

基于基因组的清酒广布乳杆菌遗传多样性研究

冉小旭,徐婷婷,包凯迪,陈仙兰,王 婵,李永波,张振东*

(黔南民族师范学院 生物科学与农学院,贵州 黔南 558000)

摘要:为分析清酒广布乳杆菌(*Latilactobacillus sakei*)的遗传多样性,深入理解其分子进化和生态适应机制,对已公开基因组序列的74株清酒广布乳杆菌开展了基因组比较分析、群体结构分析、正选择分析,并与已有的多位点分型方法进行比较。结果表明,清酒广布乳杆菌基因组大小为1.9~2.1 Mbp,GC含量为40.8%~41.2%,编码基因数量为1 776~2 134个。其基因组中含有与淀粉、纤维素、几丁质降解及葡聚糖、果糖、乳糖和半乳糖代谢相关的酶类基因,表明清酒广布乳杆菌能够适应多种环境。基于已公布的基因组数据,群体结构分析、系统发育分析和主成分分析表明,清酒广布乳杆菌可分为3个类群。正选择分析和基因组单核苷酸多态性(SNP)分布分析显示,基因*rplR*受到正选择作用,基因*agrA*的SNP位点数量最多。基于8个持家基因的多位点分型分析将74株菌株划分为46个ST型,ST7为主要型别,且基因*hemN*显著重组。研究结果为深入解析清酒广布乳杆菌的功能差异和遗传进化提供了理论依据。

关键词:基因组特征;清酒广布乳杆菌;群体结构;基因组功能

中图分类号:TS261.1

文章编号:0254-5071(2026)03-0098-07

doi:10.11882/j.issn.0254-5071.2026.03.014

引文格式:冉小旭,徐婷婷,包凯迪,等. 基于基因组的清酒广布乳杆菌遗传多样性研究[J]. 中国酿造,2026,45(3):98-104.

Genetic diversity of *Latilactobacillus sakei* based on genomic

RAN Xiaoxu, XU Tingting, BAO Kaidi, CHEN Xianlan, WANG Chan, LI Yongbo, ZHANG Zhendong*

(School of Biological Science and Agriculture, Qiannan Normal University for Nationalities, Qiannan 558000, China)

Abstract:To analyze the genetic diversity of *Latilactobacillus sakei*, and to gain a deeper understanding of its molecular evolution and ecological adaptation mechanisms, genomic comparative analysis, population structure analysis, and positive selection analysis were conducted on 74 *L. sakei* strains with published genome sequences. These analyses were then compared with existing multilocus sequence typing methods. The results showed that the genome size of *L. sakei* ranged from 1.9 to 2.1 Mbp, with GC content of 40.8% to 41.2%, and the number of coding genes ranged from 1 776 to 2 134. The genome contained genes for enzymes involved in the degradation of starch, cellulose, chitin, as well as the metabolism of glucans, fructose, lactose, and galactose, indicating the bacterium's adaptability to diverse environments. Based on the publicly available genome data, population structure, phylogenetic, and principal component analyses indicated that *L. sakei* could be divided into three major groups. Positive selection analysis and genome single nucleotide polymorphism (SNP) distribution analysis revealed that the gene *rplR* was subject to positive selection, while *agrA* exhibited the highest number of SNP sites. Multilocus sequence typing analysis based on eight housekeeping genes divided the 74 strains into 46 sequence types (STs), with ST7 being the most prevalent. Additionally, significant recombination was observed in the *hemN* gene. The findings provided theoretical insights into the functional differentiation and genetic evolution of *L. sakei*.

Key words:genomic characteristics; *Latilactobacillus sakei*; population structure; genomic function

清酒广布乳杆菌(*Latilactobacillus sakei*)是一种革兰氏染色阳性、短杆状、无鞭毛和芽孢的乳酸菌,属于乳杆菌属(*Lactobacillus*)、厚壁菌门(Firmicutes),是传统发酵食品中的常见菌种^[1-3]。清酒广布乳杆菌广泛存在于清酒、鲑广椒、米酒等发酵食品中^[4],并已被纳入允许用于食品的菌种名单^[5],被世界各国普遍认为是安全的(generally recognized as safe, GRAS)食品加工菌种。该菌在食品工业中具有重要的应用价值,广泛用于肉类发酵剂,并对大肠杆菌和金黄色葡萄球菌等具有显著抑制作用^[6]。此外,清酒广布乳杆菌在发酵过程中能降低霉鱼的腥味和酸臭味,改善发酵食品的风味^[6]。研究还发现,清酒广布乳杆菌具有益生

作用,可有效缓解结肠炎、类风湿性关节炎等免疫性疾病,以及肥胖等代谢性疾病^[7]。

随着基因组测序技术的进步,已公布的清酒广布乳杆菌基因组逐渐增多,推动了对该菌基因组的深入研究。对不同来源的清酒广布乳杆菌菌株进行基因组分析发现,其基因组中富含与果糖、蔗糖、海藻糖及葡萄糖酸盐代谢相关的基因,表明该菌能适应多种糖类的代谢^[8]。此外,仅在部分清酒广布乳杆菌菌株中发现了与核苷清除、精氨酸代谢及应对氧化还原变化相关的基因,显示出特定菌株对环境的适应性^[9]。近年来,随着消费者对发酵肉制品、乳制品和蔬菜类食品品质的要求日益提高,清酒广布乳杆菌等乳

收稿日期:2025-07-01

修回日期:2025-10-07

基金项目:黔南民族师范学院大学生创新创业训练计划项目;贵州省基础 Research 计划(自然科学类)青年引导项目(黔科合基础QN(2025)351)

作者简介:冉小旭(2003-),女,本科生,研究方向为食品生物技术。

*通讯作者:张振东(1985-),男,副教授,博士,研究方向为食品微生物。

酸菌在发酵食品中的重要作用也逐渐被认识。因此,对清酒广布乳杆菌基因组的分析有助于深入理解其在发酵食品中的分子机制、适应性及遗传进化过程,从而有助于筛选出具有优良品质的菌株。

尽管目前已有关于清酒广布乳杆菌基因组的诸多研究,但对其基因组遗传多样性的整体了解仍较为有限。为系统地分析清酒广布乳杆菌的基因组遗传多样性,本研究从美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI) RefSeq数据库下载了74株清酒广布乳杆菌基因组序列,以*Latilactobacillus sakei* subsp. *sakei* DSM 20017为参考菌株,开展了一系列基因组水平的比较分析,并与已有的多位点分型方法进行了对比,以揭示清酒广布乳杆菌的遗传进化特征,为相关菌株的功能研究与应用奠定基础。

1 材料与方法

1.1 材料与试剂

清酒广布乳杆菌基因组数据获取:从NCBI的GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>)数据库下载所有清酒广布乳杆菌(*Latilactobacillus sakei*)基因组序列(截止2024年1月前已发布的基因组数据),经过检查后去掉重复菌株的基因组,共保留74个清酒广布乳杆菌基因组数据。

1.2 仪器与设备

PowerEdge R930服务器:戴尔(中国)有限公司。

1.3 方法

1.3.1 基因组注释与分析

使用QUAST (<https://github.com/ablab/quast>)^[10]对基因组组装质量、核糖体核糖核酸(ribosomal ribonucleic acid, rRNA)数量、GC含量、基因长度等数据进行评估。由于上传到NCBI数据库中清酒广布乳杆菌的基因组的注释并未使用相同的软件,为减小因软件不同造成的差异度,对下载的所有基因组统一使用Prokka(Rapid Prokaryotic genome annotation)^[11]软件重新进行注释。使用软件run_dbcan (https://github.com/linnabrown/run_dbcan/),软件基于prokka预测的清酒乳酸杆菌基因组中编码区对应的氨基酸序列,对其碳水化合物活性酶进行注释。

1.3.2 比较基因组分析

使用pyani (<https://github.com/widdowquinn/pyani>)软件计算两两基因组间核苷酸一致性(average nucleotide identity, ANI)。使用软件EzAAI (<http://leb.snu.ac.kr/ezai>)计算两个基因组之间的平均氨基酸一致性(amino acid identity, AAI)^[12]。使用Roary(version 3.11.0)^[13]进行核心基因组和泛基因组构建。

1.3.3 基于基因组的进化树构建

以Roary提取的核心基因组基因集为基础进行进化树构建:首先对提取到的核心基因组基因集使用Mafft (<https://github.com/GSLBiotech/mafft>)来进行比对;多序列比对后,使用iqtree^[14]建树。

1.3.4 清酒广布乳杆菌遗传多样性、Ka/Ks比值、单核苷酸多态性分布与Tajima's D值分析

清酒广布乳杆菌基因组的多样性指数(polymorphism information content, PI)、TajimaD和单核苷酸多态性(single nucleotide polymorphism, SNP)位点使用vcftools (<https://github.com/vcftools/vcftools>)计算和统计;正选择分析基于清酒广布乳杆菌的核心基因组基因集,使用软件Hyphy (<https://github.com/veg/hyphy>)进行计算。使用paml v4.7j,基于零假设(Null hypothesis)计算每个持家基因的Ka/Ks(即非同义替换率与同义置换速率的比值),进行选择压力分析。

1.3.5 多位点分型分析

基于CHAILLOU S等^[15]建立的多位点序列分型方案,纳入了8个持家基因:分子伴侣蛋白基因*dnaK*、甘油通道蛋白基因*glpF*、粪卟啉原氧化酶基因*hemN*、L-乳酸脱氢酶基因*ldhL*、肌肽酶基因*PepV*、重组酶A基因*recA*、RNA聚合酶 β 亚基基因*rpoB*、延伸因子Tu基因*tuf*,通过这8个持家基因进行清酒广布乳杆菌的多位点分型分析与群体结构分析。使用ttools (<https://github.com/CJ-Chen/TBtools-II/>)提取74个菌基因组上的8个持家基因的序列,使用BioNumerics 7.6软件识别相应菌株的持家基因等位基因号和序列类型(sequence type, ST)型,构建最小生成树(minimum spanning tree, MSTree),分析各ST型与分离来源之间的关系。

1.3.6 基于多位点分型方法的遗传进化分析

将使用mafft v7.311比对过的74个清酒广布乳杆菌的8个持家基因的序列串联起来,得到一个3 334 bp的串联序列,用iqtree软件构建系统进化树,分析各ST型的系统发育关系。

1.3.7 清酒广布乳杆菌的群体结构分析

基于roary软件得到的74株清酒广布乳杆菌核心基因组基因集,使用java软件PGDSpider^[16]把mafft比对过的核心基因组基因集转为SNP文件,使用admixture v1.3.0 (<https://dalexander.github.io/admixture/>)软件分析清酒广布乳杆菌的群体结构。

2 结果与分析

2.1 清酒广布乳杆菌基因组特征

利用QUAST评估基因组组装质量,包括rRNA数量、GC含量、基因长度等,绘制了GC含量密度图和基因组大小密度图,结果见图1。由图1中GC含量密度图显示,清酒广布乳杆菌基因组的GC含量主要集中在40.8%~41.2%,74株菌株的平均GC含量为41.04%。基因组大小密度图表明,基因组大小主要分布在1.9~2.1 Mbp,平均值为1.9 Mb,高于EISENBACH L等^[18]的研究结果。编码序列(coding sequences, CDS)数量范围为1 776~2 134。模式菌株*Latilactobacillus sakei* subsp. *sakei* DSM 20017的GC含量为41.15%,主要分布在40.8%~41.2%,略低于EISENBACH L等^[18]的结果。样本量差异可能是原因之一,本研究纳入了

74株菌株,样本量显著高于其10株菌株,因此结果更具有代表性。

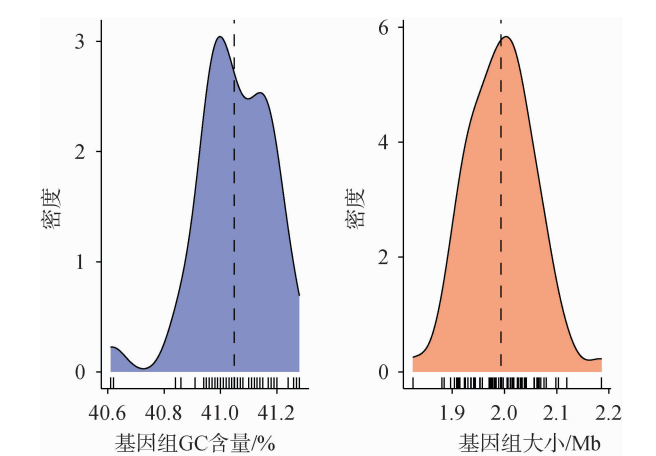


图1 基因组GC含量和基因组大小密度图
Fig. 1 Density plot of genomic GC content and genome size

2.2 基于基因组的清酒广布乳杆菌鉴定

通过计算74株清酒广布乳杆菌菌株的平均核苷酸一致性(ANI),发现菌株*L. sakei* subsp. *sakei* DS4与模式菌株*L. sakei* subsp. *sakei* DSM 20017的ANI值为92.53%,低于分类标准的95%,推测其可能为新的分类单元。而其余73株菌株与模式菌株的ANI值均高于95%,证明这些菌株均属于清酒广布乳杆菌^[17]。平均氨基酸一致性(AAI)的结果显示,NCBI下载的74株菌株与模式菌株间的AAI值为95.67%~99.98%。其中,菌株DS4与DSM 20017基因组间的AAI值最低,为95.67%,但仍高于95%。因此,综合判断,所有下载的74株菌株均被鉴定为清酒广布乳杆菌(*Latilactobacillus sakei*)。

2.3 清酒广布乳杆菌的核心基因组与泛基因组分析

为探讨清酒广布乳杆菌基因组多样性,构建了清酒广布乳杆菌核心基因组和泛基因组,并分析了独特基因,绘制了基因数量曲线图,结果见图2。

由图2A可知,随着分析菌株数量的增加,核心基因组的基因数量逐渐减少,曲线呈下降趋势。单个菌株基因组的平均基因数量为1 927个,而74株菌株核心基因组的基因数量为1 102个。这意味着清酒广布乳杆菌的核心基因组相对较为保守,核心基因在菌株间的一致性较高。由图2B可知,与之相反,清酒广布乳杆菌的泛基因组基因数量随菌株数量增加而逐渐增多。单个基因组的平均泛基因组基因数量为1 927,而74株菌株的泛基因组基因数量为6 032个。这反映了该物种的泛基因组呈开放状态,表明其通过水平基因转移等方式不断获得新基因,以适应多样环境。由图2C可知,菌株特有基因的数量也随着基因组数量增加而逐渐增多,74株菌株平均独特基因数量为1 751个。这些特有基因可能赋予清酒广布乳杆菌菌株独特的代谢或耐受能力,

进一步凸显了种内遗传多样性。综合来看,清酒广布乳杆菌的核心基因组较为保守,而泛基因组和特有基因展示了较大的多样性。这表明该菌种可能通过其可变基因或特有基因的多样性,在不同环境或宿主中具有较强的适应能力。

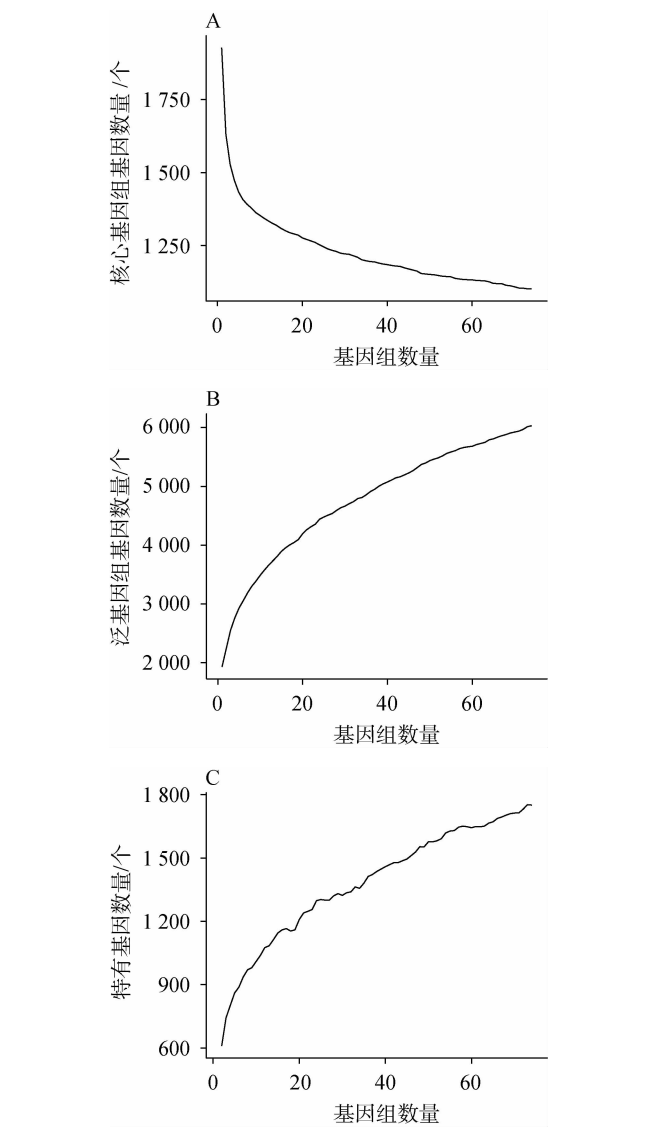


图2 清酒广布乳杆菌的核心基因组(A)、泛基因组(B)和特有基因(C)数量分析

Fig. 2 Quantitative analysis of the core genome (A), pan-genome (B), and unique genes (C) of *Latilactobacillus sakei*

2.4 清酒广布乳杆菌的碳水化合物活性酶分析

基因组包含了生物的全部遗传信息,通过比较基因组研究,能更好的了解物种的功能基因分布情况^[18-19]。利用最新的碳水化合物酶基因数据库分析清酒广布乳杆菌基因组中的碳水化合物活性酶基因,结果见图3。

由图3可知,这些碳水化合物活性酶基因可分为5大类,包括辅助活性酶(auxiliary activities,AA),碳水化合物结合模块(carbohydrate-binding modules,CBM),碳水化合物酯酶(carbohydrate esterases,CE),糖苷水解酶(glycoside

hydrolases, GH)、糖基转移酶(glycosyltransferases, GT), 以及40个亚家族。由图3A可知, 其中糖苷水解酶类和糖基转移酶类的数量最多。

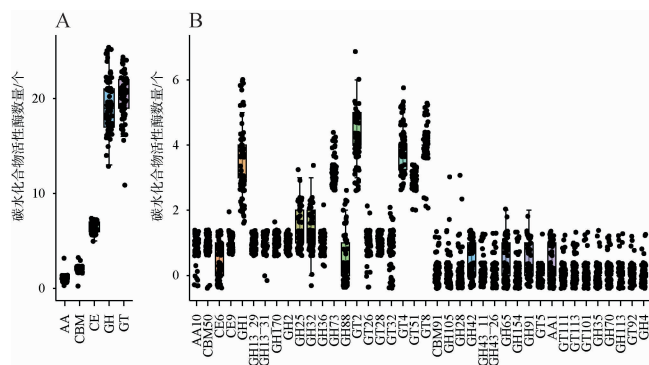


图3 碳水化合物活性酶种类与数量的一级分类(A)与二级分类(B)
Fig. 3 Types and quantitative distribution of carbohydrate enzymes:
Primary classification (A) and secondary classification (B)

由图3B可知, 在碳水化合物活性酶的二级分类中, 纤维素或几丁质合成酶(glycosyltransferase family 2, GT2)的基因数量居首, 其次依次为 β -葡萄糖苷酶(glycoside hydrolase family 1, GH1)、白脂磷酸 β -D-甘露糖基转移酶(glycosyltransferase family 4, GT4)以及半乳糖基/葡萄糖基转移酶(glycosyltransferase family 8, GT8)。

此外, 清酒广布乳杆菌基因组中还编码多种与多糖降解相关的基因。例如, 在淀粉分解方面, 该菌含有淀粉酶(glycoside hydrolase family 13, GH13_29), 它主要参与淀粉的降解过程; 支链淀粉酶(glycoside hydrolase family 13, GH13_31), 负责支链淀粉的分解; 以及磷酸化酶(glycoside hydrolase family 4, GH4), 通过水解磷酸化糖苷键来促进淀粉及其相关多糖的进一步降解。在几丁质分解方面, 相关基因包括铜依赖单加氧酶(auxiliary activity family 10, AA10), 它氧化几丁质及其衍生物以辅助降解; 几丁质结合模块(carbohydrate-binding module family 50, CBM50), 通过特异性结合几丁质来增强降解酶的亲和力和效率, 从而加速几丁质的分解; 以及几丁质水解酶(glycoside hydrolase family 154, GH154), 专一水解几丁质中的 β -1,4-葡萄糖胺键。与木质素相关降解方面, 该菌编码木质素氧化酶(auxiliary activity family 1, AA1), 它氧化木质素中的酚类物质以促进其初步降解; 以及木聚糖乙酰基水解酶(carbohydrate esterase family 6, CE6), 通过水解木聚糖中的乙酰基团来辅助木质素相关多糖(如木聚糖)的处理。在果胶利用方面, 果胶降解酶(glycoside hydrolase family 28, GH28)水解果胶中的 α -1,4-半乳糖醛酸键, 而果胶去乙酰酶(carbohydrate esterase family 9, CE9)则去除果胶中的乙酰基团, 进一步便利其降解。这些基因的存在与该菌在环境中的适应能力有关。

2.5 清酒广布乳杆菌群体结构分析

基于74株清酒广布乳杆菌菌株核心基因组的DNA序

列, 构建最大似然法系统发育树、群体结构分析结果、基于SNP数据进行主成分分析, 结果见图4。

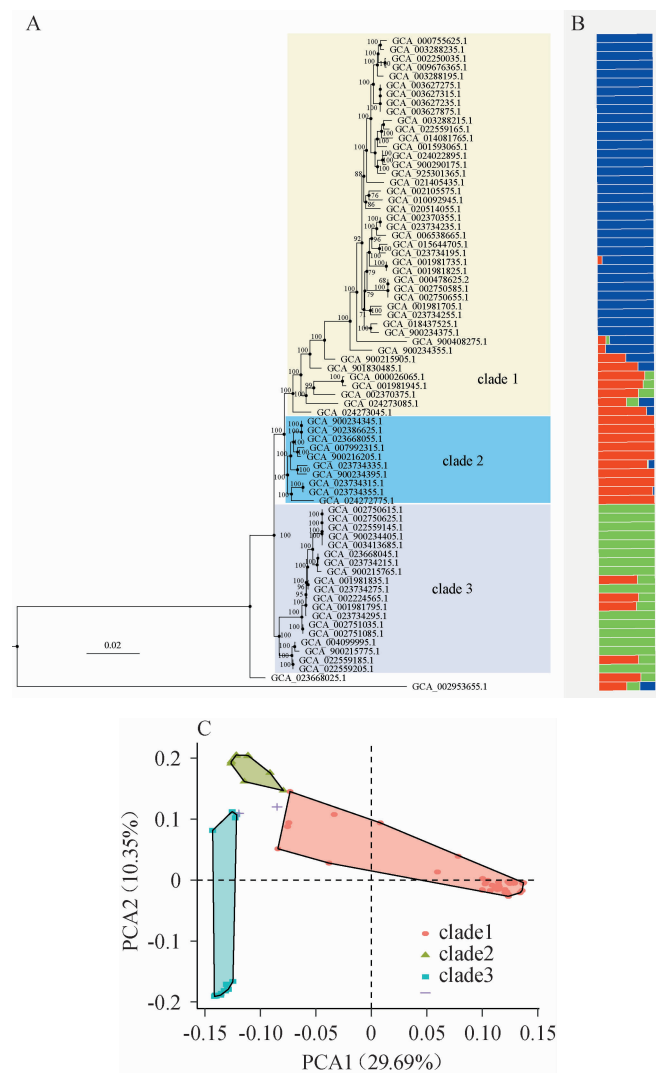


图4 清酒广布乳杆菌系统发育树(A)、群体结构分析(B)和主成分分析(C)结果

Fig. 4 Phylogenetic tree (A), population structure analysis (B) and principal component analysis (C) of *Latilactobacillus sakei*

由图4A可知, 由进化树显示自展值普遍为100, 表明系统发育树拓扑结构可信度高。74株菌株在系统发育树上分为3个主要分支(clade), 其中模式菌株*L. sakei* subsp. *sakei* DSM 20017位于第1分支(clade I), *L. sakei* subsp. *carnosus* DSM 20017位于第2分支(clade II)。由图4B可知, 不同颜色代表不同的遗传亚群, 每一行表示单个菌株在各亚群中的遗传成分比例。群体分析表明, 纳入分析的清酒广布乳杆菌可以分类3个亚群。进一步基于SNP数据进行主成分分析(principal component analysis, PCA), 结果见图4C, 由图4C可知, 这些菌株形成了3个明显的聚类, 且与系统发育分析结果一致。在主成分1和主成分2上, 各聚类间差异显著。

2.6 清酒广布乳杆菌遗传多样性、Ka/Ks、SNP分布与Tajima's D值分析

计算和展示了清酒广布乳杆菌基因组的单核苷酸多态性(SNP)位点、非同义突变率(Ka)与同义突变率(Ks)比值、Tajima's D值、遗传多样性指数(PI值),结果见图5。

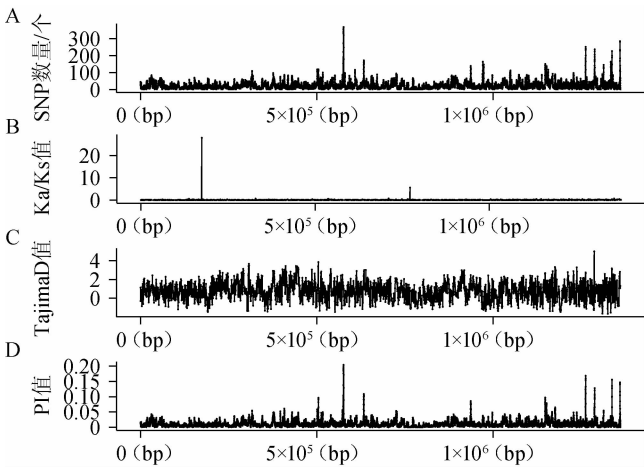


图5 基因组单核苷酸多态性分布(A)、Ka/Ks值(B)、Tajima's D值(C)和基因多样性分析(D)
Fig. 5 Genomic single nucleotide polymorphism distribution (A), Ka/Ks values (B), Tajima's D values (C) and genetic diversity analysis (D)

单核苷酸多态性(SNP)是指基因组水平上由单个核苷酸变异引起的DNA序列多样性,SNP数量越多,表明多态性越高^[20]。由图5A可知,在清酒广布乳杆菌核心基因组中,编号为g00878的基因(位于575 259~575 798 bp)的SNP数量最多,共计262个,但该基因被注释为假定蛋白。其次是基因g00388(编码agr群体感应系统的应答基因agrA,位于1 359 089~1 359 835 bp),其SNP数量为260个^[21]。

Ka/Ks比值是生物信息学中常用的指标,其中Ka指非同义替换位点替换率,Ks指同义替换位点替换率。若Ka/Ks>1(即Ka>Ks),则表明基因受正选择作用;若Ka/Ks=1,则表示基因为中性进化;而Ka/Ks<1(即Ka<Ks)则说明基因受到纯化选择^[22]。由图5B可知,在核心基因组174 783~174 950 bp区间内,编号为g00037的基因的Ka/Ks值为27.949 7,其表达产物为假定蛋白。在772 558~772 917 bp区间,基因rplR(编号g03462)的Ka/Ks值为5.600 2,其表达产物为50S核糖体蛋白L18(50S ribosomal protein L18)。Ka/Ks>1,表明这两个基因受到正选择作用(图5B)。rplR基因的表达产物为50S核糖体蛋白L18(50S ribosomal protein L18)。由于该基因受到强烈的正选择作用,推测该基因在核糖体功能优化或环境适应性方面具有重要意义^[23]。

Tajima's D值用于衡量群体观测杂合度与预期杂合度的偏差。Tajima's D>0表明群体的观测杂合度低于预期,稀有等位基因频率较低;Tajima's D<0则表明群体观测杂合度

高于预期^[24]。由图5C可知,Tajima's D值的范围为-1.652 25~5.005 18。其中最大值5.005 18出现在核心基因组1 286 163~1 286 546 bp区域,对应基因编号为g01434,其表达产物为假定蛋白;最小值-1.652 25出现在1 323 727~1 326 258 bp区域,对应基因g00442.1。该基因编码氨肽酶N,一种广泛分布的膜结合型外肽酶,在调控细胞周期、迁移、分化、增殖和凋亡等生理活动中具有重要作用^[25]。

遗传多样性指数(PI值)是衡量基因组内部变异水平的指标。PI值越高,说明基因组具有更高的遗传多样性^[26]。由图5D可知,PI值的最大值为0.204 5,位于核心基因组575 259~575 798 bp区域,对应基因编号为g00878;其次是编号为g01345的基因(位于575 701~576 200 bp),其PI值为0.203 9。两者均被注释为假定蛋白,表明这些区域在清酒广布乳杆菌基因组中具有最高的变异水平。

2.7 清酒广布乳杆菌多位点分型分析

参照CHAILLOU S等^[15]的方法,基于8个持家基因,对纳入基因组比较分析的74株清酒广布乳杆菌进行多位点序列分型(multilocus sequence typing, MLST)分析。纳入MLST分析的8个持家基因片段长度,其GC含量见表1,多位点分型结果见图6。

表1 用于多位点序列分型分析的8个持家基因特征与重组检测 Table 1 Characteristics of 8 housekeeping genes used for multilocus sequence typing analysis and recombination detection			
基因	大小/bp	GC含量/%	Phi系数(P值)
dnaK	486	39.6	0.18(P=0.207 9)
glpF	399	46.9	0.11(P=0.228 6)
hemN	369	38.7	0.28(P=0.003 346)
ldhL	442	41.3	0(P=1)
PepV	540	41.2	0(P=1)
recA	399	45.1	0.01(P=0.312 5)
rpoB	369	39.1	0.21(P=0.642 1)
tuf	330	44.1	0(P=1)

由表1可知,基于分裂重组分析,发现用于清酒广布乳杆菌多位点分型分析的基因中,基因hemN在74株清酒广布乳杆菌中发生了显著的重组(P<0.05),这一点是CHAILLOU S等^[15]研究中未发现的。该基因在卟啉和血红素合成中起重要作用,所编码酶有助于菌株适应不同氧气浓度环境,增强其生存能力^[27]。

由图6可知,多位点分型结果显示,纳入分析的74株清酒广布乳杆菌被划分为46个ST型。其中,ST7是最多的ST型,共包括10株菌,占总菌株数的13.51%,其分离来源为发酵可可、香肠、清酒发酵剂、肉类和未知来源。其次是ST3和ST13,分别有6株(占8.11%)和5株(占6.76%)。ST3的分离来源为发酵蔬菜和牛奶,而ST13的来源为肉类、发酵蔬菜和香肠。这表明,清酒广布乳杆菌的序列分型结果与分离来源之间没有显著的特异性关联。

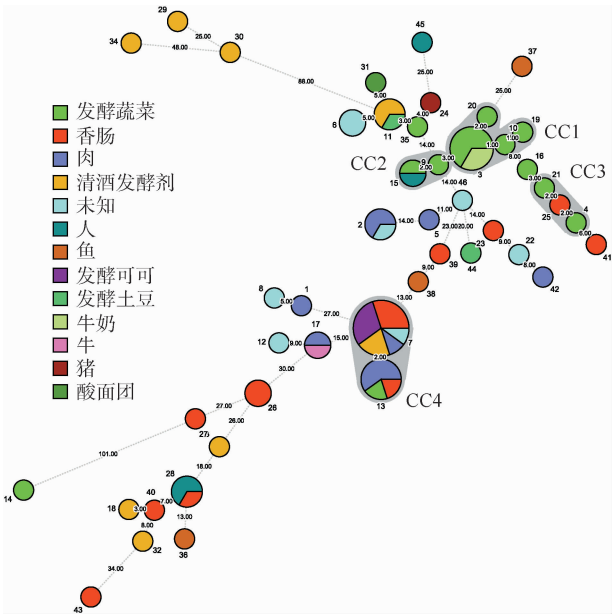


图6 74株清酒广布乳杆菌分离源最小生成树分析
Fig. 6 Minimum spanning tree analysis of 74 strains of *Lactilobacillus sakei*

进一步分析显示,74株清酒广布乳杆菌可分为4个克隆复合体(clonal complex,CC)和44个克隆单体(singleton)。4个克隆复合体包含30株菌,占总菌株数的40.54%。其中,CC1包含4个ST型(ST3、ST10、ST19和ST20)共9株菌,CC2包含2个ST型(ST9和ST15)共3株菌,CC3包含3个ST型(ST4、ST21和ST25)共3株菌,而CC4包含2个ST型(ST7和ST13)共15株菌,是主要的优势克隆复合体。

2.8 基于多位点分型法的群体结构分析

将用于多位点序列分型的8个持家基因串联起来,构建了基于最大似然法的系统发育树(图7A)。由图7A可知,44个ST型通过聚类形成3个亲缘关系较近类群。同时,ST14(菌株DS4,基因组序列号GCA_002953655.1)和ST27(菌株TMW 1.1397,基因组序列号GCA_027574105.1)形成独立分支,表明这两种ST型与其他ST型的进化距离较远,不属于上述3个类群。结合基因组鉴定结果,推测这两株清酒广布乳杆菌可能代表清酒广布乳杆菌的新亚种。

基于8个持家基因的群体结构图、分裂重组图及主成分分析结果见图7B~7D。由图7B可知,基于Admixture的群体结构分析,74株清酒广布乳杆菌可被划分为3个类群,分别用3种颜色标记。由图7C可知,SplitsTree4的分裂重组图显示,这些菌株分为3个主要的群体,使用相同的3种颜色标记,这与Admixture分析结果中的3个类群相对应。由图7D可知,主成分分析结果表明,3个类群形成了明显的3个区域,前两个主成分的结果进一步证实了这一划分;特别地,对应系统发育树中ST14和ST27的个体在PCA图上不属于这3个主要类群,突显了它们在进化上与其他清酒广布乳杆菌的差异性。

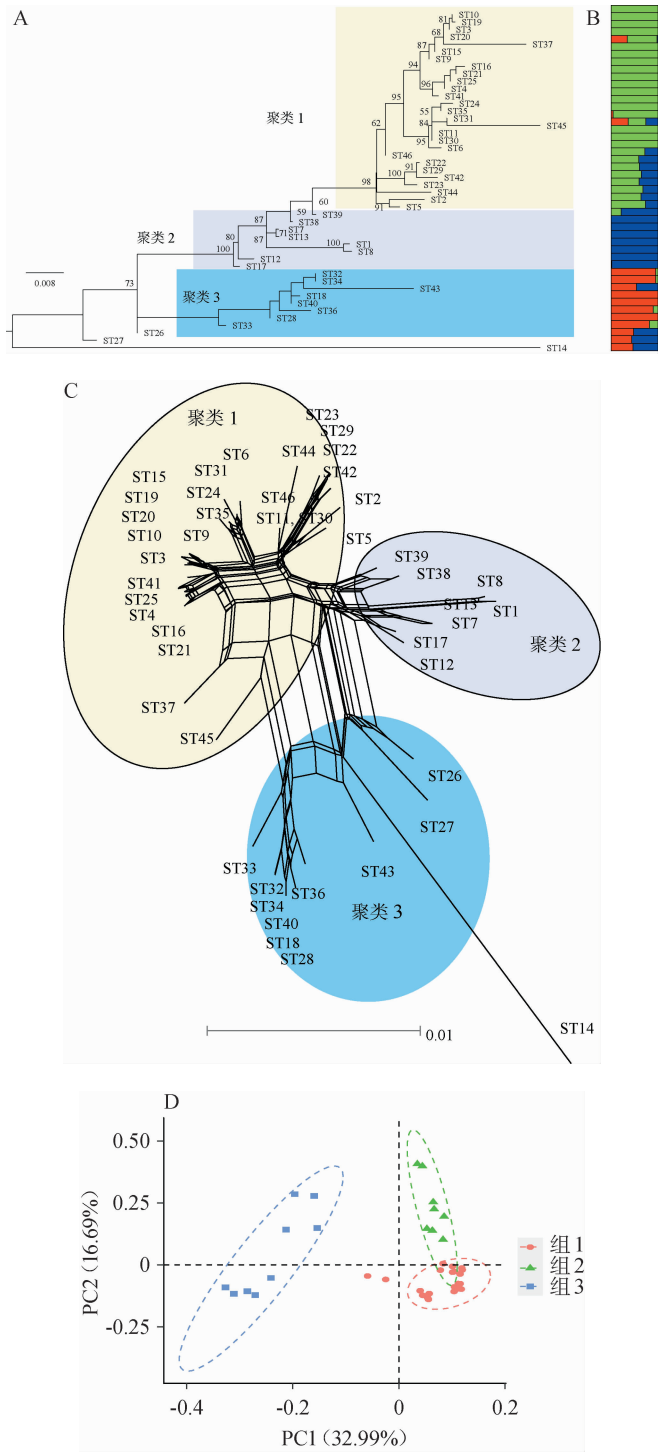


图7 基于8个持家基因46个ST型的系统进化树(A)、群体结构图(B)、分裂重组分析(C)及主成分分析结果(D)
Fig. 7 Phylogenetic evolutionary tree (A), population structure plot (B), split decomposition analysis (C), and principal component analysis (D) based on eight housekeeping genes

基于基因组和8个持家基因片段的群体结构分析,均可将清酒广布乳杆菌划分为3个主要群体,这一结果与CHAILLOU S等^[15]的研究一致,表明这8个持家基因的进化趋势与基因组进化情况较为一致,适用于清酒广布乳杆菌

菌的分型分析。值得注意的是,菌株DS4和TMW1.1397无论是基于基因组还是持家基因的系统发育分析,均显示其与其余74株清酒广布乳杆菌的进化关系较远,且与模式菌株的ANI值较低。目前清酒广布乳杆菌包含两个亚种(*L. sakei* subsp. *carneus*和*L. sakei* subsp. *sakei*),而这两株菌与这两个亚种的进化关系均较远,推测可能属于尚未命名的两个新亚种^[28]。

3 结论

本研究对74株已公开基因组序列的清酒广布乳杆菌进行了遗传进化分析。比较基因组分析结果显示,其核心基因组和泛基因组分别包含1 102和6 032个基因家族。基于8个持家基因以及全基因组数据构建的群体结构分析结果一致,将该种群划分为3个主要类群,菌株DS4和TMW1.1397株可能代表两个潜在的新亚种,不属于这三个类群。另外,在*hemN*基因中检测到显著的基因重组信号,而*rplR*基因($K_a/K_s=5.60$)被鉴定为正选择靶点。功能注释显示,该种基因组富含编码淀粉、纤维素和果胶等多糖降解酶的基因,这有助于解释其在肉类发酵和植物基食品中的生态适应性。

参考文献:

- [1] DE SOUZA E L, DE OLIVEIRA K Á R, DE OLIVEIRA M E G. Influence of lactic acid bacteria metabolites on physical and chemical food properties[J]. *Curr Opin Food Sci*, 2023, 49: 100981.
- [2] HU Y, WANG J, LIN Q, et al. Unraveling the difference in flavor characteristics of dry sausages inoculated with different autochthonous lactic acid bacteria[J]. *Food Biosci*, 2022, 47: 101778.
- [3] 赵旌茹. 发酵肉制品中菌株的筛选及对香肠风味的影响[D]. 沈阳: 沈阳农业大学, 2023.
- [4] LI Z J, LI Y H, XIAO C G, et al. Genomic and metabolic features of the *Lactobacillus sakei* JD10 revealed potential probiotic traits[J]. *Microbiol Res*, 2022, 256: 126954.
- [5] 国家卫生健康委员会. 解读关于《可用于食品的菌种名单》和《可用于婴幼儿食品的菌种名单》更新的公告(2022年第4号)[J]. *饮料工业*, 2022, 25(5): 3-4.
- [6] 曾金秀, 尹红梅, 张权等. 霉鱼源清酒乳杆菌的分离筛选及其在霉鱼发酵中的应用[J]. *食品与发酵工业*, 2024, 45(6): 56-62.
- [7] 陈颖. 具有缓解炎症性肠病功效的清酒乳杆菌的分离筛选及其作用机制探究[D]. 无锡: 江南大学, 2023.
- [8] EISENBACH L, GEISSLER A J, EHMAN M A, et al. Comparative genomics of *Lactobacillus sakei* supports the development of starter strain combinations[J]. *Microbiol Res*, 2019, 221: 1-9.
- [9] NYQUIST O L, MCLEOD A, BREDE D A, et al. Comparative genomics of *Lactobacillus sakei* with emphasis on strains from meat[J]. *Mol Genet Genomics*, 2011, 285: 297-311.
- [10] MIKHEENKO A, PRJIBELSKI A, SAVELIEV V, et al. Versatile genome assembly evaluation with QUAST-LG[J]. *Bioinformatics*, 2018, 34(13): 142-150.
- [11] SEEMANN T. Prokka: rapid prokaryotic genome annotation[J]. *Bioinformatics*, 2014, 30(14): 2068-2069.
- [12] KIM D, PARK S, CHUN J, et al. Introducing EzAAI: A pipeline for high throughput calculations of prokaryotic average amino acid identity [J]. *J Microbiol*, 2021, 59: 476-480.
- [13] PAGE A J, CUMMINS C A, HUNT M, et al. Roary: Rapid large-scale prokaryote pangenome analysis[J]. *Bioinformatics*, 2015, 31(22): 3691-3693.
- [14] MINH B Q, SCHMIDT H A, CHERNOMOR O, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era[J]. *Mol Biol Evol*, 2020, 37(5): 1530-1534.
- [15] CHAILLOU S, LUCQUIN I, NAJJARI A, et al. Population genetics of *Lactobacillus sakei* reveals three lineages with distinct evolutionary histories[J]. *PLOS One*, 2013, 8(9): e73253.
- [16] LISCHER H E L, EXCOFFIER L. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs[J]. *Bioinformatics*, 2012, 28(2): 298-299.
- [17] KIM M, OH H S, PARK S C, et al. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes[J]. *Int J Syst Evol Microbiol*, 2014, 64(2): 346-351.
- [18] PACESA M, PELEA O, JINEK M. Past, present, and future of CRISPR genome editing technologies[J]. *Cell*, 2024, 187(5): 1076-1100.
- [19] DEWAR A E, BELCHER L J, WEST S A. A phylogenetic approach to comparative genomics[J]. *Nat Rev Genet*, 2025, 26: 395-405.
- [20] BAHJAT N M, YILDIZ M, NADEEM M A, et al. Population structure, genetic diversity, and GWAS analyses with GBS-derived SNPs and silicodart markers unveil genetic potential for breeding and candidate genes for agronomic and root quality traits in an international sugar beet germplasm collection[J]. *BMC Plant Biol*, 2025, 25: 523.
- [21] BAKOVIĆ J, YU B Y K, SILVA D, et al. Redox regulation of the quorum-sensing transcription factor *AgrA* by coenzyme A[J]. *Antioxidants*, 2021, 10(6): 841.
- [22] HURST L D. The K_a/K_s ratio: Diagnosing the form of sequence evolution[J]. *Trends Genet*, 2002, 18(9): 486.
- [23] GONG T, YU J L, YANG H, et al. Prevention of nasal injury in preterm infants during positive pressure ventilation: A systematic review of interface and dressing selection[J]. *BMC pediatrics*, 2025, 25(1): 1-12.
- [24] NIELSEN R. Statistical tests of selective neutrality in the age of genomics [J]. *Heredity*, 2001, 86(Pt 6): 641-647.
- [25] BARNIEH F M, LOADMAN P M, FALCONER R A. Is tumour-expressed aminopeptidase N (APN/CD13) structurally and functionally unique[J]. *Biochim Biophys Acta Rev Cancer*, 2021, 1876(2): 188641.
- [26] TEIXEIRA J C, HUBER C D. The inflated significance of neutral genetic diversity in conservation genetics[J]. *Proc Natl Acad Sci USA*, 2021, 118(10): e2015096118.
- [27] CHENG J, LIU W Q, ZHU X Y, et al. Functional diversity of HemN-like proteins[J]. *ACS Bio Med Chem Au*, 2022, 2: 109-119.
- [28] KIM E, YANG S M, LEE S Y, et al. Classification of *Latilactobacillus sakei* subspecies based on MALDI-TOF MS protein profiles using machine learning models[J]. *Microbiol Spectr*, 2024, 12(10): e0366823.