

链霉菌基质辅助激光解吸电离飞行时间质谱数据库的建立及测试分析

孙伟*, 罗倩, 许杰, 李雷, 蒋海霞

上海交通大学 生命科学技术学院, 上海

孙伟, 罗倩, 许杰, 李雷, 蒋海霞. 链霉菌基质辅助激光解吸电离飞行时间质谱数据库的建立及测试分析[J]. 微生物学报, 2025, 65(9): 4211-4223.

SUN Wei, LUO Qian, XU Jie, LI Lei, JIANG Haixia. Establishment and testing of a MALDI-TOF MS database for the identification of *Streptomyces*[J]. *Acta Microbiologica Sinica*, 2025, 65(9): 4211-4223.

摘要: 链霉菌种类繁多, 以盛产抗生素而闻名, 是极具研发价值的放线菌类群。然而, 商业数据库中收录的链霉菌属菌种数量较少, 这限制了基质辅助激光解吸电离飞行时间质谱(matrix-assisted laser desorption/ionization-time of flight mass spectrometry, MALDI-TOF MS)技术在链霉菌鉴定中的应用。【目的】基于全基因组测序数据挖掘, 建立链霉菌属菌种的 MALDI-TOF MS 数据库, 并测试其准确性。【方法】检索基因组数据库, 提取已测序的链霉菌属菌种模式菌株的 12 种核糖体蛋白序列, 并计算其理论 MALDI 质谱峰值, 以这 12 个峰值的组合作为菌种鉴定依据建立数据库。收集模式菌株进行 MALDI-TOF MS 测试, 将实测峰值与数据库中菌种的理论峰值进行比对, 以测试建库方法及数据库比对结果的准确性。【结果】建立了包含链霉菌属 615 个菌种的 MALDI-TOF MS 数据库。测试菌株的质谱均实现了与数据库中目标菌种的准确匹配。本研究还提出了基于数据库比对的菌种鉴定方法。【结论】本研究建立的数据库为基于 MALDI-TOF MS 技术鉴定链霉菌奠定了基础。

关键词: 链霉菌; 基质辅助激光解吸电离飞行时间质谱; 数据库; 模式菌株; 核糖体蛋白质

资助项目: 上海交通大学决策咨询课题(JCZXSJA2023-04)

This work was supported by the Decision-making Consultation Project of Shanghai Jiao Tong University (JCZXSJA2023-04).

*Corresponding author. Tel: +86-21-34206834, E-mail: bacterior@sjtu.edu.cn

Received: 2025-03-02; Accepted: 2025-04-16; Published online: 2025-05-08

Establishment and testing of a MALDI-TOF MS database for the identification of *Streptomyces*

SUN Wei*, LUO Qian, XU Jie, LI Lei, JIANG Haixia

School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

Abstract: *Streptomyces* is a genus of diverse actinomycetes known for producing antibiotics, showcasing high value for research and development. However, the limited number of *Streptomyces* species included in commercial databases restricts the application of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) in the identification of *Streptomyces*. **[Objective]** To establish a MALDI-TOF MS database for the identification of *Streptomyces* based on the mining of whole genome sequencing data and test its accuracy. **[Methods]** By searching against a genome database, we selected all the genome-sequenced type strains of *Streptomyces*. Subsequently, the sequences of 12 ribosomal proteins were extracted from their genome sequences and theoretical MALDI mass/charge (m/z) values were calculated. Accordingly, a database was established based on the combination of 12 theoretically calculated m/z values, which can be used as the identification criteria for *Streptomyces* species. Finally, type strains were randomly collected for MALDI-TOF MS testing. To test the accuracy of the database construction method and comparison results, we matched the experimental m/z values with the theoretically calculated m/z values of the species in the database. **[Results]** We established a MALDI-TOF MS database including 615 *Streptomyces* species. The mass spectra of the tested strains were accurately matched with those of corresponding species in the database. In addition, we proposed a method for the identification of *Streptomyces* by database comparison. **[Conclusion]** The database established in this study laid a foundation for identifying *Streptomyces* based on MALDI-TOF MS.

Keywords: *Streptomyces*; matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS); database; type strain; ribosomal protein

基质辅助激光解吸电离飞行时间质谱 (matrix-assisted laser desorption/ionization-time of flight mass spectrometry, MALDI-TOF MS) 是一种软电离质谱技术, 具有速度快、成本低和准确率高的特点, 已广泛应用于医学诊断、环境监测和食品安全等领域的微生物鉴定^[1-6]。该技术鉴定微生物的原理是采集待测菌株的蛋白质指纹图谱, 与 MALDI-TOF MS 数据库中的菌株图谱比对, 根据比对分值鉴定至属、种或株。菌种鉴定的准确性依赖于 MALDI-TOF MS 图谱数

据库的完善性与可靠性。当前的商业数据库, 如德国布鲁克的 MALDI Biotyper 数据库、法国梅里埃的 VITEK MS 数据库以及安图生物的本土微生物数据库均以收录病原微生物为主, 存在菌种类别及数量不足, 不能满足用户鉴定需求的问题^[7]。对于不在商业数据库范围内, 但具有重要研发价值的非病原微生物建立相应数据库, 可拓展 MALDI-TOF MS 技术鉴定微生物的应用范围。

针对用户多样化的鉴定需求, 质谱仪厂商

提供了一套自建 MALDI-TOF MS 数据库方案, 包括获得建库菌种的单菌落、甲酸提取法制备样本、仪器校正、图谱采集、数据评估和图谱入库等步骤。建库过程需要严格的质量控制^[8], 建库人员需要掌握微生物学技术并经过质谱相关专业培训。因此自建数据库一般不予公开, 或者需要购买使用, 限制了自建库的共享。

早期研究表明, 细菌 MALDI-TOF MS 图谱中以核糖体蛋白质的质谱峰为主^[9-10]。本课题组在前期研究中基于放线菌全基因组测序数据注释出的核糖体蛋白质序列预测分子量, 找出了放线菌 MALDI-TOF MS 图谱中常见的核糖体蛋白质信号, 为识别核糖体蛋白质的质谱峰鉴定放线菌的方法建立提供了理论依据^[11]。通过对伦茨氏菌属 (*Lentzea*) 和糖丝菌属 (*Saccharothrix*) 2 类稀有放线菌的研究发现, 核糖体蛋白质的理论 MALDI 质谱峰值 (mass/charge, m/z) 组合可以作为菌种鉴定依据^[12-13], 证实了基于全基因组测序数据挖掘建立 MALDI-TOF MS 数据库方案的可行性。

链霉菌 (*Streptomyces*) 属于放线菌门 (*Actinomycetota*) 放线菌纲 (*Actinomycetes*) 北里氏孢菌目 (*Kitasatosporales*) 链霉菌科 (*Streptomycetaceae*), 广泛分布于土壤、淡水和海洋环境。链霉菌是放线菌门中最大的分支, 许多种类的链霉菌能够产生抗生素、抗癌化合物、抗病毒化合物、免疫抑制剂、驱虫剂和除草剂等生物活性物质^[14], 是极具研发价值的放线菌类群。在链霉菌多样性研究及天然产物资源勘探中, 菌种鉴定是一项基础工作, 最常用的鉴定方法仍是 16S rRNA 基因测序分析法^[15]。由于商业数据库收录的链霉菌属菌种数量少, MALDI-TOF MS 技术尚未应用于链霉菌鉴定。随着高通量测序技术的快速发展和测序成本的显著降低, 公开基因组测序数据的链霉菌属菌种与日俱增, 为基于全基因组测序数据挖掘建立链霉菌 MALDI-TOF MS 数据库创造了条件。前期研究发现, 分子量小于 10 kDa 的 12 种核糖

体蛋白质 (L36、L34、L33、L32、L30、L28、S14-Z、L35、L29、S18、L27 和 S20) 在放线菌 MALDI-TOF MS 图谱中以高信号强度稳定出现^[12-13]。在对 MALDI Biotyper 数据库中链霉菌属菌种的 MALDI-TOF MS 图谱进行核糖体蛋白质信号注释时, 同样发现这 12 种蛋白质的注释率高^[11]。因此本研究以链霉菌属菌种模式菌株 12 种核糖体蛋白质的理论 MALDI 质谱峰值组合作为菌种鉴定依据, 建立链霉菌 MALDI-TOF MS 数据库, 并收集模式菌株进行测试分析, 测试建库方法及数据库比对结果的准确性。

1 材料与方法

1.1 链霉菌属入库菌种选取

以“*Streptomyces*”为检索词搜索原核生物标准命名名录 (list of prokaryotic names with standing in nomenclature, LPSN) (<https://lpsn.dsmz.de/>), 获得有效发表并正确命名的链霉菌属菌种信息, 包括菌种名称及模式菌株编号。以菌种名称为检索词搜索美国国家生物技术信息中心基因组数据库 (National Center for Biotechnology Information, NCBI) (<https://www.ncbi.nlm.nih.gov/datasets/genome/>), 选取已公开模式菌株基因组测序数据的菌种入库。

1.2 入库菌种模式菌株核糖体蛋白质序列提取

以“ribosomal”为检索词, 基因类型选择蛋白质编码基因, 搜索入库菌种模式菌株的基因组组装报告, 将检索到的核糖体蛋白质按序列长度排序, 提取长度小于 90 个氨基酸、分子量小于 10 kDa 的 12 种核糖体蛋白质 (L36、L34、L33、L32、L30、L28、S14-Z、L35、L29、S18、L27 和 S20) 的序列。

1.3 核糖体蛋白质理论 MALDI 质谱峰值计算

链霉菌属于原核生物, 蛋白质翻译后 N-端的起始甲硫氨酸可能被切除。分析核糖体蛋白

质序列第 2 个氨基酸, 如符合切除规则^[16], 用去除起始甲硫氨酸的序列计算分子量, 否则用初始序列计算分子量。应用在线蛋白质基本性质分析工具 ProtParam (<https://web.expasy.org/protparam/>) 计算得到核糖体蛋白质分子量。

原核生物存在甲基化、乙酰化等翻译后修饰^[17]。前期研究发现链霉菌属菌种核糖体蛋白质 L33 发生一个甲基化修饰, 导致分子量增加 14 Da^[11], 计算 L33 的分子量需考虑这一情况。

蛋白质在基质辅助激光解吸电离下产生带 1 个电荷的阳离子 $[M+H]^+$, 核糖体蛋白质的理论分子量换算为 MALDI 质谱峰值需考虑蛋白质结合 1 个质子后分子量增加 1 Da 的情况。

综上所述, 考虑起始甲硫氨酸切除、L33 的甲基化修饰和 MALDI 质子化 3 个因素后计算得到核糖体蛋白质的理论 MALDI 质谱峰值。

1.4 链霉菌 MALDI-TOF MS 数据库建立与分析

以入库菌种模式菌株 12 种核糖体蛋白质的理论 MALDI 质谱峰值组合作为各菌种鉴定依据, 建立链霉菌 MALDI-TOF MS 数据库。

评估以 12 种核糖体蛋白质的理论质谱峰值组合区分菌种的能力。对于无法区分的组别, 将组内菌种模式菌株的基因组序列提交模式菌株基因组服务器(type strain genome server, TYGS) (<https://tygs.dsmz.de/>) 进行比对, 获得组内菌种间的数字 DNA-DNA 杂交 (digital DNA-DNA hybridization, dDDH) 值, dDDH 值 > 70% 的菌种判定为相同菌种, dDDH 值 < 70% 的菌种判定为不同菌种^[18]。

1.5 链霉菌属菌种模式菌株收集

随机收集 5 个链霉菌属菌种的模式菌株用于测试和分析, 包括委内瑞拉链霉菌 (*Streptomyces venezuelae*) ATCC 10712^T、灰色链霉菌 (*Streptomyces griseus*) J1014^T、阿维菌素链霉菌 (*Streptomyces avermitilis*) NRRL 8165^T、微白黄链霉菌 (*Streptomyces albidoflavus*) J1074^T 和

天蓝色链霉菌 (*Streptomyces coelicolor*) M145^T, 均由上海交大学生命科学技术学院李雷副教授惠赠。

1.6 菌株 MALDI-TOF MS 测试与分析

制备 ISP2 固体培养基, 28 °C 培养菌株 2–3 d, 用无菌竹签挑取单菌落涂抹到 MALDI 384 抛光不锈钢靶板上, 室温自然干燥。在单菌落涂层上加入 1 μ L 10 mg/mL 的 α -氰基-4-羟基肉桂酸基质溶液, 溶剂为乙腈:水:三氟乙酸 (50:47.5:2.5, 体积比)。室温自然干燥后放入质谱仪。

基质辅助激光解析高分辨飞行时间质谱仪购自布鲁克道尔顿公司 (Bruker Daltonik GmbH)。使用控制软件 flexControl 3.4 采集质谱图, 使用布鲁克细菌测试标准品校准方法。仪器测定参数: 采集 m/z 范围为 4 000–10 000; 激光点击数为 1 000; 测定模式为阳离子线性模式。如单次采集的质谱峰强度低于 10^4 , 通过累加模式使质谱峰强度达到 10^4 。使用分析软件 flexAnalysis 3.4 处理质谱图并标峰, 导出峰列表。

基于质谱仪在线性模式下的测量精度, 将蛋白质分子离子 m/z 的最大质量误差设为 ± 500 ppm 并按 2 种方法测试: (1) 将测试菌株的质谱峰值与其所属菌种模式菌株的 12 个理论峰值一一比对, 检测其一致性, 此方法测试建库方法的准确性; (2) 将测试菌株的质谱峰值与数据库菌种进行比对, 找到匹配数最大的菌种, 检测其是否为所属菌种, 此方法测试数据库比对结果的准确性。

基于测试分析结果建立基于数据库比对的菌种鉴定方法。

1.7 16S rRNA 基因测序分析法验证

按本研究建立的方法对上海辰山植物园土壤样品中分离出的放线菌菌株进行鉴定, 选取鉴定至种的菌株委托苏州金唯智生物科技有限公司进行测试。基因组 DNA 提取、16S rRNA 基因扩增和 Sanger 法测序均由苏州金唯智生物

科技有限公司完成。菌株的 16S rRNA 基因序列提交至 GenBank 获得登录号并在 EzBioCloud (<https://www.ezbiocloud.net/>) 数据库进行比对和鉴定。对 2 种方法得到的鉴定结果进行比较和分析。

2 结果与分析

2.1 链霉菌属入库菌种选取

截至 2025 年 1 月 31 日, 原核生物标准命名名录收录的有效发表且正确命名的链霉菌属菌种共计 738 个, 其中公开模式菌株基因组测序数据的菌种有 626 个。除去未注释基因组的 2 个、未核实来源的 3 个、基因组污染的 4 个以及未注释出 L36 的 2 个菌种, 最终数据库 1.0 版本包含链霉菌属 615 个菌种[原始数据存储于国家微生物科学数据中心(<http://nmdc.cn>), 编号为 NMDCX0002103], 覆盖率达 83.3%。

2.2 入库菌种模式菌株核糖体蛋白质序列提取

链霉菌的核糖体蛋白质在序列长度上表现出高度保守性。需要注意的是, L32、L28 和 S18 注释出不同长度的 2 套序列, L33 注释出相同长度(54 个氨基酸)的 2 套序列。基于前期研究结果^[1], L32 提取长度为 57 个氨基酸的序列, L28 提取长度为 61 个氨基酸的序列, S18 提取长度为 78 个氨基酸的序列, L33 提取以“MAATDV”起始的序列。

2.3 核糖体蛋白质理论 MALDI 质谱峰值计算

序列分析表明链霉菌的 12 种核糖体蛋白质中, L36 不符合起始甲硫氨酸切除规则, 采用初始序列计算理论分子量; 另外 11 种核糖体蛋白质则用去除起始甲硫氨酸的序列计算理论分子量。L33 在理论分子量基础上加 14 以匹配甲基化。转换为 MALDI 质谱峰值时, 将所有分子量加 1 以匹配 MALDI 质子化。最终得到 615 个菌种模式菌株的核糖体蛋白质理论 MALDI 质谱峰

值组合, 其中 606 个菌种包含 12 个峰值的组合, 另外 9 个菌种包含 11 个峰值的组合, 这是由于 1 种核糖体蛋白质未注释所致[老鹳草链霉菌(*Streptomyces geranii*)缺少 L33; 银耳链霉菌(*Streptomyces tremellae*)、产褐色链霉菌(*Streptomyces fuscigenes*)、山土链霉菌(*Streptomyces montanisoli*)、白黄黑链霉菌(*Streptomyces albiflaviniger*)、新泽西链霉菌(*Streptomyces novaecaesareae*)和海南链霉菌(*Streptomyces hainanensis*)缺少 L28; *Streptomyces solicavernae* 缺少 S18; 杀病链霉菌(*Streptomyces pathocidini*)缺少 S20]。

2.4 链霉菌 MALDI-TOF MS 数据库建立与分析

为便于数据分析和数据库检索, 将每个模式菌株的理论 MALDI 质谱峰值按从小到大排列, 入库菌种按最小峰值从小到大排列, 峰值相同则按次小峰值从小到大排列, 以此类推, 对入库菌种排序并编号, 建立链霉菌 MALDI-TOF MS 数据库 1.0 版本 (<https://biocore.sjtu.edu.cn/public/upload/20250430/%E9%93%BE%E9%9C%89%E8%8F%8CMALDI-TOF+MS%E6%95%B0%E6%8D%AE%E5%BA%93V1.0.xlsx>)。数据库中 537 个菌种(占比 87.3%)具有特征性峰值组合, 可与其他菌种区分开。数据库中存在相同峰值组合的菌种共计 34 组 78 个(编号为 NMDCX0002103), 这些组内的菌种间无法区分。

基于基因组指标的分类方法, 如数字 DNA-DNA 杂交、平均核苷酸一致性和平均氨基酸一致性分析如今已成为主流分类方法^[15]。近年来, 基于基因组指标的研究发现链霉菌属菌种间存在异名同种现象, 部分菌种名不再作为正确名称, 而被定义为“Synonym”^[19-20]。本研究对数据库中具有相同峰值组合的菌种进行基因组序列比对, 发现 23 个组的组内菌种间 dDDH 值 > 70% (表 1), 提示这些组的组内菌种应为异名同

表1 dDDH值大于70%的23组链霉菌属菌种

Table 1 Twenty-three *Streptomyces* groups in which dDDH values of the members are above 70%

Group	Organism A	Organism B	dDDH value (%)
1	<i>S. pluricologrescens</i> JCM 4602 ^T	<i>S. rubiginosohelvolus</i> JCM 4415 ^T	99.3
2	<i>S. asiaticus</i> DSM 41761 ^T	<i>S. cangkringensis</i> DSM 41769 ^T	98.4
	<i>S. asiaticus</i> DSM 41761 ^T	<i>S. indonesiensis</i> DSM 41759 ^T	98.9
	<i>S. asiaticus</i> DSM 41761 ^T	<i>S. rhizosphaericus</i> DSM 41760 ^T	96.6
	<i>S. cangkringensis</i> DSM 41769 ^T	<i>S. indonesiensis</i> DSM 41759 ^T	98.4
	<i>S. cangkringensis</i> DSM 41769 ^T	<i>S. rhizosphaericus</i> DSM 41760 ^T	96.3
	<i>S. indonesiensis</i> DSM 41759 ^T	<i>S. rhizosphaericus</i> DSM 41760 ^T	96.2
3	<i>S. peucetius</i> NRRL WC-3868 ^T	<i>S. rimosus</i> ATCC 10970 ^T	80.0
4	<i>S. niger</i> NRRL B-3857 ^T	<i>S. olivaceiscleroticus</i> JCM 4805 ^T	81.3
5	<i>S. libani</i> JCM 4322 ^T	<i>S. nigrescens</i> DSM 40276 ^T	92.0
6	<i>S. antimycoticus</i> NBRC 12839 ^T	<i>S. mordarskii</i> JCM 5052 ^T	75.6
7	<i>S. cuspidosporus</i> JCM 4316 ^T	<i>S. sparsogenes</i> DSM 40356 ^T	77.8
8	<i>S. malaysiensis</i> JCM 10672 ^T	<i>S. solisilvae</i> HNM0141 ^T	99.8
9	<i>S. collinus</i> DSM 40129 ^T	<i>S. violaceochromogenes</i> JCM 4530 ^T	94.8
10	<i>S. daghestanicus</i> JCM 4365 ^T	<i>S. griseoviridis</i> DSM 40229 ^T	96.0
	<i>S. daghestanicus</i> JCM 4365 ^T	<i>S. niveoruber</i> JCM 4234 ^T	94.6
	<i>S. griseoviridis</i> DSM 40229 ^T	<i>S. niveoruber</i> JCM 4234 ^T	94.6
11	<i>S. roseiscleroticus</i> JCM 4823 ^T	<i>S. ruber</i> JCM 3131 ^T	100.0
12	<i>S. gancidicus</i> JCM 4171 ^T	<i>S. pseudogriseolus</i> NRRL B-3288 ^T	94.2
	<i>S. gancidicus</i> JCM 4171 ^T	<i>S. rubiginosus</i> JCM 4416 ^T	95.6
	<i>S. pseudogriseolus</i> NRRL B-3288 ^T	<i>S. rubiginosus</i> JCM 4416 ^T	94.4
13	<i>S. griseoincarnatus</i> JCM 4381 ^T	<i>S. labedae</i> JCM 9381 ^T	95.2
	<i>S. griseoincarnatus</i> JCM 4381 ^T	<i>S. tunisiensis</i> JCM 17589 ^T	83.6
	<i>S. labedae</i> JCM 9381 ^T	<i>S. tunisiensis</i> JCM 17589 ^T	83.6
14	<i>S. griseoluteus</i> JCM 4765 ^T	<i>S. recifensis</i> NRRL B-3811 ^T	72.4
15	<i>S. finlayi</i> JCM 4637 ^T	<i>S. spiroverticillatus</i> JCM 4609 ^T	99.9
16	<i>S. carpaticus</i> JCM 6915 ^T	<i>S. cheonanensis</i> JCM 14549 ^T	87.3
	<i>S. carpaticus</i> JCM 6915 ^T	<i>S. harbinensis</i> CGMCC 4.7047 ^T	87.7
	<i>S. cheonanensis</i> JCM 14549 ^T	<i>S. harbinensis</i> CGMCC 4.7047 ^T	92.1
17	<i>S. bellus</i> JCM 4292 ^T	<i>S. coeruleorubidus</i> JCM 4359 ^T	81.8
18	<i>S. thermogriseus</i> JCM 11269 ^T	<i>S. thermovulgaris</i> NRRL B-12375 ^T	95.0
19	<i>S. canarius</i> JCM 4733 ^T	<i>S. olivaceoviridis</i> JCM 4499 ^T	76.0
20	<i>S. graminearus</i> JCM 6923 ^T	<i>S. murinus</i> NRRL B-2286 ^T	86.2
21	<i>S. griseoaurantiacus</i> JCM 4763 ^T	<i>S. jietaisiensis</i> CGMCC 4.1859 ^T	90.4
22	<i>S. albaduncus</i> CECT 3226 ^T	<i>S. griseoloalbus</i> JCM 4480 ^T	92.1
23	<i>S. calvus</i> CECT 3271 ^T	<i>S. virens</i> JCM 9095 ^T	92.8

种，因此具有相同的峰值组合符合预期。18个组(30个种)的组内菌种间 dDDH 值<70% (表 2)，提示这些组的组内菌种并非同种。经基因组序

列比对验证后，数据库中无法区分的菌种实际为 30 个(占比 4.9%，表 2)，这些菌种无法鉴定到种，而数据库中约 95% 的菌种可以鉴定到种。

表2 dDDH值小于70%的18组链霉菌属菌种

Table 2 Eighteen *Streptomyces* groups in which dDDH values of the members are below 70%

Group	Organism A	Organism B	dDDH value (%)
1	<i>S. bikiniensis</i> NRRL ISP-5581 ^T	<i>S. cinereoruber</i> ATCC 19740 ^T	60.7
2	<i>S. amritsarensis</i> MTCC 11845 ^T	<i>S. yangpuensis</i> fd2-tb ^T	46.6
3	<i>S. globisporus</i> JCM 4378 ^T	<i>S. pluricolaroscens</i> JCM 4602 ^T	65.8
4	<i>S. globisporus</i> JCM 4378 ^T	<i>S. rubiginosohelvolus</i> JCM 4415 ^T	66.1
5	<i>S. bacillaris</i> ATCC 15855 ^T	<i>S. cavourensis</i> JCM 4298 ^T	68.0
6	<i>S. globosus</i> JCM 13859 ^T	<i>S. toxytricini</i> JCM 4421 ^T	57.1
7	<i>S. barkulensis</i> RC 1831 ^T	<i>S. pini</i> PL19 ^T	67.3
8	<i>S. chrestomyceticus</i> JCM 4735 ^T	<i>S. monomycini</i> NRRL B-24309 ^T	49.4
9	<i>S. chrestomyceticus</i> JCM 4735 ^T	<i>S. paromomycinus</i> NBRC 15454 ^T	59.9
10	<i>S. monomycini</i> NRRL B-24309 ^T	<i>S. paromomycinus</i> NBRC 15454 ^T	50.0
11	<i>S. catenulae</i> NRRL B-2342 ^T	<i>S. ramulosus</i> JCM 4193 ^T	47.7
12	<i>S. antimycoticus</i> NBRC 12839 ^T	<i>S. yatensis</i> JCM 13244 ^T	63.9
13	<i>S. mordarskii</i> JCM 5052 ^T	<i>S. yatensis</i> JCM 13244 ^T	61.3
14	<i>S. cyaneochromogenes</i> MK-45 ^T	<i>S. shaanxiensis</i> JCM 16925 ^T	63.3
15	<i>S. neyagawaensis</i> JCM 4796 ^T	<i>S. ossamyceticus</i> NRRL B-3822 ^T	57.9
16	<i>S. neyagawaensis</i> JCM 4796 ^T	<i>S. torulosus</i> NRRL B-3889 ^T	62.1
17	<i>S. harenosi</i> PRKS01-65 ^T	<i>S. leeuwenhoekii</i> C34 ^T	56.8
18	<i>S. bangladeshensis</i> JCM 14924 ^T	<i>S. nogalater</i> JCM 4799 ^T	36.6

2.5 菌株 MALDI-TOF MS 测试与分析

本研究中 5 个测试菌株的 MALDI-TOF MS 图谱见图 1, 菌株的质谱峰列表原始数据存储于国家微生物科学数据中心(<http://nmdec.cn>), 编号为 NMDCX0002103。

2.5.1 建库方法准确性测试

用菌种名“*Streptomyces venezuelae*”搜索数据库, 找到 0256 *Streptomyces venezuelae* ATCC 10712^T。峰值比对表明, 菌株质谱图中的 11 个峰值(图 1A)与目标菌株的 12 个理论峰值匹配(表 3), m/z 6 816.72 同时匹配 6 815.83 和 6 819.14, 峰值的最大质量误差为-354.88 ppm, 实测值和理论值完全一致。

用菌种名“*Streptomyces griseus*”搜索数据库, 找到 0076 *Streptomyces griseus* DSM 40236^T。峰值比对表明, 菌株质谱图中的 11 个峰值(图 1B)与目标菌株的 12 个理论峰值匹配(表 3), m/z 6 860.70 同时匹配 6 859.85 和 6 862.15, 峰值的

最大质量误差为-345.69 ppm, 实测值和理论值完全一致。

用菌种名“*Streptomyces avermitilis*”搜索数据库, 找到 0276 *Streptomyces avermitilis* MA-4680^T。峰值比对表明, 菌株质谱图中的 12 个峰值(图 1C)与目标菌株的 12 个理论峰值匹配(表 3), 峰值的最大质量误差为-286.66 ppm, 实测值和理论值完全一致。

用菌种名“*Streptomyces albidoflavus*”搜索数据库, 找到 0490 *Streptomyces albidoflavus* DSM 40455^T。峰值比对表明, 菌株质谱图中的 11 个峰值(图 1D)与目标菌株的 11 个理论峰值匹配(表 3), 峰值的最大质量误差为-384.90 ppm。由于缺少与 m/z 6 963.19 匹配的峰值, 未实现实测值和理论值完全一致。进一步分析发现, 线性模式分辨率低, 质谱图中的高强度峰 m/z 6 971.61 遮盖了附近低强度的核糖体蛋白质峰, 导致对质谱图标峰时未将其标注上。将质谱图局部放大后, 肩峰 m/z 6 963.26 清晰可见(图 1D), 因此菌

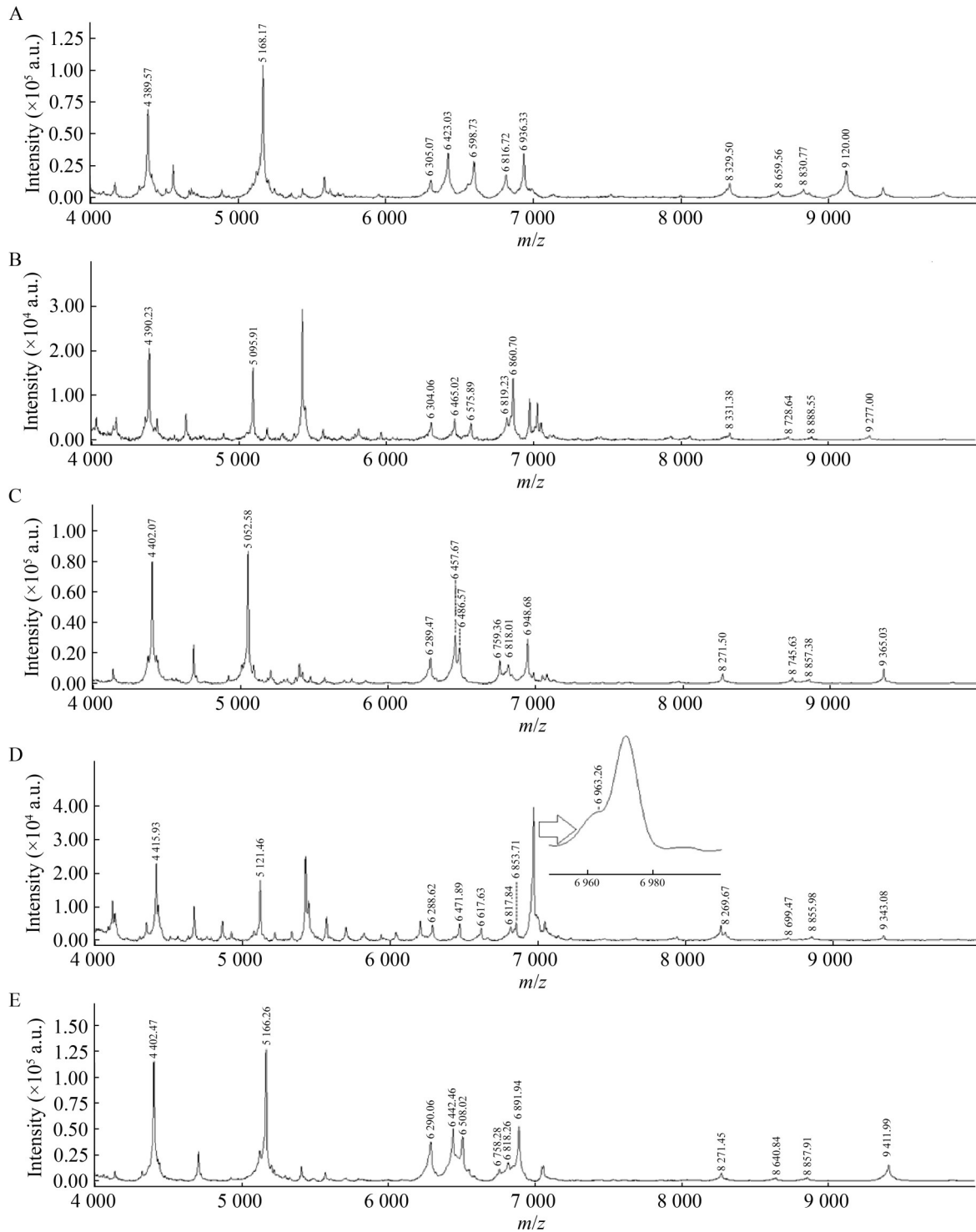


图1 链霉菌属测试菌株的MALDI-TOF MS图谱

Figure 1 MALDI-TOF MS spectra of *Streptomyces* test strains. A: *S. venezuelae* ATCC 10712^T; B: *S. griseus* J1014^T; C: *S. avermitilis* NRRL 8165^T; D: *S. albidoflavus* J1074^T; E: *S. coelicolor* M145^T. Peaks that match ribosomal proteins are marked.

株质谱实测值和理论值完全一致。

用菌种名“*Streptomyces coelicolor*”搜索数据库,未找到目标菌种。分析原因发现 LPSN 已将 *Streptomyces coelicolor* 定义为“Synonym”,菌种未入库。

对 4 个测试菌株的比对和分析表明了实测峰值与理论峰值的一致性,证实了本研究所用建库方法的准确性。

2.5.2 数据库比对结果准确性测试

基于 m/z 4 389.57 和 5 168.17,将检索范围定位至 0256–0257,进一步比对得出测试菌株与 0256 *Streptomyces venezuelae* ATCC 10712^T 的匹配数最大,为 12。

基于 m/z 4 390.23 和 5 095.91,将检索范围定位至 0064–0081,进一步比对得出测试菌株与 0076 *Streptomyces griseus* DSM 40236^T 的匹配数最大,为 12;与 0077 *Streptomyces anulatus* JCM 4721^T 的匹配数为 11。

基于 m/z 4 402.07 和 5 052.58,将检索范围定位至 0276–0277,进一步比对得出测试菌株与 0276 *Streptomyces avermitilis* MA-4680^T 的匹配数最大,为 12。

基于 m/z 4 415.93 和 5 121.46,将检索范围定位至 0484–0492,进一步比对得出测试菌株与 0490 *Streptomyces albidoflavus* DSM 40455^T 的匹配数最大,为 11;与 0491 *Streptomyces koyangensis* VK-A60^T 的匹配数为 10;与 0489 *Streptomyces diastaticus* JCM 4128^T 的匹配数为 9。

基于 m/z 4 402.07 和 5 166.26,将检索范围

定位至 0358–0377,进一步比对得出测试菌株与 0361 *Streptomyces violaceoruber* CGMCC 4.1801^T 的匹配数最大,为 12。为确定 *S. coelicolor* 的分类地位,按金标准基因组测序比对法进行鉴定。将 *S. coelicolor* A3(2)^T 的基因组序列 (GCA 000203835.1) 提交至模式菌株基因组服务器进行比对,鉴定结果为 *Streptomyces violaceoruber* (表 4),与数据库比对结果一致。

对 5 个测试菌株的数据库比对找到匹配数最大的菌种,均与所属菌种一致,证实了数据库比对结果的准确性。非正确命名的菌种 *S. coelicolor* 也得到了准确鉴定。

2.5.3 基于数据库比对的鉴定方法建立

基于以上测试分析结果,数据库可用于链霉菌属已知菌种的检测和未知菌种的鉴定。检测已知菌种的方法:用菌种名搜索数据库,找到目标菌种模式菌株,进行理论峰值和实测峰值匹配,若 12 个理论峰值全匹配,则将待检菌株确定为目标菌种。鉴定未知菌种的方法:基于 L36 (m/z 范围为 4 367.36–4 432.44) 和 L34 (m/z 范围为 5 036.88–5 248.15) 的峰值缩小检索范围,从中找出匹配数最大的模式菌株,若匹配数为 12,则将待检菌株鉴定为该模式菌株所属菌种。

2.6 16S rRNA 基因测序分析法验证

数据库比对发现,与数据库菌种的最大匹配数为 12 和 11 的菌株共有 6 株(表 5),菌株质谱峰列表原始数据存储于国家微生物科学数据中心 (<http://nmdc.cn>),编号为 NMDCX0002103。

表4 测试菌株*Streptomyces coelicolor* A3(2)的基因组比对结果

Table 4 Genome-based comparison of *Streptomyces coelicolor* A3(2)

Query	Subject	dDDH value (%)	Taxonomic status
<i>S. coelicolor</i> A3(2)	<i>Streptomyces anthocyanicus</i> JCM 5058	94.2	Synonym
<i>S. coelicolor</i> A3(2)	<i>Streptomyces coelestis</i> JCM 4739	93.9	Synonym
<i>S. coelicolor</i> A3(2)	<i>Streptomyces violaceolatus</i> JCM 4531	93.7	Synonym
<i>S. coelicolor</i> A3(2)	<i>Streptomyces violaceoruber</i> CGMCC 4.1801	93.3	Correct name
<i>S. coelicolor</i> A3(2)	<i>Streptomyces humiferus</i> JCM 3037	92.4	Synonym
<i>S. coelicolor</i> A3(2)	<i>Streptomyces rubrogriseus</i> NBRC 15455	70.2	Correct name

菌株 S1-8、S1-13 和 S1-16 与数据库菌种的最大匹配数为 12, 可鉴定至种。菌株 S1-6、S1-10 和 S1-39 与数据库菌种的最大匹配数为 11, 未匹配的 1 个峰值是由于质量轴偏移导致其超出允许最大误差(编号为 NMDCX0002103)。前期研究发现质量轴偏移时可适当调整质量误差范围^[11], 将最大误差设定为-600 ppm, 3 个菌株的最大匹配数均达到 12, 同样可鉴定至种。

六个菌株的 16S rRNA 基因序列比对结果见表 6, 其中 5 个菌株的比对结果与本研究建立方法的鉴定结果一致。鉴定结果不一致的菌株是 S1-13, 其 16S rRNA 基因序列相似度最高的是 *Streptomyces marianii* ICN19^T。进一步分析发现, LPSN 中“*Streptomyces marianii*”为非有效发表种, 提示这一结果并不准确。序列相似度第二高的是婺源链霉菌(*Streptomyces wuyuanensis*) CGMCC 4.7042^T, 鉴定结果一致。基于以上测

试分析, 本研究建立的基于数据库比对的菌种鉴定方法与 16S rRNA 基因测序分析法具有高度一致的鉴定结果。

3 讨论与结论

本研究基于基因组测序数据挖掘建立了链霉菌 MALDI-TOF MS 数据库。建库过程不依赖仪器测试, 避免了菌株培养、样本处理、上机测试、质量控制等图谱数据库构建环节, 极大简化了建库流程, 降低了建库成本。这套方法易于标准化, 只要遵循蛋白质的翻译后加工、修饰规律, 即可准确预测并计算得出目标菌种的一组核糖体蛋白质理论质谱峰值, 这些数值不受菌株培养条件、样本处理方法、仪器测量精度等因素的影响而产生偏差。因此这套建库方案具备推广条件, 可应用于其他细菌类群的

表5 六个土壤来源菌株的MALDI-TOF MS数据库比对结果

Table 5 MALDI-TOF MS database-based comparison of six strains isolated from a soil sample

Query	Top-hit type strain	Matching number (Err≤±500 ppm)	Matching number (Err≤±600 ppm)
S1-6	<i>Streptomyces virginiae</i> NRRL ISP-5094 ^T	11	12
S1-8	<i>Streptomyces thermocarboxydus</i> JCM 10368 ^T	12	12
S1-10	<i>Streptomyces lydicamycinicus</i> TP-A0598 ^T	11	12
S1-13	<i>Streptomyces wuyuanensis</i> CGMCC 4.7042 ^T	12	12
S1-16	<i>Streptomyces alboflavus</i> NRRL B-2373 ^T	12	12
S1-39	<i>Streptomyces violaceoruber</i> CGMCC 4.1801 ^T	11	12

Err: Mass error.

表6 六个土壤来源菌株的16S rRNA基因序列比对结果

Table 6 16S rRNA gene sequence comparison of six strains isolated from a soil sample

Query	Top-hit type strain	Similarity (%)	Accession No.	Taxonomic status
S1-6	<i>Streptomyces virginiae</i> NRRL ISP-5094 ^T	99.71	PV450548	Correct name
S1-8	<i>Streptomyces thermocarboxydus</i> DSM 44293 ^T	99.35	PV450549	Correct name
S1-10	<i>Streptomyces lydicamycinicus</i> NBRC 110027 ^T	100.00	PV450550	Correct name
S1-13	<i>Streptomyces marianii</i> ICN19 ^T	99.85	PV450551	Not correct name
	<i>Streptomyces wuyuanensis</i> CGMCC 4.7042 ^T	99.50		Correct name
S1-16	<i>Streptomyces alboflavus</i> NRRL B-2373 ^T	99.71	PV450552	Correct name
S1-39	<i>Streptomyces anthocyanicus</i> NBRC 14892 ^T	100.00	PV450553	Synonym
	<i>Streptomyces violaceoruber</i> NBRC 12826 ^T	100.00		Correct name
	<i>Streptomyces tricolor</i> NBRC 15461 ^T	100.00		Synonym

MALDI-TOF MS 数据库构建。

MALDI-TOF MS 鉴定微生物的样本制备方法主要包括直接涂抹法、扩展的直接涂抹法和甲酸提取法。本研究发现,直接涂抹法对培养 2-3 d 的链霉菌单菌落(营养菌丝体)制备效果好,采集到的质谱图以核糖体蛋白质峰为主,杂峰少,这对得到准确的匹配数至关重要。直接涂抹法操作简单,15 min 可完成 1 个样本制备,5 min 可完成质谱采集。用到的靶板可以清洗后反复使用,制备过程仅消耗少量的基质和试剂,鉴定一个单菌落仅需 1-2 元。长期以来,链霉菌的菌种鉴定主要应用 16S rRNA 基因测序法,需要先挑取单菌落液体培养 1-2 d,再收集菌体提取基因组 DNA,PCR 扩增 16S rRNA 基因和 Sanger 测序,一般又需要 1-2 d。此外,一个链霉菌菌株 16S rRNA 基因全长测序的成本大概在 40-50 元。随着基因组测序技术的发展,基于基因组指标的鉴定方法能够更真实地反映类群间的进化关系,但基因组测序法周期长、成本高,限制了其在菌种鉴定中的应用。因此凭借显著的价格和速度优势,MALDI-TOF MS 法鉴定链霉菌显示出很大的应用潜力。

布鲁克的 MALDI Biotyper 数据库(Bruker MBT DB7854)收录了 457 种放线菌,但其中仅有 14 种链霉菌。本研究建立的链霉菌 MALDI-TOF MS 数据库 1.0 版本包含 615 种链霉菌,达到了较高水平的覆盖程度,极大地补充了商业数据库。随着微生物基因组测序的推进,一些尚未测序的链霉菌属菌种的全基因组测序数据仍会公开。同时,国际系统与进化微生物学杂志(*International Journal of Systematic and Evolutionary Microbiology*, IJSEM)已经规定新种的发表必须提供全基因组序列。因此,通过补充新测序的菌种数据,可以继续扩充数据库。通过对更多未知菌株进行鉴定可以全面分析鉴定方法的准确度。本研究中数据库建立基于基因组测序数据、基因功能注释和手动计算,不排除个别序列错误和计算错误的可能。后续使

用中,将对发现的错误数据及时进行更正,并定期发布更新完善的数据库版本,推动自建链霉菌数据库及鉴定方法的应用。

数据库是 MALDI-TOF MS 菌种鉴定技术的核心。鉴于构建标准图谱数据库的技术壁垒高,本研究对 MALDI-TOF MS 数据库构建方法进行了创新。基于微生物基因组数据挖掘构建数据库的方法不依赖仪器设备和配套软件,不需要培养入库菌株,具备一定系统分类学和生物信息学知识即可完成快速建库。通过数据库的开放共享有助于鉴定方法的推广。目前,国内很多高校和研究机构都购置了基质辅助激光解析飞行时间质谱仪,但设备的整体使用率不高。在相关部门鼓励大型仪器对外共享的背景下,很多机构提供送样测试服务,科研人员可通过送样测试得到待检菌株的质谱,再基于自建库或公开数据库比对完成菌种鉴定。

作者贡献声明

孙伟:项目提出,数据收集与分析,设计并完成实验,文章写作与修订;罗倩:数据分析,文章审阅;许杰:项目咨询,文章审阅;李雷:提供链霉菌属模式菌株,文章审阅;蒋海霞:文章审阅。

作者利益冲突公开声明

作者声明不存在任何可能会影响本文所报告工作的已知经济利益或个人关系。

参考文献

- [1] 尹盼盼,白雯静,李海燕,彭腾腾,石晓峰. 基质辅助激光解吸电离飞行时间质谱在微生物鉴定中的研究进展[J]. *微生物学通报*, 2023, 50(10): 4655-4666.
YIN PP, BAI WJ, LI HY, PENG TT, SHI XF. Research progress of matrix-assisted laser desorption ionization time-of-flight mass spectrometry in microbial identification[J]. *Microbiology China*, 2023, 50(10): 4655-4666 (in Chinese).
- [2] 黄琨波,郭思玲,张建国. MALDI-TOF MS 在临床微生物鉴定中的应用进展[J]. *国际检验医学杂志*, 2023, 44(S2): 209-213.
HUANG KB, GUO SL, ZHANG JM. Application

- progress of MALDI-TOF MS in clinical microbiological identification[J]. *International Journal of Laboratory Medicine*, 2023, 44(S2): 209-213 (in Chinese).
- [3] ZENG X, WANG Y, SHEN X, WANG H, XU ZL. Application of matrix-assisted laser desorption ionization time-of-flight mass spectrometry for identification of foodborne pathogens: current developments and future trends[J]. *Journal of Agricultural and Food Chemistry*, 2024, 72(40): 22001-22014.
- [4] 任婷, 史敏, 刘聪慧, 黄一刚, 裴金金. 基质辅助激光解吸电离飞行时间质谱在食源性致病菌鉴定中的应用[J]. *食品安全质量检测学报*, 2024, 15(5): 62-68.
REN T, SHI M, LIU CH, HUANG YG, PEI JJ. Application of matrix assisted laser desorption ionization time of flight mass spectrometry in identification of foodborne pathogens[J]. *Journal of Food Safety & Quality*, 2024, 15(5): 62-68 (in Chinese).
- [5] 徐菲, 杨凯, 朱龙吉, 林文芳, 崔丽. 土壤病原微生物检测技术研究进展[J]. *农业环境科学学报*, 2022, 41(12): 2593-2603.
XU F, YANG K, ZHU LJ, LIN WF, CUI L. Advances in the detection technology of pathogenic microorganisms in soil[J]. *Journal of Agro-Environment Science*, 2022, 41(12): 2593-2603 (in Chinese).
- [6] SILVA-JIMÉNEZ H, ARAUJO-PALOMARES CL, MACÍAS-ZAMORA JV, RAMÍREZ-ÁLVAREZ N, GARCÍA-LARA B, CORRALES-ESCOBOSA AR. Identification by MALDI-TOF MS of environmental bacteria with high potential to degrade pyrene[J]. *Journal of the Mexican Chemical Society*, 2018, 62(2): 214-225.
- [7] 周月霞, 胡成进, 陈英剑. MALDI-TOF MS 微生物鉴定数据库应用研究进展[J]. *临床检验杂志*, 2016, 34(6): 447-450.
- [8] 罗燕萍, 徐英春, 王辉, 喻华, 崔生辉, 鲁辛辛, 马筱玲, 胡继红, 苏建荣, 陈茶, 赵建宏, 顾兵, 马小军, 林勇平, 王明贵, 王皓峰, 张樱, 周巍, 陈荣, 李松, 等. 自建 MALDI-TOF MS 微生物鉴定数据库专家共识[J]. *中华检验医学杂志*, 2019, 42(6): 414-419.
LUO YP, XU YC, WANG H, YU H, CUI SH, LU XX, MA YL, HU JH, SU JR, CHEN C, ZHAO JH, GU B, MA XJ, LIN YP, WANG MG, WANG HF, ZHANG Y, ZHOU W, CHEN R, LI S, et al. Expert consensus on laboratory developed MALDI-TOF MS microbial identification database[J]. *Chinese Journal of Laboratory Medicine*, 2019, 42(6): 414-419 (in Chinese).
- [9] ARNOLD RJ, REILLY JP. Observation of *Escherichia coli* ribosomal proteins and their posttranslational modifications by mass spectrometry[J]. *Analytical Biochemistry*, 1999, 269(1): 105-112.
- [10] RYZHOV V, FENSELAU C. Characterization of the protein subset desorbed by MALDI from whole bacterial cells[J]. *Analytical Chemistry*, 2001, 73(4): 746-750.
- [11] 孙伟, 罗倩, 张薇, 吕霖烽, 祝新德, 许杰, 张凤丽. 利用基因组学和 MALDI-TOF MS 技术鉴定放线菌纲细菌的核糖体蛋白质标志物[J]. *微生物学报*, 2022, 62(7): 2671-2685.
SUN W, LUO Q, ZHANG W, LV JY, ZHU XD, XU J, ZHANG FL. Characterization of the ribosomal protein biomarkers of *Actinobacteria* via genomics and MALDI-TOF MS techniques[J]. *Acta Microbiologica Sinica*, 2022, 62(7): 2671-2685 (in Chinese).
- [12] 孙伟, 许杰, 周万平, 罗倩, 张薇. 基于核糖体蛋白质标志物及 MALDI-TOF MS 技术快速鉴定放线菌属放线菌[J]. *微生物学报*, 2023, 63(1): 430-445.
SUN W, XU J, ZHOU WP, LUO Q, ZHANG W. Rapid identification of *Lentzea* members: based on ribosomal protein biomarkers and MALDI-TOF MS[J]. *Acta Microbiologica Sinica*, 2023, 63(1): 430-445 (in Chinese).
- [13] 孙伟, 许杰, 张薇, 罗倩, 台萃. 基于核糖体蛋白质标志物及基质辅助激光解吸电离飞行时间质谱技术快速鉴定糖丝菌属放线菌[J]. *微生物学报*, 2023, 63(12): 4800-4813.
SUN W, XU J, ZHANG W, LUO Q, TAI C. Rapid identification of *Saccharothrix* spp. based on ribosomal protein biomarkers and matrix-assisted laser desorption/ionization-time of flight mass spectrometry[J]. *Acta Microbiologica Sinica*, 2023, 63(12): 4800-4813 (in Chinese).
- [14] LACEY HJ, RUTLEDGE PJ. Recently discovered secondary metabolites from *Streptomyces* species[J]. *Molecules*, 2022, 27(3): 887.
- [15] 董雷, 韩嘉瑞, 李帅, 岳凌霄, 李文均. 链霉菌最新研究进展[J]. *微生物学报*, 2023, 63(5): 1815-1832.
DONG L, HAN JR, LI S, YUE LX, LI WJ. The latest research progress of *Streptomyces*[J]. *Acta Microbiologica Sinica*, 2023, 63(5): 1815-1832 (in Chinese).
- [16] HIREL PH, SCHMITTER MJ, DESSEN P, FAYAT G, BLANQUET S. Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1989, 86(21): 8247-8251.
- [17] KOSTAS J, PARKER KC. Using matrix-assisted laser desorption/ionization time of flight spectra to elucidate species boundaries by matching to translated DNA databases[J]. *Journal of the American Society for Mass Spectrometry*, 2020, 31(1): 73-84.
- [18] MEIER-KOLTHOFF JP, GÖKER M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy[J]. *Nature Communications*, 2019, 10: 2182.
- [19] LI CJ, WANG M, ZHANG Z, ZHI XY, YANG LL. Genome-based analyses reveal heterotypic synonyms of *Streptomyces* species and associated subspecies[J]. *Archives of Microbiology*, 2022, 204(9): 581.
- [20] MADHAIYAN M, SARAVANAN VS, SEE-TOO WS. Genome-based analyses reveal the presence of 12 heterotypic synonyms in the genus *Streptomyces* and emended descriptions of *Streptomyces bottropensis*, *Streptomyces celluloflavus*, *Streptomyces fulvissimus*, *Streptomyces glaucescens*, *Streptomyces murinus*, and *Streptomyces variegatus*[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2020, 70(6): 3924-3929.