



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

孙伟*, 罗倩, 张薇, 吕霁焯, 祝新德, 许杰, 张凤丽

上海交通大学生命科学技术学院, 上海 200240

孙伟, 罗倩, 张薇, 吕霁焯, 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. *微生物学报*, 2022, 62(7): 2671–2685.

Sun Wei, Luo Qian, Zhang Wei, Lv Jiyang, Zhu Xinde, Xu Jie, Zhang Fengli. Characterization of the ribosomal protein biomarkers of *Actinobacteria* via genomics and MALDI-TOF MS techniques. *Acta Microbiologica Sinica*, 2022, 62(7): 2671–2685.

摘要: 【目的】基质辅助激光解吸电离飞行时间质谱(MALDI-TOF MS)法基于微生物的特征蛋白指纹图谱鉴定菌种, 本研究利用基因组学和 MALDI-TOF MS 技术鉴定放线菌纲细菌的核糖体蛋白质标志物。【方法】从 MALDI-TOF MS 图谱数据库选取放线菌纲代表菌种, 在基因组数据库检索目标菌种, 获取目标菌株或其参比菌株的核糖体蛋白质序列, 计算获得分子质量理论值, 用于注释目标菌株 MALDI-TOF MS 指纹图谱中的核糖体蛋白质信号。【结果】从 8 目, 24 科, 53 属, 114 种, 142 株放线菌的 MALDI-TOF MS 图谱中总共注释出 31 种核糖体蛋白质。各菌株的指纹图谱中核糖体蛋白质信号数量差异显著。各种核糖体蛋白质信号的注释次数差异显著。总共 15 种核糖体蛋白质在超过半数图谱中得到注释, 注释次数最高的是核糖体大亚基蛋白质 L36。【结论】本研究找到了放线菌纲细菌 MALDI-TOF MS 图谱中常见的 15 种核糖体蛋白质信号, 可为通过识别核糖体蛋白质的质谱特征峰鉴定放线菌的方法建立提供依据。

关键词: 基质辅助激光解吸电离飞行时间质谱; 放线菌纲; 核糖体蛋白质; 标志物; 质谱注释

基金项目: 国家自然科学基金(81973230); 上海交通大学决策咨询课题(JCZXSJB-13)

Supported by the National Natural Science Foundation of China (81973230) and by the Decision-Making Consultation Project of Shanghai Jiao Tong University (JCZXSJB-13)

*Corresponding author. E-mail: bacterior@sjtu.edu.cn

Received: 2 November 2021; Revised: 15 January 2022; Published online: 16 February 2022

Characterization of the ribosomal protein biomarkers of *Actinobacteria* via genomics and MALDI-TOF MS techniques

SUN Wei*, LUO Qian, ZHANG Wei, LV Jiyang, ZHU Xinde, XU Jie, ZHANG Fengli

School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract: [Objective] Matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) technique provides mass spectral fingerprints of characteristic proteins for microbial identification. This study aims to characterize the ribosomal protein biomarkers from *Actinobacteria* via genomics and MALDI-TOF MS techniques. [Methods] *Actinobacteria* representatives were chosen from MALDI-TOF MS spectral library. By searching against genome database, we acquired the ribosomal protein sequences from the target or reference strains of target species and calculated the theoretical molecular masses. The mass peaks in the MALDI-TOF mass spectra of target strains were annotated with the calculated molecular masses of ribosomal proteins. [Results] Mass peaks annotated in the spectra of 142 strains of 114 species, 53 genera, 24 families from 8 orders were assigned to 31 ribosomal proteins. The number of annotated ribosomal proteins varied significantly among strains. The number of mass peak annotations also varied considerably among different subunit proteins. A total of 15 ribosomal proteins were annotated in over half of the spectra, and the ribosomal protein with the most mass peak annotations was L36. [Conclusion] This study identified 15 common ribosomal protein mass peaks in MALDI-TOF mass spectra of *Actinobacteria*. The results can support the establishment of a method for the identification of *Actinobacteria* by matching characteristic mass peaks of ribosomal protein biomarkers in MALDI-TOF mass spectra.

Keywords: MALDI-TOF MS; *Actinobacteria*; ribosomal protein; biomarker; annotation of mass peaks

基质辅助激光解吸电离飞行时间质谱 (MALDI-TOF MS) 指纹图谱法是一种新兴的微生物鉴定方法, 具有简单、快速、高通量等优势, 已应用于临床诊断、环境监测和食品安全等领域^[1-6]。这一方法建立在 MALDI-TOF MS 法测量微生物稳定表达的高丰度蛋白质(如核糖体蛋白质)的基础上^[7-8]。采集样品的 MALDI-TOF MS 图谱, 通过比对数据库中已知菌种的指纹图谱获得鉴定结果, 鉴定准确性依赖于图谱数据库的完善性与可靠性。目前, 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)收录环境微生物的指纹图谱较少, 限制了其在环境微生物鉴定中的应用^[9]。

为了解决这一问题, 厂商提供了一套标准化自建库方法, 其流程包括菌株收集、培养、前处理、图谱采集、质量控制和方法验证, 但标准化建库方法对技术要求严格, 而且操作繁琐, 费用高昂, 限制了其进一步推广^[10]。因此, 需要开发更高效、经济的微生物质谱鉴定方法, 以扩展其应用范围。

近年来不依赖图谱数据库的 MALDI-TOF MS 法鉴定微生物引起人们关注。这一方法建立在从目标菌种的基因组/蛋白质组序列中获取生物标志物的序列并预测分子质量基础上^[11]。采集样品的 MALDI-TOF MS 图谱, 通过识别生物标志物的质谱特征峰鉴定菌种。细菌全基因组测序

技术的普及为细菌基因组数据挖掘及重要基因的功能分析构建了全面的研究平台,也为这一方法的探索创造了有利条件。核糖体蛋白质是参与构成核糖体的所有蛋白质的统称,包括小亚基和大亚基核糖体蛋白质,分别以 SX 和 LX (X 为数字)命名^[12]。核糖体蛋白质呈现高度的保守性,是一类重要的生物标志物。江文静等以 13 种核糖体蛋白质作为生物标志物,以 MALDI-TOF MS 为测试手段鉴定了蓝藻样品中的模式藻株铜绿微囊藻^[13]。Cheng 等使用 10 种蛋白质(包括 8 种核糖体蛋白质)作为生物标志物,通过 MALDI-TOF MS 分析实现了 8 个常见细菌属菌种的鉴定^[14]。Kostas 和 Parker 选择 65 种蛋白质(包括 52 种核糖体蛋白质)作为生物标志物,通过 MALDI-TOF MS 分析实现了弧菌目和环境样品中一些菌种的鉴定^[15]。这些研究表明选择合适的生物标志物,可以对环境中特定类群的微生物实现快速、准确鉴定,显示出这一方法广阔的应用前景。

放线菌纲是一个常见的细菌类群,种类多样,分布广泛,在自然界物质循环中发挥重要作用,还是产生抗生素类药物及其他活性物质的重要资源^[16]。目前应用 MALDI-TOF MS 方法鉴定环境放线菌的研究很少。我们应用 MALDI-TOF MS 图谱比对法对土壤、淤泥、海洋沉积物等环境来源的放线菌进行鉴定时发现得到可信鉴定结果的菌株比例低,一般占测试菌株总数的 10%–30%。MALDI Biotyper 数据库收录的放线菌纲细菌达 457 种,但种类以致病菌为主,是环境放线菌鉴定成功率低的主要原因。目前,基因组数据库收录的放线菌纲细菌达 2 700 余种并且来源广泛,并不限于致病菌,可获得从菌株基因组注释出的核糖体蛋白质序列。因此,本研究基于全基因组测序和核糖体蛋白质基因注释,对 MALDI Biotyper 数据库中 114 种放线菌纲细菌 MALDI-TOF MS 指纹图谱中的核糖体蛋白质信

号进行注释,寻找可用于放线菌鉴定的核糖体蛋白质标志物,以期通过识别核糖体蛋白质的质谱特征峰鉴定放线菌的方法建立提供依据。

1 材料与方法

1.1 MALDI-TOF MS 图谱数据库及菌种选取

MALDI Biotyper 数据库 (Bruker MBT DB7854) 购自德国布鲁克公司,涵盖放线菌纲 8 目, 27 科, 60 属, 457 种。从数据库选取 8 目, 24 科, 53 属, 114 种, 142 株放线菌(表 1) 注释其指纹图谱中的核糖体蛋白质信号。

1.2 基因组数据库及核糖体蛋白质序列提取

美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI)基因组数据库 (<http://www.ncbi.nlm.nih.gov/genome/>) 收录了 2 700 余种放线菌纲细菌的参考基因组信息,涵盖大多数已知菌种,从中检索 114 个目标菌种,挑选目标菌株或其参比菌株(与目标菌株同种的其他菌株)。在菌株基因组组装与注释报告的蛋白质编码基因(protein coding genes, CDS)栏汇总了 NCBI 原核基因组注释系统(Prokaryotic Genome Annotation Pipeline, PGAP)注释出的蛋白质信息^[17],从中检索核糖体蛋白质并提取序列。

1.3 核糖体蛋白质序列分析

在生物体内核糖体合成的前体蛋白需进行一系列翻译后加工、修饰才能成为具有功能的成熟蛋白。为了准确预测核糖体蛋白质分子质量,必须考虑可能的翻译后加工、修饰等因素。原核生物普遍存在的一种翻译后加工是 N-端起始甲硫氨酸的切除^[18–19],当 N-端的第二个氨基酸是甘氨酸、丙氨酸、脯氨酸、丝氨酸、苏氨酸、缬氨酸或半胱氨酸时,起始甲硫氨酸被切除,当 N-端的第二个氨基酸是其他氨基酸时,起始甲硫氨酸不被切除。分析核糖体蛋白质序列,如符合切除规则,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。

表 1 从 MALDI Biotyper 数据库选取的放线菌及其质谱的核糖体蛋白质信号注释数

Table 1 The number of ribosomal protein (RP) mass peaks annotated in the MALDI-TOF mass spectra of various *Actinobacteria* chosen from MALDI Biotyper Database

Order	Family	Genus	Number of species	Number of strains	Number of annotated RPs	Average number of annotated RPs	
<i>Actinomycetales</i>	<i>Actinomycetaceae</i>	<i>Actinobaculum</i>	2	2	38	19	
		<i>Actinomyces</i>	3	8	96	12	
		<i>Arcanobacterium</i>	2	2	27	14	
		<i>Varibaculum</i>	1	1	3	3	
		<i>Mobiluncus</i>	1	1	16	16	
<i>Streptomycetales</i>	<i>Streptomycetaceae</i>	<i>Streptomyces</i>	7	7	92	13	
		<i>Kitasatospora</i>	1	1	19	19	
<i>Streptosporangiales</i>	<i>Nocardiopsaceae</i>	<i>Nocardiopsis</i>	1	1	8	8	
<i>Micromonosporales</i>	<i>Micromonosporaceae</i>	<i>Micromonospora</i>	5	7	95	14	
<i>Pseudonocardiales</i>	<i>Pseudonocardiaceae</i>	<i>Pseudonocardia</i>	1	1	15	15	
		<i>Amycolatopsis</i>	5	5	75	15	
		<i>Saccharopolyspora</i>	2	2	31	16	
		<i>Lentzea</i>	1	1	13	13	
		<i>Corynebacteriales</i>	<i>Corynebacteriaceae</i>	<i>Corynebacterium</i>	5	5	74
	<i>Dietziaceae</i>	<i>Dietzia</i>	3	3	15	5	
	<i>Mycobacteriaceae</i>	<i>Mycobacterium</i>	4	4	36	9	
	<i>Nocardiaceae</i>	<i>Nocardia</i>	3	4	61	15	
		<i>Rhodococcus</i>	3	4	54	14	
	<i>Gordoniaceae</i>	<i>Gordonia</i>	3	6	70	12	
	<i>Tsukamurellaceae</i>	<i>Tsukamurella</i>	1	5	35	7	
<i>Micrococcales</i>	<i>Micrococcaceae</i>	<i>Micrococcus</i>	3	7	93	13	
		<i>Arthrobacter</i>	3	3	47	16	
		<i>Kocuria</i>	3	5	74	15	
		<i>Rothia</i>	3	3	50	17	
		<i>Microbacterium</i>	3	4	54	14	
		<i>Agrococcus</i>	3	3	48	16	
		<i>Agromyces</i>	3	3	36	12	
		<i>Curtobacterium</i>	2	2	24	12	
		<i>Leifsonia</i>	1	1	17	17	
		<i>Clavibacter</i>	1	1	14	14	
		<i>Pseudoclavibacter</i>	1	1	11	11	
		<i>Leucobacter</i>	2	2	31	16	
		<i>Rathayibacter</i>	1	1	13	13	
		<i>Brevibacteriaceae</i>	<i>Brevibacterium</i>	3	4	53	13
		<i>Cellulomonadaceae</i>	<i>Cellulomonas</i>	3	3	44	15
		<i>Dermabacteraceae</i>	<i>Brachybacterium</i>	3	4	36	9
			<i>Dermabacter</i>	1	2	27	14
		<i>Dermacoccaceae</i>	<i>Dermacoccus</i>	1	1	12	12
		<i>Kytococcaceae</i>	<i>Kytococcus</i>	2	2	24	12
		<i>Dermatophilaceae</i>	<i>Dermatophilus</i>	1	1	9	9
			<i>Arsenicococcus</i>	1	1	18	18
		<i>Intrasporangiaceae</i>	<i>Janibacter</i>	1	1	12	12
			<i>Terrabacter</i>	1	1	10	10
		<i>Jonesiaceae</i>	<i>Jonesia</i>	1	2	14	7
		<i>Promicromonosporaceae</i>	<i>Cellulosimicrobium</i>	1	1	14	14
			<i>Oerskovia</i>	1	1	14	14
		<i>Propionibacteriales</i>	<i>Propionibacteriaceae</i>	<i>Luteococcus</i>	1	1	17
<i>Propionibacterium</i>	3			3	31	10	
<i>Propioniferax</i>	1			2	17	9	
<i>Propionimicrobium</i>	1			1	12	12	
<i>Tessaracoccus</i>	1			1	16	16	
<i>Nocardioideaceae</i>	<i>Aeromicrobium</i>			2	2	13	7
	<i>Nocardioides</i>			2	2	12	6

1.4 核糖体蛋白质理论分子量计算

应用蛋白质基本性质分析工具 ProtParam (<http://web.expasy.org/protparam/>)^[20] 计算核糖体蛋白质平均分子量作为分子量理论值。

1.5 MALDI-TOF MS 指纹图谱中的核糖体蛋白质信号注释

基于目标菌株或其参比菌株各种核糖体蛋白质的分子量理论值进行目标菌株的质谱注释。蛋白质通过基质辅助激光解吸电离(MALDI)产生带一个电荷的阳离子 $[M+H]^+$, 有时产生带两个电荷的阳离子 $[M+2H]^{2+}$ ^[7-8], 质谱注释时需考虑蛋白质电离后带一个和两个电荷的情况。此外, 一些原核生物的个别核糖体蛋白质会发生甲基化(如 L33、S11)或乙酰化(如 S18)修饰^[7,15], 导致理论分子量增加 14 或 42 Da, 质谱注释时需注意这些修饰造成的分子量变化。

考虑到 MALDI-TOF MS 质谱仪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^[15]。若质量轴偏移, 适当提高最大质量误差, 但不超过 ± 800 ppm。若质谱中有 m/z 处在某种核糖体蛋白质分子量理论值 $M+1$ 或 $(M+2)/2$ 的误差范围内, 则将该数值注释为对应的核糖体蛋白质信号, 并记录 m/z 丰度(相对百分比)及其在峰列表中的排序。依次进行目标菌株的质谱注释并统计各菌株图谱中注释的核糖体蛋白质信号数量以及各种核糖体蛋白质信号在分析菌株中的出现次数。

1.6 测试菌株

为测试前文分析未涉及的放线菌纲菌种 MALDI-TOF MS 图谱中目标核糖体蛋白质信号的出现情况, 选取从土壤样本中分离的 3 个菌株进行 MALDI-TOF MS 测试和分析, 包括根际动球菌(*Kineococcus rhizosphaerae*) D24 (16S rRNA 基因序列 GenBank 注册号: OL519199, 与模式菌株 *K. rhizosphaerae* DSM 19711^T 的序列相似

度为 99.28%)、江西伦茨菌(*Lentzea jiangxiensis*) 3-5 (OL519200, 与模式菌株 *L. jiangxiensis* CGMCC 4.6609^T 的序列相似度为 99.63%)和食醚红球菌(*Rhodococcus aetherivorans*) 4-12 (OL519201, 与模式菌株 *R. aetherivorans* DSM 44752^T 的序列相似度为 99.93%)。

1.7 菌株蛋白提取

取待测微生物样本(来自单个菌落, 取一个接种环的量)转移到盛有 20 μ L 70%甲酸水溶液的 1.5 mL Eppendorf 管中, 用移液枪反复吹打, 涡旋充分混匀。加入 20 μ L 乙腈, 用移液枪反复吹打。13 000 r/min 离心 2 min。将 1 μ L 微生物提取物上清液移取到靶板上, 室温晾干。将 1 μ L α -氰基-4-羟基肉桂酸(HCCA)基质溶液滴加覆盖样本, 室温晾干。

1.8 MALDI-TOF MS 测试

基质辅助激光解析高分辨飞行时间质谱成像系统(MALDI TOF, ultraflextreme)购自德国布鲁克公司。使用控制软件 flexControl 3.4 采集质谱图, 仪器测定参数: 采集质荷比(m/z)为 4 000–12 000; 激光点击数 1 000; 测定模式为阳离子线性模式。使用分析软件 flexAnalysis 3.4 处理质谱图并标峰。

1.9 测试菌株的核糖体蛋白质鉴定

按前文所述方法从 NCBI 基因组数据库提取 3 个参比菌株(*K. rhizosphaerae* DSM 19711^T、*L. jiangxiensis* CGMCC 4.6609^T 和 *R. aetherivorans* PSBB011)的 15 种目标核糖体蛋白质(L36、L35、L34、L33、L32、L31、L30、L29、L28、L27、S20、S19、S18、S15 和 S14)序列并计算其分子量理论值, 注释 3 个测试菌株 MALDI-TOF MS 指纹图谱中的核糖体蛋白质信号并统计注释出的信号数量。

2 结果与分析

2.1 MALDI-TOF MS 指纹图谱数据分析

MALDI Biotyper 数据库中菌株的

MALDI-TOF MS 指纹图谱数据以峰列表形式呈现,一般包含 70 或 80 个峰的信息,包括质荷比 m/z 、丰度、权重、出现频率等(表 2)。对 142 个菌株的质谱峰列表分析发现 m/z 范围在 3 000–14 000, 以 10 000 以下的峰居多, 10 000 以上的峰数目较少, 可能与样品制备时使用的基质为 HCCA 有关^[8]。质谱峰以单电荷峰为主, 但不乏一些双电荷峰存在, 尤其是高丰度蛋白质, 可同时产生单电荷和双电荷峰。

2.2 核糖体蛋白质检索及序列分析

从 NCBI 基因组数据库放线菌纲菌种的注释蛋白质列表中可检索到 54 种核糖体蛋白质^[21], 包括 50S 核糖体大亚基蛋白质 34 种(L1–L7、L9–L25 和 L27–L36)以及 30S 核糖体

小亚基蛋白质 20 种(S1–S20), 其中一些核糖体蛋白质有多种类型, 如 L31 和 S14, 前者有 L31 和 L31-B 两种类型, 后者有 S14-Z 和 S14 两种类型。大多数菌株可从其基因组注释出 L31、L31-B、S14-Z 和 S14, 而 5 个科(棒杆菌科、迪茨氏菌科、微球菌科、微杆菌科和短杆菌科)的菌株注释出 L31-B 和 S14。此外, 注释出的部分核糖体蛋白质有 2 个甚至多个序列, 如 L36、L33、L32、L31、L28 和 S18 等, 这种情况下将这些序列全部提取计算分子量。序列分析发现大多数核糖体蛋白质符合 N-端起始甲硫氨酸切除规则, 只有少数例外, 如 L36 的 N-端第二个氨基酸通常是赖氨酸, N-端起始甲硫氨酸不切除。

表 2 MALDI Biotyper 数据库中菌株质谱峰列表(以藤黄微球菌 *M. luteus* DSM 1790 为例)

Table 2 The peak lists obtained from the mass spectra of the strains in the MALDI Biotyper Database (using *M. luteus* DSM 1790 as an example)

m/z	Intensity/%	Weight/%	Frequency/%	In peak lists
5 923.33	61.41	100.00	100.0	24
5 964.91	6.67	100.00	100.0	24
6 061.45	27.05	100.00	100.0	24
6 114.05	18.33	100.00	100.0	24
6 193.93	10.22	100.00	100.0	24
6 246.00	75.08	100.00	100.0	24
6 385.35	6.62	100.00	100.0	24
6 429.87	93.14	100.00	100.0	24
6 538.60	7.63	100.00	100.0	24
6 778.08	13.55	100.00	100.0	24
6 884.48	16.32	100.00	100.0	24
7 233.45	50.26	100.00	100.0	24
7 316.53	6.52	100.00	100.0	24
7 386.13	67.86	100.00	100.0	24
7 577.98	8.55	100.00	100.0	24
7 624.90	11.17	100.00	100.0	24
7 932.02	8.38	100.00	100.0	24
8 107.38	48.69	100.00	100.0	24
8 354.55	24.36	100.00	100.0	24
8 418.95	29.10	100.00	100.0	24

2.3 MALDI-TOF MS 指纹图谱核糖体蛋白质信号注释

142 个放线菌纲菌株 MALDI-TOF MS 指纹图谱的核糖体蛋白质信号注释结果见附表 1, 信号数量统计结果见附表 2。

结果表明 97 个菌株的质谱中核糖体蛋白质峰的质量误差范围为-500~+500 ppm。另外 45 个菌株的质谱因质量轴偏移导致误差范围改变, 如 048 混浊红球菌 (*Rhodococcus opacus*) DSM 43251 质谱中各个核糖体蛋白质峰的实测值均比理论值低, 误差范围为-700~-100 ppm, 075 苏别蒂山农霉菌 (*Agromyces subbeticus*) HKI 340_DSM 16689^T 质谱中各个核糖体蛋白质峰的实测值均比理论值高, 误差范围为 0~+700 ppm。为了不遗漏核糖体蛋白质信号, 对这类质谱调整误差范围后进行注释。

质谱注释发现 29 个菌株的质谱中出现不同核糖体蛋白质 m/z 接近的情况, 包括两种核糖体蛋白质的单电荷峰 m/z 接近以及一种核糖体蛋白质的单电荷峰和另一种核糖体蛋白质的双电荷峰 m/z 接近(表 3)。以 001 马赛放杆菌 (*Actinobaculum massiliense*) FC3 为例, L33 和 L30 的理论分子量分别为 6 436.37 Da 和 6 435.44 Da, 对于质谱中 m/z 6 436.23 的峰, 无法判断其匹配两种(L33 和 L30)还是一种蛋白质(L33 或 L30)。以 010 栗褐链霉菌 (*Streptomyces badius*) B192 UFL 为例, S14-Z 和 S12 的理论分子量分别为 6 818.14 Da 和 13 639.83 Da, 对于质谱中 m/z 6 821.29 的峰, 无法判断其匹配两种(S14-Z 的单电荷峰和 S12 的双电荷峰)还是一种蛋白质(S14-Z 的单电荷峰或 S12 的双电荷峰)。为了不遗漏核糖体蛋白质信号, 将这类 m/z 注释为两种核糖体蛋白质。

表 3 质谱中一个 m/z 匹配两种核糖体蛋白质分子离子的菌株

Table 3 The list of the strains in the mass spectra of which one m/z matched two ribosomal protein ions

Strain	Ribosomal protein ion	Theoretical mass (m/z)	Experimental mass (m/z)	Error	
				Da	ppm
<i>Actinobaculum massiliense</i> ME0003421_c LDW	L30 ⁺	6 436.44	6 436.23	-0.21	-33
	L33 ⁺	6 437.37		-1.14	-177
<i>Kitasatospora phosalacinea</i> HKI 222 HKJ	L27 ⁺	8 810.89	8 812.34	1.45	165
	S18 ⁺	8 813.30		-0.96	-109
	L30 ⁺	6 825.83	6 826.90	1.07	157
	S12 ²⁺	6 828.45		-1.55	-226
<i>Mycobacterium gordonae</i> 0714 BSI	L32 ⁺	6 350.47	6 349.73	-0.74	-117
	L33 ⁺	6 351.29		-1.56	-246
<i>Nocardia araoensis</i> DSM 44729 ^T	L33 ⁺	6 493.45	6 493.29	-0.16	-25
	L30 ⁺	6 495.51		-2.22	-342
<i>Arthrobacter crystallopoietes</i> DSM 20117 ^T	L28 ⁺	8 583.98	8 585.21	1.23	143
	S18 ⁺	8 584.13		1.08	126
<i>Arthrobacter stackebrandtii</i> DSM 16005 ^T	L27 ⁺	8 844.00	8 847.32	3.32	375
	S20 ⁺	8 845.28		2.04	231
<i>Kocuria marina</i> CICC 23948	L19 ⁺	13 531.47	13 532.33	0.86	64
	S12 ⁺	13 531.94		0.39	29
<i>Kocuria marina</i> CICC 24030	L19 ⁺	13 531.47	13 529.70	-1.77	-131
	S12 ⁺	13 531.94		-2.24	-166

(待续)

(续表 3)

<i>Agromyces hippuratus</i> HKI 11533_DSM 8598 ^T	S18 ⁺	9 222.75	9 225.89	3.14	340
	S20 ⁺	9 224.69		1.2	130
<i>Brevibacterium marinum</i> DSM 18964 ^T	L31-B ⁺	9 546.75	9 548.13	1.38	144
	S20 ⁺	9 548.93		-0.8	-84
<i>Oerskovia turbata</i> B258 UFL	L27 ⁺	8 814.83	8 816.09	1.26	143
	L29 ⁺	8 818.02		-1.93	-219
<i>Arsenicococcus bolidensis</i> DSM 15745 ^T	L28 ⁺	6 801.91	6 803.96	2.05	301
	S14-Z ⁺	6 807.08		-3.12	-459
<i>Aeromicrobium massiliense</i> DSM 25782 ^T	L33 ⁺	6 661.64	6 661.32	-0.32	-48
	L30 ⁺	6 661.66		-0.34	-51
<i>Streptomyces badius</i> B192 UFL	S14-Z ⁺	6 819.14	6 821.29	2.15	315
	S12 ²⁺	6 820.92		0.37	55
<i>Streptomyces chartreusis</i> HKI 249 HKJ	S14-Z ⁺	6 819.14	6 818.05	-1.09	-160
	S12 ²⁺	6 820.92		-2.87	-420
<i>Streptomyces galilaeus</i> HKI 22 HKJ	S14-Z ⁺	6 819.14	6 819.58	0.44	65
	S12 ²⁺	6 820.92		-1.34	-196
<i>Streptomyces phaeochromogenes</i> B265 UFL	S14-Z ⁺	6 819.14	6 821.71	2.57	377
	S12 ²⁺	6 820.92		0.79	117
<i>Streptomyces violaceoruber</i> B263 UFL	S14-Z ⁺	6 819.14	6 820.14	1.00	147
	S12 ²⁺	6 820.92		-0.78	-114
<i>Micromonospora viridifaciens</i> DSM 43909 ^T	L34 ⁺	5 198.04	5 197.81	-0.23	-44
	S17 ²⁺	5 199.54		-1.73	-332
	L30 ⁺	6 829.94	6 830.62	0.68	100
	S12 ²⁺	6 829.97		0.65	95
<i>Saccharopolyspora erythraea</i> HKI 184 HKJ	L28 ⁺	6 790.91	6 789.25	-1.66	-244
	S12 ²⁺	6 788.40		0.85	125
<i>Rhodococcus yunnanensis</i> DSM 44837 ^T	L28 ⁺	6 645.66	6 642.18	-3.48	-524
	L7/L12 ²⁺	6 641.59		0.59	89
<i>Gordonia rubripertincta</i> DSM 43197 ^T	S14-Z ⁺	6 775.08	6 771.93	-3.15	-465
	S13 ²⁺	6 772.35		-0.42	-62
<i>Gordonia rubripertincta</i> DSM 46066	S14-Z ⁺	6 775.08	6 774.13	-0.95	-140
	S13 ²⁺	6 772.35		1.78	263
<i>Clavibacter michiganensis</i> subsp. <i>tessellarius</i> DSM 20741 ^T	L33 ⁺	6 484.45	6 484.92	0.47	72
	L19 ²⁺	6 485.52		-0.6	-92
<i>Pseudoclavibacter helvolus</i> DSM 20419 ^T	L33 ⁺	6 558.62	6 560.70	2.08	317
	S6 ²⁺	6 558.91		1.79	273
<i>Dermacoccus nishinomiyaensis</i> DSM 20448 ^T	L34 ⁺	5 192.08	5 192.77	0.69	133
	S19 ²⁺	5 194.46		-1.69	-325
<i>Janibacter limosus</i> DSM 11140 ^T	L33 ⁺	6 486.45	6 483.43	-3.02	-466
	L18 ²⁺	6 482.46		0.97	150
<i>Terrabacter tumescens</i> DSM 20308 ^T	L30 ⁺	6 643.64	6 643.81	0.17	26
	L22 ²⁺	6 644.64		-0.83	-125
<i>Nocardioides simplex</i> DSM 20130 ^T	L32 ⁺	6 627.69	6 626.42	-1.27	-192
	L7/L12 ²⁺	6 628.56		-2.14	-323

质谱注释核糖体蛋白质 L33 时发现链霉菌科菌株(6 株链霉菌和 1 株北里孢菌)和皮杆菌科菌株(2 株短状杆菌和 2 株皮杆菌)出现与其他菌株不同的情况, 其质谱中均缺少与 L33 理论分子量相匹配的 m/z , 但出现比理论值增加了 14 ± 2 的 m/z (表 4)。鉴于有研究发现了一些原核生物中核糖体蛋白质 L33 的甲基化修饰现象^[7-8], 推测这些菌株的 L33 发生甲基化修饰并将这些 m/z 注释为 L33。

对于存在 2 种类型的核糖体蛋白质 L31 和 S14, 大多数菌株的质谱中注释出 L31 和 S14-Z, 极少注释出 L31-B 和 S14, 而 L31 和 S14-Z 缺失的微球菌科和微杆菌科菌株的质谱中普遍注释出 L31-B 和 S14。这一结果表明放线菌纲细菌基因组中一些核糖体蛋白质存在 2 套基因, 当其中 1 套缺失时, 另一套会表达合成核糖体蛋白质。

质谱注释显示各类群放线菌的 MALDI-TOF MS 指纹图谱中核糖体蛋白质信号数量差异显著。从属水平分析, 各属菌株的质谱中注释出的核糖体蛋白质平均数量差异显著(表 1), 最多高达 19 种(放杆菌和北里孢菌), 最少仅 3 种(弯曲短杆菌)。53 个分析属中 51 个属注释出的核

糖体蛋白质平均数量不少于 6 种, 另外 2 个(迪茨氏菌和弯曲短杆菌)注释出的平均数量少于 6 种。从菌株水平分析, 不同菌株的质谱中注释出的核糖体蛋白质数量差异显著, 最多高达 20 种 [001 马赛放杆菌(*Actinobaculum massiliense*) FC3 和 064 空气罗氏菌(*Rothia aeria*) CCUG 25688], 最少仅 2 种 [038 肉桂色迪茨氏菌(*Dietzia cinnamea*) 117 RLT]。同一菌种不同菌株间注释出的核糖体蛋白质数量显示出一定程度的波动。以 003 口腔放线菌(*Actinomyces oris*)为例, 4 个分析菌株(CCUG 54866、Cory_54 IBS、BK495_10 ERL 和 BK03176_10 ERL)注释出的核糖体蛋白质数量分别为 14、16、11 和 14。类似情况也在其他菌种中出现, 一方面体现出菌株样本间核糖体蛋白质提取效率的差异, 另一方面, 同一菌种不同菌株的某(几)种核糖体蛋白质可能因个别氨基酸突变导致分子量变化而未成功注释。142 个菌株中注释出至少 10 种核糖体蛋白质的菌株共计 111 个, 占菌株总数的 78.2%, 注释出至少 6 种核糖体蛋白质的菌株共计 134 个, 占菌株总数的 94.4%, 注释出少于 6 种核糖体蛋白质的菌株合计 8 个, 占菌株总数的 5.6%。

表 4 核糖体蛋白质 L33 可能发生甲基化修饰的菌株

Table 4 The list of the strains whose subunit L33 was likely methylated

Family	Strain	L33 [M+H] ⁺		
		Theoretical mass (m/z)	Experimental mass (m/z)	Error/Da
<i>Streptomycetaceae</i>	<i>Streptomyces badius</i> B192 UFL	6 292.24	6 307.07	14.83
	<i>Streptomyces chartreusis</i> HKI 249 HKJ	6 276.24	6 288.49	12.25
	<i>Streptomyces galilaeus</i> HKI 22 HKJ	6 294.27	6 307.36	13.09
	<i>Streptomyces lavendulae</i> B264 UFL	6 310.27	6 325.01	14.74
	<i>Streptomyces phaeochromogenes</i> B265 UFL	6 276.24	6 292.15	15.91
	<i>Streptomyces violaceoruber</i> B263 UFL	6 276.24	6 289.77	13.53
	<i>Kitasatospora phosalacinea</i> HKI 222 HKJ	6 293.22	6 305.18	11.96
<i>Dermabacteraceae</i>	<i>Brachybacterium faecium</i> DSM 4810 ^T	6 295.26	6 307.24	11.98
	<i>Brachybacterium muris</i> 7 RLT	6 329.29	6 341.98	12.69
	<i>Dermabacter hominis</i> DSM 7083 ^T	6 329.27	6 341.89	12.62
	<i>Dermabacter hominis</i> 13099930_2_g MVD	6 329.27	6 343.19	13.92

从 142 个菌株的 MALDI-TOF MS 图谱中总共注释出 31 种核糖体蛋白质, 包括 19 种大亚基蛋白质和 12 种小亚基蛋白质, 其信号在分析菌株中的出现次数差异显著(表 5)。低分子量(<10 000 Da)核糖体蛋白质(L36、L35、L34、L33、L32、L31、L30、L29、L28、L27、L31-B、S20、S18 和 S14-Z)信号的出现次数高于高分子量 (>10 000 Da)核糖体蛋白质信号。L36 是分子量最低的核糖体蛋白质, 由 37–40 个氨基酸构成, 分子量范围在 4 280–4 710 Da。L36 的信号在分析菌株中出现次数最高, 为 127 次, 且普遍呈现高丰度, 其丰度在 44 个菌株的质谱峰中排第一, 在 85 个菌株的质谱峰中排前三。总共 15 种核糖体蛋白质(L36、L29、L28、L32、L27、L30、L33、L34、S18、S14、S20、L35、L31、S15 和 S19)信号在超过半数菌株的图谱中出现, 可通过这些质谱特征峰的组合表征各菌种。

2.4 核糖体蛋白质在菌株水平的鉴别能力分析

为评估核糖体蛋白质在菌株水平的鉴别能

力, 本研究分析了部分菌种的多个菌株, 发现大多数情况下核糖体蛋白质的质谱特征峰组合可以表征菌种甚至区分菌株。以 057 藤黄微球菌 (*Micrococcus luteus*) 为例(表 6), 8 个分析菌株中序列完全一致的核糖体蛋白质有 5 种(L36、S18、L27、L23 和 S14), 它们可作为菌种鉴定的依据。其他核糖体蛋白质显示出不同程度的序列差异, 其中分子量接近的核糖体蛋白质无法用于区分菌株, 如 L35 无法区分 AS2 与其他菌株。分子量表现出一定差异的核糖体蛋白质可用于区分菌株, 如 L32、L31-B、L29 和 L24 等。质谱注释同样支持这一观点, 测试菌株 *M. luteus* DSM 1790 和同种 8 个参比菌株的核糖体蛋白质匹配度呈现显著差异, 与 FDAARGOS_677 的匹配度最高, 为 18 种核糖体蛋白质, 和其他菌株的匹配度在 7-14 种核糖体蛋白质之间。此外, 一些亲缘关系很近的菌株, 各种核糖体蛋白质均呈现出序列一致性, 无法基于核糖体蛋白质进行区分, 如 018 白色拟诺卡菌(*Nocardiosis alba*)、020 黄铜色小单孢菌(*Micromonospora chalybeata*)等, 需要通过其他指标进行区分。

表 5 注释出的各种核糖体蛋白质信号的出现次数

Table 5 The number of mass peak annotations among different ribosomal proteins

Ribosomal protein	Number of mass peak annotations	Ribosomal protein	Number of mass peak annotations	Ribosomal protein	Number of mass peak annotations
L36	127	L35	90	S13	13
L29	118	L31 and L31-B	83	L19	8
L28	115	S15	81	S6	6
L32	112	S19	80	L22	4
L27	111	S17	50	L18	4
L30	109	S12	47	L25	1
L33	108	L24	45	L14	1
L34	106	L21	25	S11	1
S18	106	L7/L12	18	S16	1
S14 and S14-Z	100	L23	14		
S20	94	S10	13		

表 6 *M. luteus* 核糖体蛋白质分子量标注及菌株 DSM 1790 质谱的核糖体蛋白质信号注释

Table 6 The list of the masses predicted for 18 ribosomal proteins of 8 *M. luteus* strains and the annotation of ribosomal protein mass peaks in the mass spectrum of *M. luteus* DSM 1790

<i>M. luteus</i> strain	L36	L34	L33	L35	L32	L30	L28	S18	L27	S20	L31-B	L29	S15	S19	L23	S14	L24	S12
AS2 (7)	4 332.24	5 104.94	6 444.41	7 315.74	7 414.53	7 943.00	8 417.78	8 625.17	8 965.03	9 142.67	9 591.71	9 717.89	10 316.99	10 544.31	11 107.68	11 424.12	12 235.97	13 590.91
SA211 (11)	4 332.24	5 177.99	6 428.41	7 315.70	7 384.50	8 106.23	8 447.87	8 625.17	8 965.03	9 142.67	9 618.78	9 729.99	10 316.99	10 544.31	11 107.68	11 424.12	12 192.00	13 548.83
R17 (12)	4 332.24	5 177.99	6 428.41	7 315.70	7 384.50	8 106.23	8 417.78	8 625.17	8 965.03	9 142.67	9 618.78	9 729.99	10 316.99	10 544.31	11 107.68	11 424.12	12 192.00	13 548.83
SB1254 (13)	4 332.24	5 177.99	6 428.41	7 315.70	7 414.53	8 096.20	8 417.78	8 625.17	8 965.03	9 170.68	9 599.74	9 776.02	10 288.89	10 528.31	11 107.68	11 424.12	12 192.00	13 548.83
NCTC 2665 (13)	4 332.24	5 177.99	6 428.41	7 315.70	7 380.51	8 096.20	8 417.78	8 625.17	8 965.03	9 142.67	9 599.74	9 776.02	10 288.89	10 528.31	11 107.68	11 424.12	12 222.03	13 548.83
NCTC 7563 (14)	4 332.24	5 177.99	6 428.41	7 315.70	7 414.53	8 106.23	8 417.78	8 625.17	8 965.03	9 140.65	9 599.74	9 745.99	10 288.89	10 528.31	11 107.68	11 424.12	12 192.00	13 548.83
NCCP 16831 (14)	4 332.24	5 177.99	6 428.41	7 315.70	7 384.50	8 106.23	8 417.78	8 625.17	8 965.03	9 168.70	9 618.78	9 729.99	10 302.96	10 544.31	11 107.68	11 424.12	12 222.03	13 548.83
FDAARGOS_677 (18)	4 332.24	5 177.99	6 428.41	7 315.70	7 384.50	8 106.23	8 417.78	8 625.17	8 965.03	9 170.68	9 604.76	9 745.99	10 288.89	10 528.31	11 107.68	11 424.12	12 222.03	13 548.83
DSM 1790	4 332.71	5 179.55	6 429.87	7 316.53	7 386.13	8 107.38	8 418.95	8 626.48	8 965.98	9 172.73	9 605.73	9 747.25	10 291.96	5 264.86	5 556.67	5 713.46	12 222.18	6 778.08

2.5 测试菌株的核糖体蛋白质鉴定

三个测试菌株 MALDI-TOF MS 指纹图谱的核糖体蛋白质信号注释结果见附表 3。菌株 D24 的质

谱中出现匹配参比菌株 *K. rhizosphaerae* DSM 19711^T 全部 15 种目标核糖体蛋白质的 15 个特征峰 (图 1A), 可据此鉴定菌株为 *K. rhizosphaerae*, 质

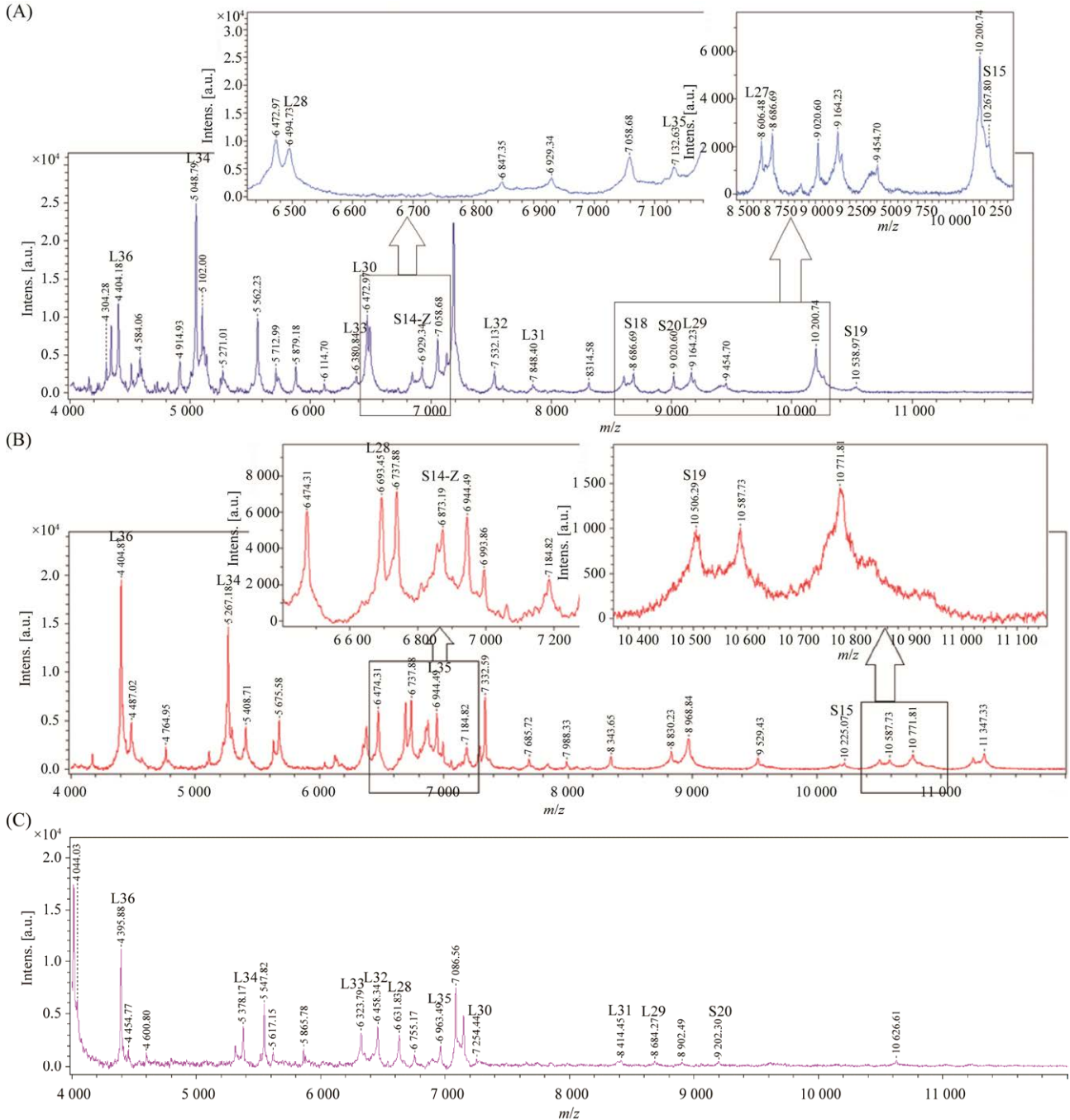


图 1 测试菌株根际动球菌 *K. rhizosphaerae* D24 (A)、江西伦茨菌 *L. jiangxiensis* 3-5 (B) 和食醚红球菌 *R. aetherivorans* 4-12 (C) 的 MALDI-TOF MS 图谱

Figure 1 MALDI-TOF mass spectra of *K. rhizosphaerae* D24 (A), *L. jiangxiensis* 3-5 (B) and *R. aetherivorans* 4-12 (C). The annotated ribosomal protein mass peaks annotated are labeled in red.

谱法与 16S rRNA 基因序列比对法的结果一致。菌株 3-5 的质谱中出现匹配参比菌株 *L. jiangxiensis* CGMCC 4.6609^T 核糖体蛋白质的 7 个特征峰(图 1B), 菌株 4-12 的质谱中出现匹配参比菌株 *R. aetherivorans* PSBB011 核糖体蛋白质的 10 个特征峰(图 1C)。参照对藤黄微球菌的分析, 同一菌种不同菌株间至少应匹配 7 个核糖体蛋白质, 因此这两个测试菌株均可能与参比菌株同种, 但由于缺失了部分特征峰的信息, 基于部分特征峰的匹配能否准确鉴定到种, 需要将质谱图与同属更多参比菌种的核糖体蛋白质进行匹配和分析得出结论。

3 讨论

本研究所用的质谱注释方法依赖于 MALDI-TOF MS 图谱数据库以及基因组数据库的信息, 菌种名称以及从基因组提取的核糖体蛋白质序列准确度都影响注释结果。本研究中共 2 个菌株[008 *Varibaculum cambriense* 110324_100 PNU 和 015 疮痂链霉菌(*Streptomyces scabiei*) DSM 40611] 均注释出 3 种核糖体蛋白质信号, 与同科或同属其他分析菌株差异显著, 注释结果可信度低, 不排除菌种名称错误的可能。此外, 质谱仪分辨率的局限性也影响注释结果。菌株蛋白质组中蛋白数量繁多, 难免出现 m/z 接近的分子离子, 以 MALDI-TOF MS 线性模式下的分辨率无法做到基线分离, 因此, 质谱中的一个 m/z 可能匹配不止一种蛋白质。一种情况是两种核糖体蛋白质 m/z 接近, 若同时存在两种蛋白质信号, 注释准确; 若只存在一种蛋白质信号, 会造成过多注释。另一种情况是一种核糖体蛋白质和一种非核糖体蛋白质 m/z 接近, 若两种蛋白质信号同时存在, 不影响注释结果; 若只有非核糖体蛋白质存在, 会导致错误注释。因此, 对于出现次数低的核糖体蛋白质信号, 如 L22、L18、S16、

S11、L25 和 L14, 存在错误注释的可能。对于出现次数高的或者多个类群中都注释出的核糖体蛋白质信号, 注释结果可信度高。

质谱注释表明放线菌纲细菌 MALDI-TOF MS 指纹图谱中的核糖体蛋白质信号一般不超过质谱峰的半数。除了核糖体蛋白质的单电荷和双电荷峰, 还有一些质谱峰未注释。推测原因如下, 第一, 基因注释的精确度有限, 一些开放读码框架(ORF)无法确定功能, 第二, 一些蛋白质的翻译后修饰未知, 无法准确预测分子量, 第三, 蛋白质在提取过程中发生降解, 蛋白质降解产物的 m/z 无法确知。核糖体蛋白质的翻译后加工、修饰相对简单, 掌握其规律可以基于序列准确预测分子量, 而测试菌株 D24 的质谱注释结果也验证了本研究所用分子量计算方法的准确性。

本研究中 3 个测试菌株与其参比菌株的核糖体蛋白质匹配数量呈现出一定差异, 主要是测试菌株与其参比菌株的亲缘关系远近差异所致。以前文中分析的藤黄微球菌为例, 参比菌株为 FDAARGOS_677 时, 测试菌株 DSM 1790 可匹配 18 种核糖体蛋白质, 参比菌株为 AS2 时, 测试菌株可匹配 7 种核糖体蛋白质, 参比菌株为其他菌株时, 匹配数为 7-18。由于 NCBI 基因组数据库中很多菌种只有一个菌株的全基因组测序信息, 无法获取更多参比菌株, 因而测试菌株匹配参比菌株的核糖体蛋白质少于 15 种的情况将很常见。这种情况下测试样本的质谱中是否产生足够数量种水平的核糖体蛋白质特征峰将至关重要, 这就需要对目标属的更多菌种进行比较和分析找出种水平的核糖体蛋白质特征峰, 结合核糖体蛋白质的匹配数和种水平的特征峰得出鉴定结果。

MALDI Biotyper 数据库收录的放线菌纲菌种以致病菌为主, 对一些环境中常见的类群, 尤其是具有抗生素等活性物质合成能力的链霉菌

目、小单孢菌目、假诺卡氏菌目、链孢囊菌目等的属、种收集的指纹图谱较少，限制了MALDI-TOF MS方法在环境放线菌鉴定中的应用。迄今已提交全基因组序列的放线菌纲菌种多样、来源广泛，开发不依赖图谱数据库的、识别核糖体蛋白质质谱特征峰的鉴定方法可扩展MALDI-TOF MS鉴定方法的应用范围。相比自建库标准方法，该方法构建数据库只要获得参比菌株的基因组信息即可，无需收集大量菌株和采集质谱图，具有简单快速、成本低廉、可公开等优势，易于推广普及，因而具有广阔的应用前景。本研究找到15种核糖体蛋白质标志物并对3个菌株进行了初步测试分析，可为通过识别核糖体蛋白质的质谱特征峰鉴定放线菌的方法建立提供依据。

4 结论

本研究利用基因组学技术注释放线菌纲细菌MALDI-TOF MS指纹图谱中的核糖体蛋白质信号，找到15种核糖体蛋白质标志物，可为通过识别核糖体蛋白质的质谱特征峰鉴定放线菌的方法建立提供依据。

参考文献

- [1] Patel R. MALDI-TOF MS for the diagnosis of infectious diseases. *Clinical Chemistry*, 2015, 61(1): 100–111.
- [2] Singhal N, Kumar M, Kanaujia PK, Viridi JS. MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. *Frontiers in Microbiology*, 2015, 6: 791.
- [3] 戴颖欣, 李敏. MALDI-TOF MS在临床微生物检验中的应用. *检验医学*, 2015, 30(2): 102–107.
Dai YX, Li M. Application of MALDI-TOF MS technology in clinical microbial determination. *Laboratory Medicine*, 2015, 30(2): 102–107. (in Chinese)
- [4] 李楠, 刘文森, 刘林娜, 郝镗, 李树民, 孟轲音, 何扬, 万忠海, 王承宇, 李吉平. MALDI-TOF质谱技术及Biotyper数据库在养殖场空气微生物鉴定中的应用. *中国兽医学报*, 2014, 34(12): 2031–2034.
Li N, Liu WS, Liu LN, Hao Z, Li SM, Meng KY, He Y, Wan ZH, Wang CY, Li JP. The application of the MALDI-TOF MS and biotyper database in the identification of farm airborne microorganisms. *Chinese Journal of Veterinary Science*, 2014, 34(12): 2031–2034. (in Chinese)
- [5] Elbehiry A, Marzouk E, Hamada M, Al-Dubaib M, Alyamani E, Moussa IM, Al Rowaidhan A, Hemeg HA. Application of MALDI-TOF MS fingerprinting as a quick tool for identification and clustering of foodborne pathogens isolated from food products. *New Microbiologica*, 2017, 40(4): 269–278.
- [6] 顾春华. MALDI-TOF MS技术及其在食品微生物检测方面的应用. *中国酿造*, 2019, 38(9): 24–27.
Gu CH. MALDI-TOF MS technique and its application challenges in food microbiology detection. *China Brewing*, 2019, 38(9): 24–27. (in Chinese)
- [7] Arnold RJ, Reilly JP. Observation of *Escherichia coli* ribosomal proteins and their posttranslational modifications by mass spectrometry. *Analytical Biochemistry*, 1999, 269(1): 105–112.
- [8] Ryzhov V, Fenselau C. Characterization of the protein subset desorbed by MALDI from whole bacterial cells. *Analytical Chemistry*, 2001, 73(4): 746–750.
- [9] 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. *临床检验杂志*, 2016, 34(6): 447–450.
- [10] 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, 杨文航. 自建MALDI-TOF MS微生物鉴定数据库专家共识. *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, Yang WH. *Expert consensus on laboratory developed MALDI-TOF MS microbial identification database*. *Chinese Journal of Laboratory Medicine*, 2019, 42(6): 414–419. (in Chinese)
- [11] Pineda FJ, Antoine MD, Demirev PA, Feldman AB, Jackman J, Longenecker M, Lin JS. Microorganism identification by matrix-assisted laser/desorption ionization mass spectrometry and model-derived ribosomal protein biomarkers. *Analytical Chemistry*, 2003, 75(15): 3817–3822.
- [12] Ban N, Beckmann R, Cate JH, Dinman JD, Dragon F, Ellis SR, Lafontaine DL, Lindahl L, Liljas A, Lipton

- JM, McAlear MA, Moore PB, Noller HF, Ortega J, Panse VG, Ramakrishnan V, Spahn CM, Steitz TA, Tchorzewski M, Tollervey D, Warren AJ, Williamson JR, Wilson D, Yonath A, Yusupov M. A new system for naming ribosomal proteins. *Current Opinion in Structural Biology*, 2014, 24: 165–169.
- [13] 江文静, 张军毅, 杜阳, 孙丽伟. 以核糖体蛋白质鉴别铜绿微囊藻的应用分析. *生物技术通报*, 2017, 33(10): 117–124.
Jiang WJ, Zhang JY, Du Y, Sun LW. Application analysis of *Microcystis aeruginosa* identification based on ribosomal proteins. *Biotechnology Bulletin*, 2017, 33(10): 117–124. (in Chinese)
- [14] Cheng D, Qiao L, Horvatovich P. Toward spectral library-free matrix-assisted laser desorption/ionization time-of-flight mass spectrometry bacterial identification. *Journal of Proteome Research*, 2018, 17(6): 2124–2130.
- [15] Kostas J, Parker KC. Using matrix-assisted laser desorption/ionization time of flight spectra to elucidate species boundaries by matching to translated DNA databases. *Journal of the American Society for Mass Spectrometry*, 2020, 31(1): 73–84.
- [16] de Simeis D, Serra S. Actinomycetes: a never-ending source of bioactive compounds—an overview on antibiotics production. *Antibiotics*, 2021, 10(5): 483.
- [17] Tatusova T, Di Cuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*, 2016, 44(14): 6614–6624.
- [18] 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. *Proceedings of the National Academy of Sciences of the United States of America*, 1989, 86(21): 8247–8251.
- [19] Gonzales T, Robert-Baudouy J. Bacterial aminopeptidases: properties and functions. *FEMS Microbiology Reviews*, 1996, 18(4): 319–344.
- [20] Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExpASY server. *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press, 2005: 571–607.
- [21] Yutin N, Puigbò P, Koonin EV, Wolf YI. Phylogenomics of prokaryotic ribosomal proteins. *PLoS One*, 2012, 7(5): e36972.

(本文责编 张晓丽)

补充材料

附表 1 142 个放线菌纲菌株 MALDI-TOF MS 指纹图谱的核糖体蛋白质信号注释结果

附表 2 142 个放线菌纲菌株 MALDI-TOF MS 指纹图谱中注释出的核糖体蛋白质信号数量统计

附表 3 3 个测试菌株 MALDI-TOF MS 指纹图谱的核糖体蛋白质信号注释结果

本文补充材料见网络版 <http://journals.im.ac.cn/actamicrocn>。补充材料为作者提供的原始数据, 作者对其学术质量和内容负责。