#### 生物信息学分析

使用FastQC（Version 0.12.1）【[http://www.bioinformatics.babraham.ac.uk/projects/fastqc](http://www.bioinformatics.babraham.ac.uk/projects/fastqc" \t "https://www.sciencedirect.com/science/article/pii/_blank)】工具评估原始测序读长的质量，所有样品都通过了常用的质量标准。使用Fastp（Version 0.23.2）【PMID：38868435】，进行reads过滤修剪,所有参数设置为默认值。然后使用 BowTie2（Version 2.3.5.1） 【PMID：22388286】提取映射到人类基因组（GRch38 版本，从NCBI GenBank下载）的所有读数来去除人类DNA序列。

使用Kraken2（Version 2.1.3）【PMID：24580807，PMID：[31779668](https://pubmed.ncbi.nlm.nih.gov/31779668/)】软件对成对的非宿主读数进行分类并分配分类标记 [。](https://www.sciencedirect.com/science/article/pii/S2213716519302024?via=ihub" \l "bib0100)读数 （基于k-mer） 与预构建的Kraken2\_DB\_standard数据库进行比对，Kraken2为宏[基因组](https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/metagenomics" \o "Learn more about metagenomic from ScienceDirect's AI-generated Topic Pages)读数的分配最特异性的分类标签进行分类。但是，有时标签分配不在物种级别。特别是在多个物种中发现几乎相同的读数的情况下，Kraken2会将其分配给最低的共同祖先，这可能是属级别或更高。因此，在包含一个或多个高度相似物种的样品中，物种特异性读数的数量会被低估。为了克服这个问题，使用 Bracken（Version 2.8）【doi：10.7717/peerj-cs.104】（贝叶斯重新估计Kraken2分类后丰度）重新估计了 Kraken2分类后的[物种丰度](https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/population-abundance" \o "Learn more about species abundance from ScienceDirect's AI-generated Topic Pages)[。](https://www.sciencedirect.com/science/article/pii/S2213716519302024?via=ihub" \l "bib0105)

* 使用megahit（Version 1.2.9）【PMID：25609793】组装由fastp质量过滤的读数。使用 Prodigal（Version 2.6.3）【PMID：20211023】对组装的重叠群进行功能预测，使用KEGG数据库【PMID：10592173】进行功能注释。使用 quast（Version 5.0.2）【PMID：23422339】评估基因组组装。

#### 宏基因组统计分析

为了展现样本中的物种组成情况，使用 RStudio 和 ggplot2 库作物种组成柱状图，物种丰度改变diff图。使用VennDiagram库作物种种类Venn图。

使用RStudio和vegan,ggplot2库计算alpha多样性（Shannon，Simpson，Richness）指数,使用wilcox.test分析两组数据的均值是否有差异。以及计算beta多样性（非度量多维排列NMDS,PCoA主坐标分析,基于Bray-Curtis相似性的ANOSIM检验）进行组间比较。

使用RStudio和ggplot2,DESeq2，EnhancedVolcano，pheatmap库进行差异物种分析和差异基因分析，作火山图和热图。并使用microeco库，使用线性判别分析效应大小（LEfSe）将样本组相互比较，以计算哪些分类群（如果有）如何区分组，显著性P< 0.05，LDA 评分> 2.5（[http://huttenhower.sph.harvard.edu/lefse/](http://huttenhower.sph.harvard.edu/lefse/" \t "https://pmc.ncbi.nlm.nih.gov/articles/PMC9134649/_blank)）。作LDA score图，丰度图和系统发育图。使用clusterProfiler库进行差异功能基因富集分析，作富集通路柱状图。

使用RStudio和psych库计算差异物种，差异代谢物之间的相关性、p值等信息，使用pheatmap库作相关性热图，使用igraph库作网络图并使用Gephi软件进行优化。