

益生菌基因组学在乳酸菌筛选和功能评价中的应用

刘文俊, 张和平*

(内蒙古农业大学 乳品生物技术与工程教育部重点实验室 农业农村部奶制品加工重点实验室
内蒙古自治区乳品生物技术与工程重点实验室 乳酸菌与发酵乳制品省部共建协同创新中心 呼和浩特 010018)

摘要 乳杆菌和双歧杆菌是益生菌筛选的主要微生物类群,在促进人类、动物甚至植物健康方面都有重要的应用价值,被广泛应用于多种功能性食品、动物饲料和微生物菌肥。然而,这些细菌对宿主健康产生积极影响的分子机制还远未完全了解。另外,传统的益生菌筛选技术主要依靠表型和生理生化特性的测定以及随机对照试验,这些方法不仅耗时、费力,而且数据难以重复。不同研究小组之间数据难以共享,不能实现菌株有益特性的均一性和标准化验证,严重制约了益生菌科学研究的发展和菌株的产业化开发利用。基于新的测序技术和强大的计算机分析方法,被称为益生菌基因组学新学科的出现,允许研究者在短时间内通过筛选大量的生物学和基因组数据,显著提高益生菌筛选的效率和准确性。本文综述了通过基因组数据的分析,帮助人们高效、快速地确定益生菌,开辟益生菌新功能的评价途径。总结益生菌基因组学在阐明益生特性功效评价的分子基础等方面的最新研究成果,以期扩展研究人员对这些有益微生物生物学的认知,揭示益生菌促进健康特性的分子机制提供新视角。

关键词 益生菌基因组学; 乳酸菌; 功能; 筛选; 人工智能

文章编号 1009-7848(2024)01-0001-11 **DOI:** 10.16429/j.1009-7848.2024.01.001

益生菌被广泛应用于乳制品、发酵食品、饮料、膳食补充剂、活菌药物、化妆品、动物生态制剂和微生物菌肥等不同的产品中,已经发展为产值超过千亿元人民币的朝阳产业^[1]。近年来,随着分子生物学的发展,益生菌促进宿主健康特性背后的分子机制也逐渐被揭示出来,这些机理、机制的发现大大增加了人们对益生菌的认识,为益生菌的应用拓宽了范围。基因组学和功能基因组学以及高通量测序技术的出现,使人们对益生菌的产业化用途和代谢特性有了广泛的了解,这有助于发现新的值得探索的益生菌应用领域和潜力。2009年,一门名为益生菌基因组学(Probiogenomics)的新学科应运而生^[2]。益生菌基因组学旨在深入了解益生菌的多样性,揭示其促进宿主健康特性的分子机制^[3]。全基因组测序、基因组数据挖掘和比较基因组学为了解微生物遗传内涵、差异和相似性提供了新的见解,并为可能的基因功能挖掘提供重要线索。

近年来,乳酸菌基因组信息呈爆发式增长态势,在公共数据平台发布的乳酸菌全基因组数量已经达到 8 万余个(图 1),这些基因组信息对研究人员认识乳酸菌的生理特征、有益功能、工业化加工特性的分子机制具有重要作用。益生菌基因组学结合体内、外随机对照试验,对发现新的益生菌菌株和表征益生乳酸菌的新功能发挥着越来越大的作用。益生菌基因组学和计算机深度机器学习模型相结合的分析技术已应用于益生菌可视化筛选、预测^[4],益生乳酸菌安全相关特征的表征(比如:毒力因子、溶血素基因、抗菌素耐药基因和可移动原件)等方面取得了重要的研究成果^[5]。鉴于此,有必要及时总结乳酸菌基因组学研究的最新进展,并基于基因组分析技术了解乳酸菌的产业化相关特征。本文从基因组学的角度论述益生乳酸菌(特别是工业化应用乳酸菌)相关的益生特性、生理生化等表型差异形成的分子机制。综述乳酸菌基因组学和依托基因组大数据的深度机器学习模型,以及人工智能技术在优良乳酸菌的筛选和功能评价中的应用研究进展,以期乳酸菌筛选,益生菌基因组和健康食品领域的研究提供新的见解。

收稿日期: 2024-01-15

基金项目: 国家自然科学基金项目(U22A20540,32272289);
内蒙古自然科学基金杰青项目(2023JQ07)

第一作者: 刘文俊,男,博士,研究员

通信作者: 张和平 E-mail: hepingdd@vip.sina.com

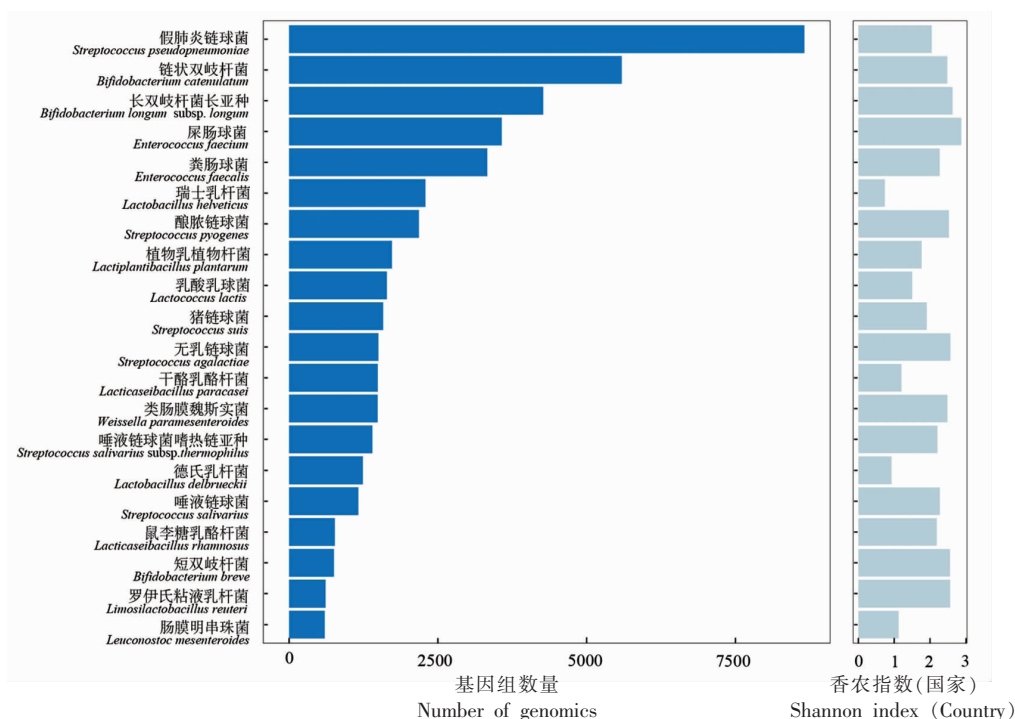


图1 乳酸菌基因组数量统计

Fig.1 Quantitative statistics of lactic acid bacteria genome

1 乳酸菌基因组学研究进展

1.1 乳酸菌的种属

乳酸菌是指发酵可利用的碳水化合物形成以乳酸为主产物的一类细菌的总称,它不是一个分类学上的名称,包含大量的种属,来源广泛,存在于各种栖息环境,包括乳制品、肉制品、蔬菜、谷物和植物环境以及人和动物肠道^[6]。乳酸菌种属内异质性非常丰富,在过去,乳酸菌常见属主要有:乳杆菌属(*Lactobacillus*)、乳球菌属(*Lactococcus*)、双歧杆菌属(*Bifidobacterium*)、明串珠菌属(*Leuconostoc*)、片球菌属(*Pediococcus*)、酒球菌属(*Oenococcus*)、魏斯氏菌属(*Weissella*)、链球菌属(*Streptococcus*)、肠球菌属(*Enterococcus*)和肉食菌属(*Caroobacterium*)等^[7]。2020年,依据基因组核心基因系统发育关系、(保守的)成对平均氨基酸一致性、分支特异性特征基因、生理生化指标和菌株的栖息地生态学等特征等将乳杆菌属重新划分为25个属^[8],包括修订后的乳杆菌属、*Paralactobacillus*和23个新属。目前,原乳杆菌属已包含332个种和亚种,如表1所示。在过去,乳酸菌主要应用于乳制品工业和其它食品的发酵。现在,乳酸菌被广泛应

用于各种饮料、酸奶或其它以特定健康益处为卖点的产品中。近年来,随着对益生菌促进人体健康作用研究的深入,乳酸菌成为食品工业创新发展的动力源泉,是食品工业研究的热点^[9]。

1.2 乳酸菌群体基因组研究

基因组的概念最早是由德国汉堡大学植物学家 Winkler 于上世纪 20 年代提出,随着科学家对生物体基因层面研究的深入,基因组研究范围逐渐扩大,现在已经发展为一门综合性学科^[10]。基因组学是以某个生物体的全部基因作为研究对象,以分子生物学为基础,通过全基因组序列分析,利用计算机、信息技术以及相关生物学技术等手段,研究生物系统内部结构、功能及进化的科学,包括结构基因组学、功能基因组学和比较基因组学 3 个分支^[11]。第 1 个细菌基因组流感嗜血杆菌 (*Haemophilus influenzae*) 的序列发表于 1995 年,开启了微生物的基因组学研究时代^[12]。乳酸乳球菌 IL1403 基因组于 2001 年发布,是首次报道的乳酸菌基因组序列^[13]。2008 年笔者及研究团队完成了国内第 1 株益生乳酸菌 *L. casei* Zhang 的全基因组测序工作,开启了国内益生乳酸菌基因

表 1 乳酸菌修订后的属包含种的数量及其模式种

序号	属名	种的 数量/个	模式种
1	乳杆菌属 (<i>Lactobacillus</i>)	46	德氏乳杆菌 (<i>Lactobacillus delbrueckii</i>)
2	霍尔托普菲尔氏菌属 (<i>Holzschufelia</i>)	1	栖花霍尔托普菲尔氏菌 (<i>Holzschufelia floricola</i>)
3	淀粉乳杆菌属 (<i>Amylolytobacillus</i>)	2	嗜淀粉淀粉乳杆菌 (<i>Amylolytobacillus amylophilus</i>)
4	蜂巢乳杆菌属 (<i>Bombilactobacillus</i>)	6	意蜜蜂乳杆菌 (<i>Bombilactobacillus mellifer</i>)
5	伴生乳杆菌属 (<i>Companilactobacillus</i>)	34	消化伴生乳杆菌 (<i>Companilactobacillus dimentarius</i>) 对应于以前消化乳杆菌 (<i>L. dimentarius</i>) 类群
6	石墙乳杆菌属 (<i>Lapidilactobacillus</i>)	8	曲面石墙乳杆菌 (<i>Lapidilactobacillus concavus</i>) 对应于以前 <i>L. concavus/dextrinicus</i> 类群
7	农田乳杆菌属 (<i>Agrilactobacillus</i>)	3	堆肥农田乳杆菌 (<i>Agrilactobacillus composti</i>)
8	施莱弗乳杆菌属 (<i>Schleiferilactobacillus</i>)	3	恶味施莱弗乳杆菌 (<i>Schleiferilactobacillus perolens</i>) 对应于以前的 <i>L. perolens</i> 类群
9	腐败乳杆菌属 (<i>Loigolactobacillus</i>)	9	棒状腐败乳杆菌 (<i>Loigolactobacillus coryniformis</i>) 对应于以前的棒状乳杆菌 (<i>L. coryniformis</i>) 类群
10	乳酪杆菌属 (<i>Lacticaseibacillus</i>)	30	干酪乳酪杆菌 (<i>Lacticaseibacillus casei</i>) 对应于以前的干酪乳杆菌 (<i>L. casei</i>) 类群
11	广泛乳杆菌属 (<i>Latilactobacillus</i>)	6	清酒广泛乳杆菌 (<i>Latilactobacillus sakei</i>) 对应于以前的清酒乳杆菌 (<i>L. sakei</i>) 类群
12	德拉格里奥氏菌属 (<i>Dellaglioia</i>)	2	低温德拉格里奥氏菌 (<i>Dellaglioia algida</i>)
13	液体乳杆菌属 (<i>Liquorilactobacillus</i>)	13	苹果液体乳杆菌 (<i>Liquorilactobacillus mali</i>) 对应于以前的唾液乳杆菌 (<i>L. salivarius</i>) 类群
14	联合乳杆菌属 (<i>Ligilactobacillus</i>)	19	唾液联合乳杆菌 (<i>Ligilactobacillus salivarius</i>) 对应于以前的唾液乳杆菌 (<i>L. salivarius</i>) 类群
15	乳植杆菌属 (<i>Lactiplantibacillus</i>)	18	植物乳植杆菌 (<i>Lactiplantibacillus plantarum</i>) 对应于以前的植物乳杆菌 (<i>L. plantarum</i>) 类群
16	糠乳杆菌属 (<i>Furfurilactobacillus</i>)	4	罗西糠乳杆菌 (<i>Furfurilactobacillus rossiae</i>) 对应于以前的 <i>L. rossiae</i> 类群
17	寡食乳杆菌属 (<i>Paucilactobacillus</i>)	7	牛粪寡食乳杆菌 (<i>Paucilactobacillus vaccinostereus</i>) 对应于以前的牛粪乳杆菌 (<i>L. vaccinostereus</i>) 类群
18	粘液乳杆菌属 (<i>Limosilactobacillus</i>)	33	发酵粘液乳杆菌 (<i>Limosilactobacillus fermentum</i>) 对应于以前的发酵乳杆菌 (<i>L. fermentum</i>)
19	果实乳杆菌属 (<i>Fructilactobacillus</i>)	10	食果糖果实乳杆菌 (<i>Fructilactobacillus fructivorans</i>) 对应于以前的 <i>L. fructivorans</i> 类群
20	醋乳杆菌属 (<i>Acetilactobacillus</i>)	1	金山醋乳杆菌 (<i>Acetilactobacillus jinshanensis</i>)
21	蜜蜂乳杆菌属 (<i>Apilactobacillus</i>)	11	昆氏蜜蜂乳杆菌 (<i>Apilactobacillus kunkeei</i>) 对应于以前的 <i>L. kunkeei</i> 类群
22	促生乳杆菌属 (<i>Levilactobacillus</i>)	29	安格连促生乳杆菌 (<i>Levilactobacillus angrenensis</i>) 对应于以前 <i>Lactobacillus angrenensis</i>
23	次乳杆菌属 (<i>Secundilactobacillus</i>)	15	坏发酵次乳杆菌 (<i>Secundilactobacillus malefermentans</i>) 对应于以前的丘状菌落 (<i>Lactobacillus collimoides</i>) 类群
24	迟缓乳杆菌属 (<i>Lenilactobacillus</i>)	21	布氏迟缓乳杆菌 (<i>Lenilactobacillus buchneri</i>) 对应于以前布氏乳杆菌 (<i>L. buchneri</i>) 类群
25	副乳杆菌属 (<i>Paralactobacillus</i>)	1	雪兰莪州副乳杆菌 (<i>Paralactobacillus selangorensis</i>) 对应于以前的 <i>L. selangorensis</i>
	合计	332	

组、蛋白质组学研究的序幕^[14]。2015年,在国际上首次完成了乳杆菌属、双歧杆菌属等8个属289个模式菌株基因组的解析,揭示乳酸菌不同物种分化伴随的功能基因进化历程,发现了乳糖代谢和蛋白水解等代谢特征与栖息环境适应性密切相关的规律,开创了乳酸菌群体基因组研究的先河^[15]。自2018年,笔者研究团队启动“乳酸菌万株基因组计划”以来,已经完成了德氏乳杆菌(*Lactobacillus delbrueckii*)^[16]、植物乳植杆菌(*Lactiplantibacillus plantarum*)^[17]和嗜热链球菌嗜热亚种(*Streptococcus thermophilus*)^[18]等17 897株乳酸菌基因组的解析,阐明了乳酸菌适应生态位发生还原性衰退的群体进化规律。基于此,建成了全球最大的乳酸菌基因组数据库(<https://www.imhpc.com/iLAB-db>),收录81 638个基因组序列,旨在为乳酸菌基因组的深入分析、可视化和共享提供一个可访问的平台,为乳酸菌的遗传进化、功能基因解析和代谢通路挖掘提供帮助^[19]。通过基因组学研究对乳酸菌进化、分化历程有了更深入的了解。早期的微生物基因组分析表明:基因组的研究内容通常反映生物体的代谢、生理、生物合成能力以及对不同条件和环境的适应性。乳酸菌基因组学的研究彻底改变了人们对其代谢过程、生物加工能力的认识,以及了解其在人类身体健康和心理健康方面的潜在作用^[20]。

1.2.1 乳酸菌比较基因组学 比较基因组学是通过比较不同物种或个体的基因组来揭示它们之间相似性和差异性的学科^[21]。在乳酸菌基因组分析中,比较基因组学发挥着关键作用,为深入理解乳酸菌的生物学特性、进化、分化历程以及生态适应性提供了丰富的信息。通过比较不同乳酸菌菌株的基因组,揭示它们之间的遗传多样性,包括单核苷酸多态性(SNPs)、插入缺失(Indels)等变异类型,为研究群体内部和群体间的遗传结构、演化动态提供基础^[22]。Sun等^[19]通过对乳杆菌属213株模式株的基因组精细图谱进行测序,发现乳杆菌属基因组多样性丰富,接近于“科”的水平。比较基因组学还能够识别和注释基因组中的功能元件,如基因、蛋白质、RNA等。通过比较乳酸菌的基因组,可以识别与代谢途径、生态适应性等相关的基因,有助于了解乳酸菌在不同环境中的适应性和

代谢特征。此外,比较相关物种或不同菌株的基因组,推断它们的进化关系和演化历史,为了解乳酸菌的起源、分化以及与其它微生物的关系提供分子生物学信息。群体基因组研究发现德氏乳杆菌保加利亚亚种(*Lactobacillus delbrueckii* subsp. *bulgaricus*)群体可划分为4个分支,且具有较高的遗传多样性和重组异质性^[16]。Zhao等^[18]对嗜热链球菌嗜热亚种(*Streptococcus thermophilus*)进行全基因组测序,发现其遗传进化具有地域和分离源相关性。通过比较基因组分析,还可以鉴定出具有特定性状或功能的关键基因,对选择优良菌株、代谢工程或改良乳酸菌的性状具有指导意义,如:改进乳制品发酵过程,增强菌株对特定环境的适应性等。总之,比较基因组学在乳酸菌研究中的应用,提供了全面的遗传信息,推动了研究人员对乳酸菌生物学特性和应用潜力的深入了解。

1.2.2 乳酸菌功能基因组学 功能基因组学,又称后基因组学,是指以全基因组测序技术得到的基因组序列为基础,利用公共数据库进行功能基因注释,通过解析基因图谱来预测基因及其产物的功能,解析不同生态位菌种的代谢机制,将功能基因与菌株表型特性相关联^[23]。近年来,随着乳酸菌资源收集保护数量的增加和基因组数据的积累,功能基因的深度挖掘也逐渐展开。Suzuki等^[24]通过对植物乳植杆菌(*Lactiplantibacillus plantarum* LOC1)的功能基因进行分析,发现该菌株具有与磷壁酸和脂蛋白生物合成相关的黏附因子基因,这部分基因参与菌株的免疫调节。功能基因组学不仅可以帮助人们了解不同生态位乳酸菌对其生长环境的反应,对于了解乳酸菌在自然发酵和工业化发酵中的适应性以及它们与人类宿主的相互作用具有重要意义。此外,还可以通过功能基因组学分析,实现潜在的益生乳酸菌的挖掘。2022年,Sun等^[25]对44株植物乳植杆菌(*Lactiplantibacillus plantarum*)的功能基因组进行分析,发现菌株的抗氧化活性、黏附活性和在模拟胃肠道环境中的生存能力分别与抗氧化酶编码基因、细胞表面蛋白编码基因和应激反应基因密切相关。菌株中存在的功能基因的数量可能决定它们在益生菌评估中的表现。

1.3 基因组学技术在益生乳酸菌研究中的应用

我国益生菌的定义是:活的微生物,当摄入充足量时,对宿主产生健康益处^[26]。益生菌要具体到“株”的水平,每个益生菌菌株都应具有相应的菌株号。特指食品范畴的有益微生物,1个菌株,不管其来源如何,只有进行分类学上的鉴定、安全性评价及功能试验后,符合益生菌的概念,才能称为益生菌^[27]。基于全基因组测序技术的平均核苷酸一致性、单核苷酸多态性、核心基因或全基因组多位点序列分型,正成为乳酸菌的精准鉴定的主要标准。菌株分类学精准鉴定是益生菌评价中最关键的环节。大多数益生菌是由乳酸菌筛选评价而来,在我国卫生健康委员会公布的《可用于食品中添加的菌株名单》中乳杆菌和双歧杆菌是最主要的菌株。益生菌的筛选和确定的10个标准包括:明确的益生菌定义、活菌、对菌株有充分的了解、菌株不含抗生素耐药基因、能在胃肠道存活并定殖、菌株与肠道菌群产生有益互作、菌株安全可靠^[28]。益生乳酸菌对人类或动物的健康有有益的影响,近来也被证明具有预防或治疗某些疾病的能力。

1.3.1 基因组学在益生乳酸菌安全性评价中的应用

乳酸菌应用历史悠久、安全,被认为是一般安全的(Generally regarded as safe,GRAS)物种^[29]。几十年来,乳酸菌在传统食品和发酵产品,比如:酸奶、泡菜、香肠和饮料等中得到广泛的应用。然而,随着益生菌乳酸菌应用范围的扩大,安全性问题日益受到关注,可能出现严重感染、耐药基因转移等不良反应。因此,工业化应用前,对所选择的菌株和发酵食品的安全性评价仍然是乳酸菌筛选和研究的一个关键任务。由于安全性是首要考虑的问题,因此菌株特性信息应集中于鉴定产品中所有细菌菌株的潜在不良特性上。除其遗传背景外,菌株特征信息应包括有关毒力因子或毒素和抗菌素抗性基因存在的详细信息^[30]。目前,体外检测仍是评估乳酸菌特性的主要方法。然而,在食品安全概念不断发展的背景下,这些方法不能满足高通量快速的菌株风险评估的需要。分析完整的遗传信息,包括潜在的毒力基因和对健康有负面影响等其它决定因素,可以评估菌株在产品中的使用安全性,为此可以采用乳酸菌单菌株的全基因组测序(Whole genomics sequencing,WGS)进行

研究。全基因组生物信息学分析是一种理想的,具有低成本、高效益的菌株特性评价方法,用于微生物安全风险评估。同时,该技术也可以增加人们对食品安全和益生乳酸菌特性评估的了解。全基因组测序方法能够提供菌株基因组的全部信息,不仅针对毒力和耐药性潜力,而且还表明菌株特异性特征,包括用于专利保护的代谢和系统基因组标记等^[31-32]。基因组测序技术也可以对质粒、噬菌体、整合子和转座子介导的水平基因转移潜力进行额外的评估和CRISPR分析,为基因组稳定性提供初步证据。研究人员应将注意力转向解决编码具有安全风险的遗传因素转移到其他生物体而造成潜在安全风险的问题。因此,在益生菌筛选中应采用具有低成本、高收益的全基因组测序技术作为常规步骤,并与表型研究联合应用,作为益生菌乳酸菌安全性评估的先进的评估准则^[33]。

1.3.2 基因组学在益生乳酸菌代谢产物研究中的应用

微生物细胞内、外产生的初级代谢物、信号分子激素和次级代谢物是其生理生化特征表型的直接证据,也能反映菌株对宿主代谢的影响。微生物代谢组研究不仅能阐明各种代谢途径的网络,揭示其调控和工程改造靶点,而且可以阐明微生物与宿主之间的互作机制,因此越来越受到人们的关注^[34]。作为食品发酵剂或益生菌被广泛应用的乳酸菌,有机酸、小分子肽以及其它一系列次级代谢产物的研究是优良菌株筛选和体现益生特性的重要切入点^[35]。基因组学和功能基因组学以及代谢组学的深入研究,使得人们对乳酸菌的工业用途和代谢特性有了广泛的了解,使得它们成为未来生物精炼中最有希望作为微生物细胞工厂的候选者^[36]。作为生物精炼细胞工厂,乳酸菌是用于生产大宗和精细化学品(包括乳酸、多元醇、维生素和食品添加剂)最有前途的微生物,因为它们对环境胁迫下具有稳定性和多用途的代谢特性^[37]。代谢工程策略设计乳酸菌细胞工厂主要集中在丙酮酸代谢的重构,以产生大量重要的终产物,如有机酸、甜味剂、芳香化合物,以及决定维生素和胞外多糖生产的复杂生物合成途径^[32,38]。植物乳植杆菌(*Lactiplantibacillus plantarum*)WCFS1的基因组分析表明,该菌株具有2个预测山梨糖醇-6-磷酸脱氢酶基因(*srlD1*和*srlD2*)。在缺乏L-和D-乳酸

脱氢酶活性的突变菌株中,由于 *srlD* 基因的过量表达,显著增加了果糖-6-磷酸对山梨醇的糖酵解通量,实现了山梨醇的高产^[39]。Capozzi 等^[40]利用有毒核黄素类似物——玫瑰黄素,从硬粒小麦粉样品中分离出 2 株过量生产核黄素的植物乳杆菌。将这 2 株菌用于面包和面食的制备,面包和面食中的核黄素含量分别提高了 3 倍和 2 倍。此外,参与乳酸菌生物合成核黄素的基因已经从几个乳酸菌中鉴定出来,比较基因组分析表明,乳酸菌生物合成核黄素的操纵子具有物种和/或菌株特异性^[41]。这些结果表明,基因组学和比较基因组学结合代谢工程方法可以作为发现和提高乳酸菌工业化应用潜力的有效工具。

2 基于基因组大数据和深度机器学习的益生菌筛选技术

发现和评价乳酸菌的益生特性是一个复杂的过程,菌株从基础研究走向应用需要经历漫长的周期。传统筛选方法依赖于多次反复试验,需要大量的人力、物力和财力。益生菌基因组学分析方法在乳酸菌筛选和安全相关的遗传特征(比如:毒力因子、溶血基因、抗生素耐药基因和移动原件)的分析和预测方面具有很大的优势^[42]。基因组大数据和人工智能技术的出现,不仅体现了它在噪声

数据集中挖掘微生物功能基因特征的先天优势,也彰显出它在乳酸菌益生机制解析方面的良好应用前景。

2.1 益生菌菌株的筛选和确定方法

乳酸菌来源广泛,菌株之间特性差异大,从不同的栖息地和分离原获得的乳酸菌中筛选出具有优良特性的菌株,是其开发利用的第 1 步。益生乳酸菌从分离、鉴定到产业化应用要进行大量的基础性工作(图 2)。基于体内和体外试验研究的经典方法耗时且成本高,缺乏高通量和比较能力。目前,下一代测序(Next generation sequencing, NGS)技术因具有节省劳动力和时间、高通量能力、高丰度以及可用于广泛研究、记录遗传信息等优点而越来越受欢迎。随着测序成本的不断降低和大量的基因组数据公布,益生菌基因组学与基于随机对照试验进行的表型评价相结合,越来越广泛地应用于了解益生菌乳酸菌的特性,有助于发现未来值得探索的乳酸菌新的潜在应用领域^[30]。在单菌株全基因组测序技术的支持下,迄今为止,许多益生乳酸菌菌株已被测序并公开,并且使用基因组方法定义它们的基本特征^[31,43-44]。因此,与生物信息学分析相结合的益生菌基因组学可作为研究益生乳酸菌有益特性及其食品安全风险评估的一种新颖、经济的方法。

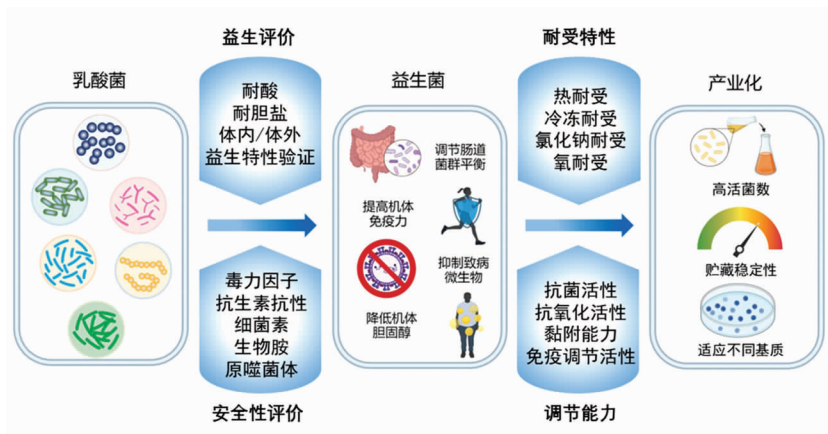


图 2 益生乳酸菌筛选及功能评价流程

Fig.2 Screening and functional evaluation process of probiotic lactic acid bacteria

2.2 人工智能在益生菌筛选中的应用进展

人工智能筛选技术是指利用人工智能相关的技术和方法,通过对大规模数据的处理、分析和学习,实现对特定目标、性质或特性的物质、生物体

或信息的筛选。国际上,人工智能应用最成功的案例当属谷歌的 AlphaFold 系统。这个系统利用卷积神经网络算法精确预测了蛋白质的三维结构,甚至可以分析 X-射线晶体学很难解决的楔入细胞

膜中的蛋白质结构^[45]。Sabater 等^[46]结合现有文献和公共数据库资源,提出了利用蔬菜废弃物和副产品靶向调控肠道微生物代谢底物的方法,建立了基于人工智能筛选蔬菜来源益生元的思维框架。目前人工智能在乳酸菌研究领域的应用取得了一些重要进展,尤其是在筛选和鉴定乳酸菌方面,包括乳酸菌的基因组数据分析、药用价值预测、代谢产物分析及微生物数据分析等方面(图3)。与此类似,McCoubrey 等^[47]利用人工智能建立了精准益生菌制剂开发模型,这个模型可以有效预测浓度如何影响益生副干酪乳杆菌在肠道中的增殖情况。我国西湖大学的研究者利用人工智能

在队列横断面分析中识别出 II 型糖尿病的肠道微生物组特征,成功构建了一个具有识别特征的微生物组风险评估系统,可被用作糖尿病的诊断、治疗和预防^[48]。中国矿业大学的研究人员通过人工智能研究健康人肠道菌群、代谢组和身体指数之间的关系,利用筛选获得的差异特征可以有效预测成人的身体指数水平^[49]。华大研究院的学者首次通过大样品基因组测序探索了宿主基因组对口腔微生物的影响,发现 5 个影响口腔微生物组的显著性位点,并应用人工智能探索基因组和微生物对牙周疾病的预测性能,取得了较好的结果。

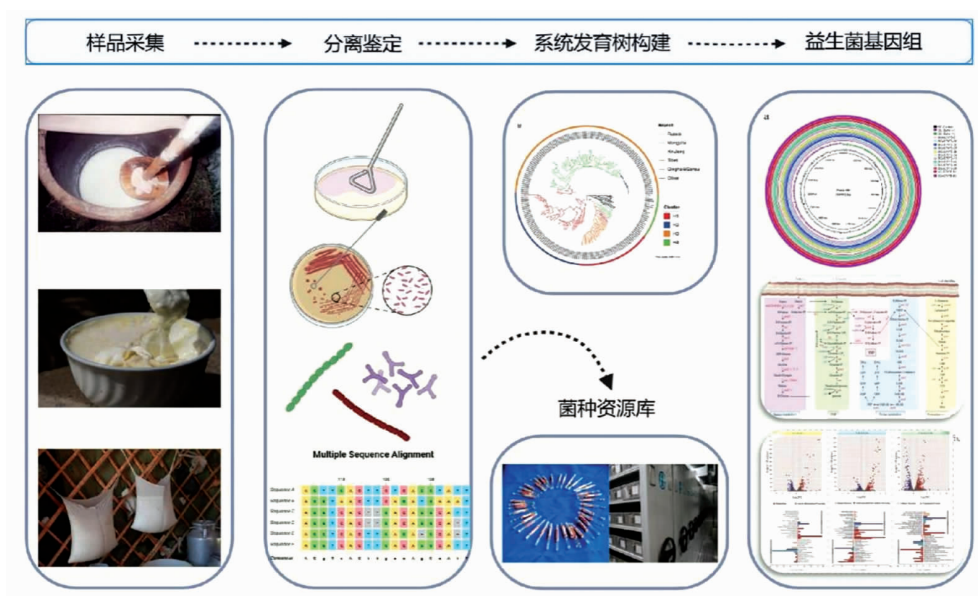


图 3 基于基因组学技术乳酸菌的筛选和确定

Fig.3 Screening and identification of lactic acid bacteria based on genomics technology

笔者及研究团队以大量的乳酸菌基因组数据为基础,应用功能基因组学和深度机器学习方法,通过解析功能基因组特征来筛选潜在的益生乳酸菌菌株,开发了基于机器学习模型的益生菌快速筛选平台 iProbiotics (<http://bioinfor.imau.edu.cn/iprobiotics>)。为益生菌的筛选和益生机制的研究提供了新的导向^[4]。运用该平台用户可以根据自己的数据或 iProbiotics 的数据进行搜索,与全部益生菌基因组序列数据进行匹配,并进行可视化结果反馈和模型预测。虽然乳酸菌功能基因组学的研究仍处于起步阶段,但是其广阔的应用前景和强劲

的发展态势,已经成为益生乳酸菌筛选和机理、机制研究等方面极具应用价值的手段。

3 结语

乳酸菌在促进人类健康方面的有益作用正陆续被证实,其研究热度和应用领域正逐步扩大。理想的益生乳酸菌通常能够抵御产品加工、流通环节的胁迫环境,菌株基因组中丰富的环境应激、表面蛋白和黏附基因可以帮助它们在胃肠道转运中存活以保持高的活菌数。目前报道的几乎所有益生乳杆菌和双歧杆菌的初步筛选和特性评价都是

基于应用技术稳定性的,比如加工耐受性和贮藏稳定性;或者基于容易测定的一些表型指标,比如:胆盐耐受性和人工胃肠液存活能力等。然而,益生菌产品的市场监管要求已经转向明确益生菌有益人类健康的分子机制。因此,过去关于益生乳酸菌的经典研究方法很难适应当前快速发展的益生菌产业需求和益生菌产品监管的行业标准要求。基于基因组测序和功能分析的“组学”方法对菌株的益生特性、遗传背景、生物安全和加工特性等进行表征,可能是满足未来监管要求的重要途径。

参 考 文 献

- [1] Grand View Research. Market analysis report, probiotics market size, share & trends analysis report by product (food & beverages, dietary supplements), by ingredient (bacteria, yeast), by distribution channel, by end-use, by Region, and segment forecasts, 2023–2030 [R/OL]. [2024–01–05]. <https://www.grandviewresearch.com/industry-analysis/probiotics-market>.
- [2] VENTURA M, O'FLAHERTY S, CLAESSEON M J, et al. Genome-scale analyses of health-promoting bacteria: Probiogenomics[J]. *Nature Reviews Microbiology*, 2009, 7: 61–71.
- [3] VENTURA M, TURRONI F, VAN SINDEREN D. Probiogenomics as a tool to obtain genetic insights into adaptation of probiotic bacteria to the human gut[J]. *Bioengineered*, 2012, 3: 73–79.
- [4] SUN Y, LI H, ZHENG J L, et al. iProbiotics: A machine learning platform for rapid identification of probiotic properties from whole-genome primary sequences[J]. *Briefings in Bioinformatics*, 2022, 23: bbab477.
- [5] CARVALHO R, GUÉDON E, ABURJAILE F F, et al. Probiogenomics of classic and next-generation probiotics[J]. *Frontiers in Microbiology*, 2022, 13: 982642.
- [6] WANG Y, WU J, LV M, et al. Metabolism characteristics of lactic acid bacteria and the expanding applications in food industry[J]. *Frontiers in Bioengineering and Biotechnology*, 2021, 9: 612285.
- [7] 张刚. 乳酸细菌: 基础、技术和应用[M]. 北京: 化学工业出版社, 2007: 22–74.
- [8] ZHANG G. *Lactic acid bacteria: Fundamentals, techniques and applications* [M]. Beijing: Chemical Industry Publishing, 2007: 22–74.
- [9] ZHENG J S, WITTOUCK S, SALVETTI E, et al. A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae* [J]. *International Journal of Systematic and Evolutionary Microbiology*, 2020, 70: 2782–2858.
- [10] CUEVAS-GONZÁLEZ P, GONZÁLEZ-CÓRDOVA A, VALLEJO-CORDOBA B, et al. Protective role of lactic acid bacteria and yeasts as dietary carcinogen-binding agents—a review[J]. *Critical Reviews in Food Science and Nutrition*, 2022, 62: 160–180.
- [11] 李伟, 印莉萍. 基因组学相关概念及其研究进展[J]. *生物学通报*, 2000, 35: 1–3.
- [12] LI W, YING L P. Genomics related concepts and research progress[J]. *Bulletin of Biology*, 2000, 35: 1–3.
- [13] CASTRO-LÓPEZ C, GARCÍA H S, MARTÍNEZ-ÁVILA G C G, et al. Genomics-based approaches to identify and predict the health-promoting and safety activities of promising probiotic strains—A probiogenomics review[J]. *Trends in Food Science & Technology*, 2021, 108: 148–163.
- [14] FLEISCHMANN R D, MARK D A, OWEN W, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd [J]. *Science*, 1995, 269: 496–512.
- [15] BOLOTIN A, WINCKER P, MAUGER S, et al. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403 [J]. *Genome Research*, 2001, 11: 731–753.
- [16] WU R N, WANG W, YU D L, et al. Proteomics analysis of *Lactobacillus casei* Zhang, a new probiotic bacterium isolated from traditional home-made koumiss in Inner Mongolia of China [J]. *Molecular & Cellular Proteomics*, 2009, 8: 2321–2338.
- [17] SUN Z H, HARRIS H M, MCCANN A, et al. Expanding the biotechnology potential of *Lactobacilli* through comparative genomics of 213 strains and associated genera [J]. *Nature Communications*, 2015, 6: 8322.
- [18] SONG Y Q, ZHAO J, LIU W J, et al. Exploring

- the industrial potential of *Lactobacillus delbrueckii* ssp. *bulgaricus* by population genomics and genome-wide association study analysis[J]. *Journal of Dairy Science*, 2021, 104: 4044–4055.
- [17] LI K N, WANG S, LIU W J, et al. Comparative genomic analysis of 455 *Lactiplantibacillus plantarum* isolates: Habitat-specific genomes shaped by frequent recombination[J]. *Food Microbiology*, 2022, 104: 103989.
- [18] ZHAO J, WU L J, LI W C, et al. Genomics landscape of 185 *Streptococcus thermophilus* and identification of fermentation biomarkers[J]. *Food Research International*, 2021, 150: 110711.
- [19] JIN H, MA T, CHEN L, et al. The iLABdb: A web-based integrated lactic acid bacteria database[J]. *Science Bulletin*, 2023, 68: 2527–2530.
- [20] SCHROETER J, KLAENHAMMER T. Genomics of lactic acid bacteria[J]. *FEMS Microbiology Letters*, 2009, 292: 1–6.
- [21] NOBREGA M A, PENNACCHIO L A. Comparative genomic analysis as a tool for biological discovery[J]. *The Journal of Physiology*, 2004, 554: 31–39.
- [22] OJALA T, KANKAINEN M, CASTRO J, et al. Comparative genomics of *Lactobacillus crispatus* suggests novel mechanisms for the competitive exclusion of *Gardnerella vaginalis*[J]. *BMC Genomics*, 2014, 15: 1–21.
- [23] MAIDEN M C, HARRISON O B. Population and functional genomics of *Neisseria* revealed with gene-by-gene approaches[J]. *Journal of Clinical Microbiology*, 2016, 54: 1949–1955.
- [24] SUZUKI M, ALBARRACIN L, TSUJIKAWA Y, et al. *Lactiplantibacillus plantarum* LOC1 isolated from fresh tea leaves modulates macrophage response to TLR4 activation[J]. *Foods*, 2022, 11: 3257.
- [25] SUN Y W, ZHANG S Y, LI H, et al. Assessments of probiotic potentials of *Lactiplantibacillus plantarum* strains isolated from Chinese traditional fermented food: Phenotypic and genomic analysis[J]. *Frontiers in Microbiology*, 2022, 13: 895132.
- [26] 向雪松, 杨瑞馥. 益生菌、益生元、后生元与健康[J]. *营养学报*, 2023, 45(2): 105–106.
- XIANG X S, YANG R F. Probiotics, prebiotic, postbiotics and health[J]. *Acta Nutrimenta Sinica*, 2023, 45(2): 105–106.
- [27] 中国食品科学技术学会益生菌分会. 益生菌“量效关系”的科学综述[J]. *中国食品学报*, 2022, 22(7): 443–454.
- Probiotics Society of the Chinese Institute of Food Science and Technology. A scientific review of the ‘Dose-response relationship’ of probiotics[J]. *Journal of Chinese Institute of Food Science and Technology*, 2022, 22(7): 443–454.
- [28] HILL C, GUARNER F, REID G, et al. The international scientific association for probiotics and prebiotics consensus statement on the scope and appropriate use of the term probiotic[J]. *Nature Reviews Gastroenterology & Hepatology*, 2014, 11: 506–514.
- [29] ZOMMITI M, FEUILLOLEY M G, CONNIL N. Update of probiotics in human world: A nonstop source of benefactions till the end of time[J]. *Microorganisms*, 2020, 8: 1907.
- [30] DE MELO PEREIRA G V, DE OLIVEIRA COELHO B, JÚNIOR A I M, et al. How to select a probiotic? A review and update of methods and criteria[J]. *Biotechnology Advances*, 2018, 36: 2060–2076.
- [31] ZHANG Z Y, LIU C, ZHU Y Z, et al. Safety assessment of *Lactobacillus plantarum* JDM1 based on the complete genome[J]. *International Journal of Food Microbiology*, 2012, 153: 166–170.
- [32] WU C D, HUANG J, ZHOU R Q. Genomics of lactic acid bacteria: Current status and potential applications[J]. *Critical Reviews in Microbiology*, 2017, 43: 393–404.
- [33] PENG X, ED-DRA A, YUE M. Whole genome sequencing for the risk assessment of probiotic lactic acid bacteria[J]. *Critical Reviews in Food Science and Nutrition*, 2023, 63: 11244–11262.
- [34] YE D Y, LI X W, SHEN J Z, et al. Microbial metabolomics: From novel technologies to diversified applications[J]. *TrAC Trends in Analytical Chemistry*, 2022, 148: 116540.
- [35] ABDALLAH A, ELEMBA E, ZHONG Q, et al. Gastrointestinal interaction between dietary amino acids and gut microbiota: with special emphasis on host nutrition[J]. *Current Protein and Peptide Science*, 2020, 21: 785–798.
- [36] GASPAR P, CARVALHO A L, VINGA S, et al. From physiology to systems metabolic engineering for

- the production of biochemicals by lactic acid bacteria[J]. *Biotechnology Advances*, 2013, 31: 764–788.
- [37] MAZZOLI R, BOSCO F, MIZRAHI I, et al. Towards lactic acid bacteria based biorefineries [J]. *Biotechnology Advances*, 2014, 32: 1216–1236.
- [38] TSUJI A, OKADA S, HOLS P, et al. Metabolic engineering of *Lactobacillus plantarum* for succinic acid production through activation of the reductive branch of the tricarboxylic acid cycle[J]. *Enzyme and Microbial Technology*, 2013, 53: 97–103.
- [39] LADERO V, RAMOS A, WIERSMA A, et al. High-level production of the low-calorie sugar sorbitol by *Lactobacillus plantarum* through metabolic engineering[J]. *Applied and Environmental Microbiology*, 2007, 73: 1864–1872.
- [40] CAPOZZI V, MENGA V, DIGESU A M, et al. Biotechnological production of vitamin B₂-enriched bread and pasta[J]. *Journal of Agricultural and Food Chemistry*, 2011, 59: 8013–8020.
- [41] CAPOZZI V, RUSSO P, DUEÑAS M T, et al. Lactic acid bacteria producing B-group vitamins: A great potential for functional cereals products[J]. *Applied Microbiology and Biotechnology*, 2012, 96: 1383–1394.
- [42] KLAENHAMMER T R, KULLEN M J. Selection and design of probiotics[J]. *International Journal of Food Microbiology*, 1999, 50: 45–57.
- [43] PANTHEE S, PAUDEL A, BLOM J, et al. Complete genome sequence of *Weissella hellenica* 0916-4-2 and its comparative genomic analysis[J]. *Frontiers in Microbiology*, 2019, 10: 1619.
- [44] SYROKOU M K, PARAMITHIOTIS S, DROSINOS E H, et al. A comparative genomic and safety assessment of six *Lactiplantibacillus plantarum* subsp. *argenteratensis* strains isolated from spontaneously fermented greek wheat sourdoughs for potential biotechnological application [J]. *International Journal of Molecular Sciences*, 2022, 23: 2487.
- [45] JUMPER J, EVANS R, PRITZEL A, et al. Highly accurate protein structure prediction with AlphaFold [J]. *Nature*, 2021, 596: 583–589.
- [46] SABATER C, CALVETE -TORRE I, VILLAMIEL M, et al. Vegetable waste and by-products to feed a healthy gut microbiota: Current evidence, machine learning and computational tools to design novel microbiome-targeted foods[J]. *Trends in Food Science & Technology*, 2021, 118: 399–417.
- [47] MCCOUBREY L E, SEGOBIN N, ELBADAWI M, et al. Active machine learning for formulation of precision probiotics[J]. *International Journal of Pharmaceutics*, 2022, 616: 121568.
- [48] MCCOUBREY L E, ELBADAWI M, ORLU M, et al. Machine learning uncovers adverse drug effects on intestinal bacteria[J]. *Pharmaceutics*, 2021, 13: 1026.
- [49] ZHAO Y, WANG C, CHEN X. Microbes and complex diseases: from experimental results to computational models[J]. *Briefings in Bioinformatics*, 2021, 22: 158.

Application of Probiogenomics in Screening and Functional Evaluation of Probiotics Lactic Acid Bacteria

Liu Wenjun, Zhang Heping*

(Key Laboratory of Dairy Biotechnology and Engineering, Ministry of Education, Key Laboratory of Dairy Products Processing, Ministry of Agriculture and Rural Affairs, Inner Mongolia Key Laboratory of Dairy Biotechnology and Engineering, Collaborative Innovative Center of Ministry of Education for Lactic Acid Bacteria and Fermented Dairy Products, Inner Mongolia Agricultural University, Hohhot 010018)

Abstract *Lactobacillus* and *Bifidobacterium* are the main microbial groups for probiotics screening, and have important application value in promoting human, animal, and even plant health. They are widely used in various functional foods, animal feed, and microbial fertilizers. However, the molecular mechanisms by which these bacteria have a positive impact on host health are still far from fully understood. In addition, traditional probiotic screening techniques mainly rely on the determination of phenotype and physiological and biochemical characteristics, as well as randomized controlled tri-

als. These methods are not only time-consuming and laborious, but also difficult to replicate data. It seriously restricted the development of probiotics science and the utilization of strains because the difficulty in sharing data among different research groups, as well as the inability to achieve uniformity and standardized validation of beneficial characteristics of strains. The emergence of a new discipline called Probiogenomics, based on new sequencing techniques and powerful computer analysis methods, allowed researchers to significantly improve the efficiency and accuracy of probiotic screening by screening large amounts of biological data in a short period of time, greatly promoted the identification of probiotic characteristics and mechanism characterization. The latest research results on how genomic data analysis can help people efficiently and quickly identify probiotics, open up new functional evaluation pathways for probiotics, and evaluate the molecular basis of probiotic characteristic functions were reviewed in the present paper. The aim was to provide a new perspective for expanding our understanding of these beneficial microbial genomics and revealing the molecular mechanisms by which probiotics promote health characteristics.

Keywords probiogenomics; lactic acid bacteria; function; screening; artificial intelligence