

Q1.宏基因组与微生物多样性的区别？

宏基因组将基因组 DNA 随机打断成若干条小片段，然后在片段两端加通用引物进行 PCR 扩增测序，将得到的 reads 进行组装后进行基因预测得到基因序列，众多基因构成环境中微生物的基因集

微生物多样性主要针对核糖体小亚基基因序列进行测序（16s rDNA 或者 18s rDNA），该基因既存在高度保守的区域还包括高变区，通过特异性引物对某一段高变区（如 V3 区）或某几段高变区（如 V3-V4 区）进行扩增测序，然后与数据库比对，可特异性识别细菌种类。总的来说，微生物多样性主要告诉我们环境里有什么微生物，而宏基因组主要告诉我们环境里的微生物能做什么。

Q2.为什么宏基因组与微生物多样性物种注释结果存在差异？

微生物多样性和宏基因组都会分析环境中物种的种类，但是物种的丰度在二者之间存在一定的差异，第一种原因是微生物多样性是基于核糖体小亚基基因序列进行物种注释和丰度统计，但是在不同物种中核糖体基因的拷贝数不一样，这会带来一些误差；第二种原因是二者在建库测序过程中会进行 PCR 反应，在 PCR 这个过程中也会存在一些误差。

Q3.微生物多样性中 OTU 是指什么？

OTU (Operational Taxonomic Unit) 即分类操作单元，是在系统发生学研究或群体遗传学研究中，为了便于进行分析，人为给某一个分类单元（品系，种，属，分组等）设置的同一标志。在微生物多样性分析中，根据不同的相似度水平，对所有序列进行 OTU 划分，一般情况下，如果序列之间的相似性高于 97%（种水平）就可以把它定义为一个 OTU，每个 OTU 代表一个物种。

Q4. Beta 多样性四种距离算法的差异?

Beta 多样性分析主要采用 binary jaccard、bray curtis、unweighted Unifrac(限细菌)、weighted Unifrac (限细菌)等 4 种算法计算样品间的距离, 那么这四种算法都有什么差别呢?

4 种距离算法比较

	基于独立 OTU	基于系统发生树
加权	bray curtis	weighted UniFrac
非加权	binary jaccard	unweighted UniFrac

非加权的计算方法, 主要考虑的是物种的有无, 即如果两个群体的物种类型都一致, 表示两个群体的样本距离最小; 加权方法, 则同时考虑物种有无和物种丰度两个问题。比如如样品 A 由 3 个物种 a 和 2 个物种 b 组成, 样品 B 由 2 个物种 a 和 3 个物种 b 组成, 则通过非加权方法计算, 因为样品 A 与样品 B 的物种组成完全一致, 都只由物种 a 和 b 组成, 因此它们之间的样本距离为 0。但通过加权方法计算, 虽然样品 A 与样品 B 的物种组成一致, 但物种 a 和 b 的数目却不同, 因此两个群体的 β 多样性则并非一致。

基于独立 OUT 的方法认为 OTU 之间不存在进化上的联系, 每个 OTU 间的关系平等; 基于系统发生树计算的方法, 会根据 16s 的序列信息对 OTU 进行进化树分类, 因此不同 OTU 之间的距离实际上有“远近”之分。

Q5. PCA 与 PCoA 的区别?

主成分分析(Principal Component Analysis, PCA)是一种分析和简化数据集的技术, 通过将方差进行分解, 将多组数据的差异反映在二维坐标图上; 主坐标分析法(Principal coordinates analysis, PCoA)是一种与 PCA 类似的降维排序方

法。PCoA 与 PCA 的区别在于 PCA 是基于原始的物种组成矩阵所做的分析，使用的是欧式距离，仅仅比较的是物种丰度的不同，而 PCoA 首先根据不同的距离算法计算样品之间的距离，然后对距离矩阵进行处理，使图中点间的距离正好等于原来的差异数据，实现定性数据的定量转换。