大豆根系微生物的基因鉴定与基因功能预测

——竞赛单元“2-1基因组”

团队名称：水哥微生物小队

指导老师：郑金水

团队成员：张敦彪，颜旭，张子栋，姚代洪

摘要

高通量培养与鉴定可用于从给定环境和植物物种的本地根微生物组样本中建立一个分类学上全面的细菌培养收集，通过16S rRNA基因扩增子测序对根样品的细菌多样性进行评价，筛选出具有相应16S rRNA基因序列的培养菌。通过对大豆根系微生物的高通量培养筛选可以鉴定其基因的类型，进而对其基因功能进行预测。

关键字：高通量培养，基因鉴定，功能预测

1. 引言
   1. 根系微生物群

植物根系组成了一个分类结构的微生物群落，称为根系微生物群。基于标记基因的扩增子图谱和宏基因组测序已被用于描述包括模式和作物物种在内的多种植物根系微生物区系的分类组成和基因含量。最近的功能和机制研究揭示了根共生微生物为宿主提供的服务例如，根部细菌微生物群保护植物免受土壤传播的真菌和卵菌病原体的侵害并参与营养素使用寄主植物选择性地调节其根系微生物群，而根系微生物群反过来又影响植物的各种生理过程从根系微生物群中培养出来的微生物是这些功能和机理研究的核心资源。

为了研究根系微生物群在特定环境中的功能，必须从同一环境（如同一土壤类型）中生长的植物根系样品中分离和培养微生物。 虽然已经分离出数千种培养的微生物，并储存在国际资源中心，但这些微生物来自各种各样的环境和宿主物种。此外，微生物适应预计将有助于从不同生境和宿主取样的微生物之间的基因型和功能分化。 由于这些原因，通常需要从给定的自然土壤和寄主物种组合中创造新的细菌种群，以解剖该系统中植物和根系微生物群之间的相互作用。

* 1. 高通量培养

植物根系微生物组的研究主要依赖于高通量扩增子和宏基因组测序技术，对微生物组的物种分类和基因组成进行描述。高通量就是一代测序一次测序只能够对1个基因进行测序，而高通量测序能够一次检测几十甚至几百个基因。与其他方法相比不需要昂贵的设备、比菌落挑选途径产量更高、避免了使用菌落挑选反复获得快速生长细菌的问题（防止生长快的细菌抑制生长缓慢的细菌生长）、two-sided barcode PCR system（双面条码技术PCR系统）比之前使用454焦磷酸测序的识别过程更加准确、高通量。局限性：只适用于细菌，不适合丝状真核生物、不适合与不在液体培养基中生产的细菌一起使用、不适合培养厌氧细菌。

1. 材料与方法
   1. 培养基的制作

实验中需要三种不同的培养基：

TSB 培养基：30 g/L TSB ，蒸馏水定容。

10% TSB培养基： TSB 培养基按照体积分数10% 进行稀释。

10% TSB培养基＋丙酮酸钠：30 g/L TSB 、10 mM/L丙酮酸钠，蒸馏水定容。

三种培养基均高压蒸汽灭菌。

* 1. 大豆根系菌群的获取

将在超菌工作台上得到的根内样本与根际样本均置于4℃ 储存。

* 1. 分液

目的：将培养基分液到384孔板，每个孔7μL。 μ

* 1. 流式细胞分选

目的：将384孔板中的每一个孔中加入1个细菌，便于后续培养。

* 1. PCR处理与电泳鉴定

将得到的八连管每个孔中按比例加入1 F引物、1μl R引物、7μl 无菌水、10μl mix与原1μl菌液形成20μl体系,按照对应程序（如表1）将八连管放入pcr仪器处理。

|  |  |  |
| --- | --- | --- |
| 预热 |  | 98℃ 10min |
| 循环  (35x) | 变性 | 95℃ 30s |
| 复性 | 55℃ 30s |
| 延伸 | 72℃ 1min30s |
| 延伸完整 |  | 72℃ 5min |

**表1 PCR程序设定**

实验采用双面条形码PCR引物设计，用799F和1193R引物测定了培养菌V5-V7基因区的16SrRNA基因序列，并将其与根样本相同可变区的16SrRNA基因扩增子测序产生的16SrRNA基因序列进行了交叉对照。799F引物不扩增植物叶绿体DNA，与细菌16S rRNA基因相比，从植物线粒体DNA中产生更大的PCR产物，因此广泛用于排除植物质体的扩增为了对培养菌进行高通量鉴定。

通过PCR后的产物电泳接过与DNAMarker进行对比来检验PCR产物是否合格。

* 1. 产物的回收与浓度测定
     1. 产物的回收

1. 短暂离心 PCR 产物
2. 移液枪测体积，转移至灭菌了的2ml离心管
3. 加入 Buffer DP ，混匀15S
4. >l00 bp 加入2-3倍 buffer DP
5. <100 bp 加入2-3倍 buffer DP和1倍异丙醇
6. 离心收集管壁液滴
7. HiPure柱子套在收集管中，混合液转移至柱子，10000xg离心15-60s,
8. 若混合液＞700ul，倒掉滤液，重复步骤5
9. （柱子吸附量有限）5-6次重复以后则需进行步骤6
10. 倒掉滤液，柱子套到回收管中，加入500ul Buffer DW2,10000xg，离心15-60s
11. 倒掉滤液，套在回收等中，10000g离心2min
12. 柱子套在2mL的离心管中，加入7-30uL Elution Buff至柱子的膜中央，静置1min,
13. 10000xg离心1min ［若要产率高可重复步骤7]
14. 得到产物，-20℃储存
    * 1. 浓度测定

使用分光光度计来测定得到产物的浓度，测定浓度的目的是便于后续按浓度比例混合得到的产物（如400ng/μl加入5ul，则100ng/μl加入20μl）

* 1. 测序

本次实验采用的三代测序原理：DNA聚合酶和模板结合,4色荧光标记 4 种碱基（即是dNTP）,在碱基配对阶段,不同碱基的加入,会发出不同光,根据光的波长与峰值可判断进入的碱基类型。三代测序其实就是对二代测序的一个升级，简单来说就是它同样一次能测好多序列，但是测序的长度达到了10kb左右，并且不需要PCR富集序列，直接测序，这就解决了信息的丢失，以及碱基错配的问题。

* 1. 高通量培养与鉴定
     1. 使用高通量培养与鉴定的原因

在用菌落采摘和桑格测序等传统方法分离和鉴定根相关细菌时，主要面临着三个困难：(i) 使用合理的时间和成本获得不同生长速度的细菌。 例如，快速生长的细菌可能会抑制生长缓慢者的生长，通常导致生长迅速者优先被隔离。

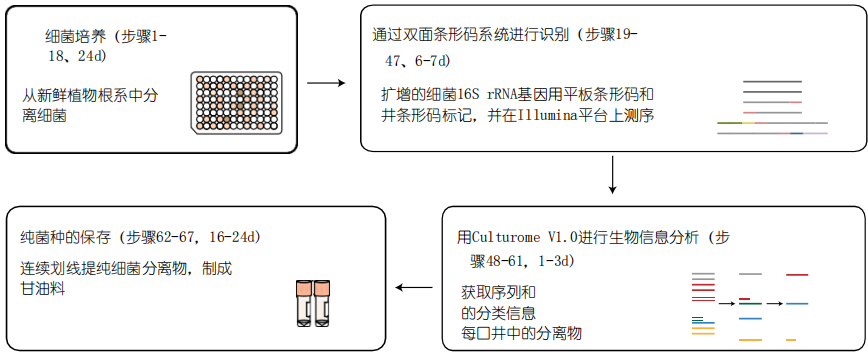
(ii) 高通量鉴定纯细菌或与其他细菌相关的培养细菌。 Sanger测序需要克隆培养的纯DNA模板，从密切相关的细菌群中获得序列信息效率低。

(iii) 高效处理和评估细菌鉴定数据。

* + 1. 意义与优点

对比以前研究根系微生物群的功能和机制的方法更简单和有效：

1. 可以使用普通的实验室设备进行，不需要使用昂贵的设备，如荧光激活细胞 分选仪器。
2. 分离程序可以在一天内完成，而菌落采摘可能需要连续几周的菌落分离，这取 决于菌落生长速度。
3. 避免了使用菌落采摘法重复获得快速生长菌的问题。用液体培养基对根系微生 物样本进行有限稀释（在96孔细胞培养板中分离细菌）细菌是在不同的孔中分离的， 从而防止快速生长的细菌抑制缓慢生长的细菌生长。
4. 双面条形码PCR系统基于Illumina测序技术，比以往采用454焦磷酸测序的鉴 定过程更准确，即通过避免单面条形码设计中获得的嵌合序列的形成来提高测序深度 和准确性。



**图1 高通量细菌培养鉴定系统概述**

* 1. 16s rRNA分析

16S rRNA 是原核核糖体小亚基的组成部分。原核核糖体的两个亚基是50S大亚基和30S小亚基。它们形成 70S 核糖体。小亚基由与 21 种蛋白质结合的 16S rRNA 组成。

* + 1. 优点

1. 16s rRNA基因分布广泛；
2. 16s rRNA基因序列丰度超过其他基因；
3. 可用于测定不同类群之间的系统发育关系；
4. 当扩增和测序成本负担得起时，水平基因转移不是大问题；
5. 成本较低。
   * 1. 原因
6. 16S rRNA基因是细菌基因组中普遍存在的基因。 由于 16S rRNA 功能在翻译过 程中对细菌细胞至关重要，因此几乎所有的细菌基因组都由 16S rRNA 基因组成。
7. 16S rRNA基因序列高度保守。 由于16S rRNA的功能更为普遍，因此16S rRNA 基因序列高度保守, 基因序列的变化可以被认为是时间的测量。
8. 16S rRNA 基因的大小 (1550 bp) 足以用于生物信息学目的。
9. 16S rRNA 基因是细菌基因组中被充分研究的基因。 由于 16S rRNA 基因的功 能对细胞至关重要，因此对其进行了许多研究。
10. 16S rRNA 序列也由可变区域组成，允许鉴定细菌种类。
    1. 基因功能预测
       1. 方法

主要是应用生物信息学分析软件或数据库进行结构和功能的预测：

1. ORF/CDS预测

不管真核还是原核都需要首先进行ORF预测。真核基因由于有内含子，必须使用专用的预测软件。原核的是直接找ATG和终止密码子。

1. Blast/序列比对工具使用

Blast/序列比对工具使用Blast的主要目的是找到相似性比较高的蛋白序列，这些蛋白的功能相似或相同的可能性更高。往往一致性或相似性要达到一定的阈值才行。

* + 1. 意义

把物种的“身份”和它们的“功能”对应起来。根据菌群代谢功能预测结果，我们一方面能一窥菌群功能谱的概貌，发挥菌群多样性组成谱测序性价比高的优势；另一方面也能帮助指导后续宏基因组De novo鸟枪法测序的实验设计，更合理地筛选用于后续研究的样本。

参考文献

1. Jingying Z,Xin Y L,Xiaoxuan G, et al. High-throughput cultivation and identification of bacteria from the plant root microbiota[J]. Nature Protocols,2021,16(2).
2. Nadav B,Grant G,H C W, et al. Genome-wide prediction of disease variant effects with a deep protein language model.[J]. Nature genetics,2023.
3. 李亭玉,李太元.益生菌发酵工艺中蜡样芽孢杆菌的分离鉴定[J].当代畜禽养殖业,2019(06):12-13.
4. Aysun U,BiancaMaria C,L A M, et al. SAP: Synteny-aware gene function prediction for bacteria using protein embeddings.[J]. bioRxiv : the preprint server for biology,2023.