

团 体 标 准

T/SZAS 22—2020

基于全基因组测序的益生菌菌株分型鉴定 指南

The guide to genotyping of the probiotics at strain level by whole
genome sequencing

2020-11-04 发布

2020-12-05 实施

深圳市标准化协会 发布

目 次

前言	II
1 范围	1
2 规范性引用文件	1
3 术语与定义	1
4 缩略语	2
5 鉴定原理	2
6 实验条件	2
7 实验步骤	2
8 数据分析	4
9 报告要求	5
10 质量控制	5
11 方法局限性	5
附 录 A（资料性附录） 常见分析软件和参数	6
附 录 B（资料性附录） 常见数据库	7
参考文献	8

前 言

本文件按照GB/T 1.1-2020给出的规则起草。

本文件由深圳华大生命科学研究院提出。

本文件由深圳市标准化协会归口。

本文件主要起草单位：深圳华大生命科学研究院、中国食品发酵工业研究院有限公司、内蒙古蒙牛乳业（集团）股份有限公司、武汉工控工业技术研究院有限公司、深圳万和制药有限公司、深圳华大基因科技有限公司。

本文件主要起草人：赵姣、黄锦群、井晓欢、何旭珩、孙建波、李启沅、姚粟、程坤、康小红、王国宏、马俊、刘峰、付晓超、王博、王韧、吴昊、李倩一。

基于全基因组测序的益生菌菌株分型鉴定指南

1 范围

本文件提供了基于全基因组测序（WGS）的益生菌菌株分型鉴定的原理、实验条件、实验步骤、数据分析和报告的指南。

本文件用于指导食品、药品、保健品、化妆品和饲料中包含一种或多种不同的益生菌的菌株鉴定。适用于相关的检测机构或研究机构等。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

- GB 4789.35 食品安全国家标准 食品微生物学检验 乳酸菌检验
- GB/T 6682 分析实验室用水规格和实验方法
- GB 19489-2008 实验室生物安全通用要求
- GB/T 19495.2-2004 转基因产品检测实验室技术要求
- GB/T 19495.3-2004 转基因产品检测 核酸提取纯化方法
- GB/T 19495.5-2018 转基因产品检测核酸定量 PCR检测方法
- GB/T 29859-2013 生物信息学术语
- GB/T 30099-2013 实验室离心机通用技术条件
- GB/T 30989-2014 高通量基因测序技术规程
- GB/T 35537-2017 高通量基因测序结果评价要求
- SN/T 2102.2-2008 食源性原体PCR检测技术规范 第2部分：PCR仪性能试验要求
- SN/T 2497.21-2010 进出口危险化学品安全试验方法 第21部分：琼脂糖凝胶电泳试验
- YY/T 1539-2017 医用洁净工作台

3 术语与定义

下列术语和定义适用于本文件。

3.1

益生菌 probiotics

活的微生物，当摄取足够量时，对宿主健康有益。

3.2

菌株 strain

同种微生物不同来源的纯培养物。

3.3

基因组genome

一个生物体所包含的DNA（部分病毒是RNA）的全部遗传信息。

3.4

序列比对 sequence alignment

比较两个或两个以上核苷酸或者氨基酸序列间的相似性的过程。

[来源：GB/T 29859-2013, 2.2.1]

4 缩略语

下列缩略语适用于本文件。

ANI：平均核酸一致性（Average Nucleotide Identity）

DNA：脱氧核糖核酸（Deoxyribonucleic Acid）

GATK：基因组分析工具包（The Genome Analysis Toolkit）

SNP：单核苷酸多态性（Single Nucleotide Polymorphism）

WGS：全基因组测序（Whole Genome Sequencing）

5 鉴定原理

遗传基因的相似性是菌株鉴定的重要依据之一。通过对食品、保健品或药品等的样本中分离得到的益生菌株进行全基因组测序（WGS），测序结果与现有数据库菌株或参考菌株的基因组序列实施序列比对，从而判断样品中的益生菌株型相似性。

6 实验条件

6.1 实验室设施和环境应符合 GB 19489-2008 和 GB/T 19495.2-2004 的规定。

6.2 应满足鉴定工作的需要,主要常用设备包括但不限于:

- a) 超净工作台：符合 YY/T 1539-2017 要求；
- b) 恒温箱：温度可设置为 37 ℃；
- c) 测序仪：可为二代或三代测序仪；
- d) PCR 仪：应符合 SN/T 2102.2-2008 要求；
- e) 冷冻离心机：应符合 GB/T 30099-2013 要求，且最大离心力宜满足 12000 g 以上；
- f) 微量移液器；
- g) 紫外分光光度计：检测范围 190 nm~600 nm；
- h) DNA 荧光定量仪；
- i) 电泳仪；
- j) 凝胶成像系统；
- k) 水浴锅；
- l) 低温冰箱：温度宜保持在 4 ℃或-20 ℃。

7 实验步骤

7.1 样品采集

7.1.1 样品采集应经过生物安全评价，人体来源样本还需在具备生物安全的基础上进行伦理审查。

7.1.2 采集过程应避免样品被其它物质污染。

7.1.3 宜根据益生菌产品的状态、菌株组成、含菌量、包装等采用相应的采集方案，确保可从样品中培养出益生菌。

7.2 样品保存与运输

7.2.1 单一菌株保存条件及运输方式：

- a) 斜面保存的菌株可冷藏或常温条件下运输；
- b) 甘油溶液保存的菌株，应于-20℃或干冰环境下运输，甘油的浓度通常为10%~20%（浓度体积比）；

7.2.2 益生菌产品保存条件及运输方式：运输温度宜保持在常温或低温条件下（如4℃环境中）；

7.2.3 样品运输不宜超过48h，收到样品后保存于低温环境，尽快安排下游检测。

7.3 样品制备

7.3.1 制备过程均应遵循无菌操作程序。

7.3.2 制备前应核对样品编号，记录样品信息，检查样品状态。确保样品无泄漏或者污染等情况，否则该样品应废弃并重新采样。

7.3.3 固体和半固体样品可按以下两种无菌操作之一制备1:10样品匀液：

- a) 方法一：均质杯法：
 - 1) 称取25g样品；
 - 2) 将其置于装有225mL生理盐水的无菌均质杯内；
 - 3) 5000g~10000g均质1min~2min，制成样品匀液。
- b) 方法二：均质袋法：
 - 1) 称取25g样品；
 - 2) 将其置于225mL生理盐水的无菌均质袋中；
 - 3) 用拍击式均质器拍打1min~2min，制成样品匀液。

7.3.4 液体样品应先将其充分摇匀后，采用无菌吸管吸取样品25mL至已装有225mL生理盐水和适量无菌玻璃珠的无菌锥形瓶中，将其充分振摇制成样品匀液。

7.4 分离培养

7.4.1 益生菌分离培养均应遵循无菌操作程序。

7.4.2 宜采用细菌通用固体培养基（如LB培养基）进行分离培养，或可根据目标菌株采用特定的培养基。

7.4.3 宜吸取1mL的样品匀液参考GB 4789.35-2016要求或者7.1.3确定的方案进行分离和培养。

7.4.4 应根据菌株代谢时对分子氧的需求情况,选择相应的需氧或厌氧的环境进行分离培养。

7.4.5 培养条件宜控制在 37 °C 培养 24 h~72 h。

7.4.6 在获得菌株纯培养物后,需进行涂片镜检或 16S 测序检测,以确定其菌株纯度。

7.5 提取

7.5.1 脱氧核糖核酸(DNA)提取宜采用单克隆菌株的菌体。

7.5.2 宜使用稳定高度可重复的方法,或 DNA 提取试剂盒提取 DNA。

7.5.3 提取过程中应设置阴性对照和重复对照。

7.5.4 DNA 浓度和纯度可按 GB/T 19495.3-2004 和 GB/T 19495.5-2018 描述的方法检测 DNA 样品,并按 SN/T 2497.21-2010 描述的方法检测 DNA 样品的完整性。

7.5.5 提取后的 DNA 应避免反复冻融,可在-20 °C 条件下保存不长于 6 个月或在-80 °C 条件下长期保存。

7.6 测序

7.6.1 测序应遵照 DNA 测序仪的操作规范进行,并对操作步骤和参数进行详细记录。测序条件应进行合理控制,使数据质量达到测序平台要求。

7.6.2 如用二代高通量平台测序,应按 GB/T 30989-2014 描述的方法测序,质量评价可参考 GB/T 35537-2017 要求。

8 数据分析

8.1 数据质量控制

8.1.1 质量值为 Q20 (测序错误率小于 0.01) 的测序数据应占原始数据 80%以上。

8.1.2 原始数据经过过滤之后,测序深度(数据量/基因组大小)应不小于 100 X。

8.1.3 通过软 K-mer 分析评估基因组数据的一致性。正常的 K-mer 分布应该服从泊松分布。

8.1.4 目标菌株的测序序列对全基因组比对的覆盖率应不低于 90%,至少达到高质量基因组草图的标准。

8.2 单核苷酸多态性(SNP)

8.2.1 通过基因组分析工具包(GATK)获得目标菌株的 SNP,并将 SNP 替代到参考基因组序列当中,获得目标菌株的基因组序列。

8.2.2 参考基因组的选择原则:

- a) 目标菌株有假定的物种信息,根据物种拉丁文名搜索基因组数据库,获得参考基因组;
- b) 目标菌株属于未知物种信息,将其测序数据和数据库中所有参考基因组进行比对,比对过滤条件:覆盖率不小于 95%,深度不小于 100X;选取覆盖率(coverage)最高的菌株作为参考基因组。

8.3 比较基因组分析与结果评价

8.3.1 目标菌株的基因组序列应与现有数据库菌株实施序列比对,并计算菌株的平均核酸一致性(ANI)指数,ANI 指数越高,则菌株间的关系越为接近。

8.3.2 目标菌株通过 GATK 获得的 SNP 应与数据库菌株构建系统发育树,系统发育树置信值应大于 70,系统发育树中菌株越靠近,则菌株的关系越接近。

9 报告要求

9.1 报告应描述鉴定出的最可能的菌株,列出菌株全名(学名)、ANI 值及系统发育树。

9.2 报告应注明使用的测序仪器与测序方法。

9.3 报告应对所使用的分析方法(包括但不限于软件,参数,流程)和数据库进行描述,以满足可重复性要求。常见分析软件和参数可参考附录 A。

9.4 报告应注明使用的数据库,数据库宜尽可能全面包含样品中可能存在的物种的基因序列,常见数据库参见附录 B。

10 质量控制

10.1 相关人员应接受生物安全、检测技术等专业培训,持证上岗。

10.2 配备必要的基础设施和适宜的检测环境,对设施和环境进行监控。

10.3 稳定的、已知基因组序列的乳酸菌株可以作为参考标准品,确保每次测序和分析的稳定性。

11 方法局限性

由于目前技术手段的限制,基于全基因组测序的益生菌菌株分型鉴定还存在技术上的局限。这些局限性包括但不限于:

- a) 测序引入的随机或系统性错误;
- b) 注释使用的数据库不能包含所有的微生物物种。
- c) 目前对于益生菌的分离方法不包括所有菌种或菌株。

附 录 A
(资料性附录)
常见分析软件和参数

A.1 全基因组比对

```
bwa-0.7.12 aln -o 1 -e 63 -i 90 -L -k 2 -l 31 -t 4 -q 10; SOAP.coverage Version: 2.7.7
```

A.2 GATK

```
The Genome Analysis Toolkit, https://gatk.broadinstitute.org/hc/en-us;  
--filterExpression "DP < 20 || QUAL<30.0 || QD < 2.0 || MQ < 40.0 || FS > 60.0 || SOR > 3.0"  
--filterName "Filter"
```

A.3 构建系统发育树

```
iqtree 1.6.12 m -st DNA -pre IQtree -ntmax 16 -nt 16 -v -m GTR+FO+I+G -b 1000 -safe -cmax  
123
```

附 录 B
(资料性附录)
常见数据库

B.1 美国立生物技术信息中心数据库

访问地址为<ftp://ftp.ncbi.nlm.nih.gov/blast/db/>

B.2 美国立生物技术信息中心细菌基因组数据库

访问地址为<http://www.ncbi.nlm.nih.gov/genome>

B.3 美国立生物技术信息中心物种分类数据库

访问地址为<http://www.ncbi.nlm.nih.gov/taxonomy>

B.4 美国能源部联合基因组研究中心 (DOE-JGI) 的微生物基因组数据库

访问地址为<https://img.jgi.doe.gov/>

参 考 文 献

[1] Kim M, Oh HS, Park SC, et al. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. IJSEM. 64(Pt2): 346-351 (2014).

[2] P. S. G. Chain, et al. Genome Project Standards in a New Era of Sequencing. SCIENCE, 326:236-237 (2009).
