 1905.215 | 107.3545 | 1991.769 | 27.76945 | 2.320835 | 0.748823 | 3.981257 | 187.2901 | 0.98749 | 83.76591 |
| 5 | N4 | 1269 | 1869.326 | 70.04475 | 1973.501 | 25.7462 | 3.543064 | 0.879035 | 8.266853 | 255.911 | 0.985466 | 87.48753 |
| 6 | N5 | 1476 | 3558.463 | 205.4642 | 3871.936 | 42.30747 | 3.672437 | 0.906002 | 10.63848 | 350.9132 | 0.960336 | 108.8472 |
| 7 | C1 | 2474 | 5232.98 | 213.6192 | 5370.267 | 48.0617 | 5.273585 | 0.96831 | 31.55611 | 734.9728 | 0.931595 | 173.49 |
| 8 | C2 | 1508 | 2997.425 | 148.6043 | 3049.523 | 36.00587 | 5.114641 | 0.96498 | 28.55547 | 464.8287 | 0.928322 | 117.1774 |
| 9 | C3 | 1031 | 1967.257 | 115.0812 | 1970.695 | 27.38169 | 4.234734 | 0.942508 | 17.39387 | 264.5926 | 0.958157 | 81.66251 |
| 10 | C4 | 1627 | 3553.823 | 183.5143 | 3754.043 | 40.33072 | 4.722335 | 0.963098 | 27.09861 | 425.3139 | 0.950514 | 117.9665 |
| 11 | C5 | 1301 | 2277.626 | 105.1276 | 2395.35 | 30.14243 | 4.740532 | 0.963418 | 27.33551 | 364.2822 | 0.947665 | 102.0252 |
| 12 | A1 | 1509 | 3244.31 | 168.3639 | 3486.951 | 39.15169 | 4.204227 | 0.918518 | 12.27272 | 416.521 | 0.941238 | 118.0936 |
| 13 | A2 | 1250 | 2141.005 | 100.9823 | 2235.462 | 28.88362 | 4.713872 | 0.967571 | 30.83624 | 294.5526 | 0.970533 | 94.99168 |
| 14 | A3 | 1478 | 2491.675 | 104.4638 | 2561.094 | 30.58095 | 4.141219 | 0.935428 | 15.4865 | 348.9131 | 0.970091 | 112.6887 |
| 15 | A4 | 793 | 1469.813 | 89.39046 | 1621.631 | 26.49706 | 3.236383 | 0.816605 | 5.452705 | 196.5992 | 0.959464 | 68.72864 |
| 16 | A5 | 1307 | 2388.525 | 111.8081 | 2602.821 | 32.47245 | 2.52447 | 0.629149 | 2.696497 | 317.97 | 0.962349 | 106.7448 |

分组结果：

| | A | B | C | D | E | F | G |
|----|----|--------|-----------|---|----------|----------|----------|
| 1 | | Group | Measure | N | Mean | SD | SE |
| 2 | 1 | Group1 | ACE | 5 | 2427.97 | 812.8467 | 363.5161 |
| 3 | 2 | Group1 | Fisher | 5 | 269.351 | 62.91114 | 28.13472 |
| 4 | 3 | Group1 | Coverage | 5 | 0.9811 | 0.01164 | 0.005206 |
| 5 | 4 | Group1 | Chao1 | 5 | 2317.276 | 709.2142 | 317.1702 |
| 6 | 5 | Group1 | InvSimpso | 5 | 9.063612 | 3.120132 | 1.395365 |
| 7 | 6 | Group1 | Shannon | 5 | 3.444446 | 0.666935 | 0.298263 |
| 8 | 7 | Group1 | PD | 5 | 94.83355 | 11.06475 | 4.948305 |
| 9 | 8 | Group1 | Observed | 5 | 1305.8 | 204.8504 | 91.6119 |
| 10 | 9 | Group1 | Simpson | 5 | 0.871004 | 0.069636 | 0.031142 |
| 11 | 10 | Group2 | ACE | 5 | 3307.976 | 1336.041 | 597.4958 |
| 12 | 11 | Group2 | Fisher | 5 | 450.798 | 175.897 | 78.66354 |
| 13 | 12 | Group2 | Coverage | 5 | 0.943251 | 0.012779 | 0.005715 |
| 14 | 13 | Group2 | Chao1 | 5 | 3205.822 | 1291.268 | 577.4728 |
| 15 | 14 | Group2 | InvSimpso | 5 | 26.38791 | 5.331613 | 2.38437 |
| 16 | 15 | Group2 | Shannon | 5 | 4.817165 | 0.403393 | 0.180403 |
| 17 | 16 | Group2 | PD | 5 | 118.4643 | 34.11141 | 15.25508 |
| 18 | 17 | Group2 | Observed | 5 | 1588.2 | 544.4113 | 243.4681 |
| 19 | 18 | Group2 | Simpson | 5 | 0.960463 | 0.010247 | 0.004583 |
| 20 | 19 | Group3 | ACE | 5 | 2501.592 | 676.3644 | 302.4793 |
| 21 | 20 | Group3 | Fisher | 5 | 314.9112 | 80.47088 | 35.98767 |
| 22 | 21 | Group3 | Coverage | 5 | 0.960735 | 0.011915 | 0.005329 |
| 23 | 22 | Group3 | Chao1 | 5 | 2347.065 | 640.2562 | 286.3313 |
| 24 | 23 | Group3 | InvSimpso | 5 | 13.34893 | 11.0381 | 4.936387 |
| 25 | 24 | Group3 | Shannon | 5 | 3.764034 | 0.873665 | 0.390715 |
| 26 | 25 | Group3 | PD | 5 | 100.2495 | 19.59938 | 8.765109 |
| 27 | 26 | Group3 | Observed | 5 | 1267.4 | 287.0754 | 128.384 |
| 28 | 27 | Group3 | Simpson | 5 | 0.853454 | 0.13752 | 0.061501 |

注意：

导出文件可进行重命名，但必须是英文字符，如字符间需添加间隔可用半角下划线“_”，不可使用非英文字符、*、#、?、\$、\、/等特殊字符。

导出文件后缀为“*.csv”，不可更改此后缀。

6. 术语简述：

Alpha 多样性：也被称为生境多样性 (within-habitat diversity)，指一个特定区域或生态系统内的多样性，用于反映区域内物种丰富度和均匀度的指标。通常我们获得的一个样品中的物种多少、物种的分布是否均匀等指标被称之为这个样品的 Alpha 多样性，是这一个样品所代表的区域的本身属性。

Alpha 多样性指数：我们在描述一个区域或者样品的 Alpha 多样性时会使用一些计算方法计算出一个数值指

数来表征这个样品的 Alpha 多样性。常用的 Alpha 多样性指数包括：观测到的物种数目、Shannon 指数、Simpson 指数、Chao1 指数、ACE 指数、faith PD 指数等。云平台中提供的多样性指数及其说明如下：

Observed species:

物种丰富度指数 (Observed species) 为群落中丰度大于 0 的物种数之和，值越大表明群落中物种种类越丰富。物种丰富度指数在计算中对所有存在的物种 (无论优势物种或稀有物种) 等权重看待，只关注物种存在与否，与它们的相对丰度无关。此外，丰富度指数对抽样深度所造成的差异也非常敏感。

Chao1:

Chao1 指数在生态学中作为度量“物种丰富度”的指标，其值越高代表群落物种越丰富。Chao1 指数基于假设：当在群落中随机抽取个体时，若不断有新的物种被发现，则表明群落中尚存一些稀有物种还未被观测到；直到已经抽取到的所有物种均保证至少被抽到两次时，即未再出现新的物种被发现时，则可以认为该群落中的所有物种已经全部被观测到。据此可用于估算群落物种总数，且对稀有物种很敏感。

ACE:

ACE 指数在生态学中同样作为度量物种丰富度的指标，其值越高代表群落物种越丰富。在群落物种测量中，根据丰富/稀有物种 (根据指定阈值为界限) 以及仅包含 1 个个体的物种，计算 ACE 指数用于估算群落中尚未被观测到的物种数量。其值越大，代表该群落中真实物种种类越多。

Shannon

香农指数 (Shannon index) 同时考虑了物种丰富度以及均匀度。它反映了能够预测在群落中随机选择的个体属于哪些物种的不确定性。如果群落仅由单一物种组成 (种群)，那么随机选择的个体必定为那个唯一的物种，此时不确定性就为零；否则将无法得知随机被选择的个体究竟属于什么物种，并且不确定性也会随着群落物种种类数的增多而增加。但是，如果群落中存在一种或少数几种物种占据了优势地位 (与其它物种相比，它们在丰度上具有明显的优势)，那么不确定性就不会那么高。当群落完全均匀，即群落中所有物种丰度完全一致时，Shannon 指数的值达到最大。

Simpson

辛普森指数 (Simpson index) 同样考虑了物种丰富度以及均匀度，但与 Shannon 指数相比，它更受均匀度的影响。经典 Simpson 指数代表了在群落中两个随机选择的个体属于同一物种的概率，当群落物种丰富度增加时，这种概率降低，即 Simpson 指数随着物种丰富度的增加而降低。

invSimpson

由于经典 Simpson 指数与物种丰富度相反的趋势不直观，如今常用演变而来的 Gini-Simpson 指数表示 Simpson 指数，即用 1 减去经典 Simpson 指数数值后得到。此时 Simpson 指数随着丰富度的增加而增加 (二者保持一致的趋势)。

Fisher

Fisher 指数假设物种的丰富度符合对数系列分布，可以衡量物种多样性。

Coverage:

Coverage 常用在微生物 16S/18S/ITS 测序中，作为反映测序深度的指标。其值越接近于 1，说明测序深度越合理，测序深度已经基本覆盖到样品中所有的物种；反之，测序深度不高，许多物种仅被测到了一次，暗示着很多低丰度物种可能尚未被测序测到。

PD

谱系多样性 (Phylogenetic diversity, PD) 定义为在系统发育树上跨越给定的一组分类群所经过的所有系统发育分支的最小总长度。而在实际计算群落谱系多样性时,除了进化距离外,物种丰度数据也是必不可少的。并且可以预期,较大的 PD 值对应于更大的群落物种多样性。

7. 常见问题解答:

Q: 应该选择什么 Alpha 多样性指数?

A: 选择 Alpha 多样性指数有两个原则: 1、根据研究目的选择合适的。例如: 如果关注丰度的变化可以考虑选择跟丰度关系密切的多样性指数, 如 Chao1 指数。如更关注均匀度可以考虑 Simpson 指数等; 2、比较多个计算结果, 看哪种指数更能反映样品或分组的趋势。