

木糖葡萄球菌的多位点序列分型分析



陆民锋, 包凯迪, 肖秋杨, 陈晓璇, 徐婷婷, 张振东*

黔南民族师范学院生物科学与农学院, 贵州都匀 558000

摘要: 【目的】为了探究木糖葡萄球菌的亲缘远近和遗传进化关系, 以为木糖葡萄球菌的快速检测和鉴定提供基础。【方法】本论文以下载自 NCBI 的所有木糖葡萄球菌的全部菌株为研究对象, 基于 8 个持家基因 (*arcC*、*bglA*、*dnaJ*、*groL*、*purA*、*pyrD*、*recA* 和 *tpiA*) 创建了一个多位点序列分型 (Multi-locus sequence typing, MLST) 流程, 进行多位点序列分析。【结果】结果表明, 所创建的 MLST 方法能较好的对木糖葡萄球菌进行分型分析, 其中纳入分析的 72 株木糖葡萄球菌可以归类为 53 个 ST 型, 进而被划分为 3 个亚群。另外, 所选的 8 个持家基因中 *arcC* 和 *recA* 显著发生了基因重组。【结论】总之, 建立的多位点分型方法能够用于木糖葡萄球菌的遗传进化分析。

关键词: 木糖葡萄球菌; 多位点序列分型; 持家基因; 群体结构; 重组分析

DOI: [10.57237/j.life.2025.02.001](https://doi.org/10.57237/j.life.2025.02.001)

Multilocus Sequence Typing Analysis of *Staphylococcus xylosus*

Lu Minfeng, Bao Kaidi, Xiao Qiuyang, Chen Xiaoxuan, Xu Tingting, Zhang Zhendong*

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

Abstract: Objective: In order to explore the genetic distance and evolutionary relationship of *Staphylococcus xylosus*, and to provide a basis for rapid detection and identification of *Staphylococcus xylosus*. Method: This paper focuses on all strains of *Staphylococcus xylosus* downloaded from NCBI as the research object. Based on 8 housekeeping genes (*arcC*, *bglA*, *dnaJ*, *groL*, *purA*, *pyrD*, *recA*, and *tpiA*), a multilocus sequence typing (MLST) process was created for Multilocus sequence analysis. Result: The results indicate that the MLST method created can effectively perform typing analysis on *Staphylococcus xylosus*. Among them, the 72 strains of *Staphylococcus xylosus* included in the analysis can be classified into 53 ST types, which can be further divided into 3 subgroups. In addition, among the 8 selected housekeeping genes, *arcC* and *recA* underwent significant genetic recombination. Conclusion: In summary, the established MLST method can be used for genetic evolution analysis of *Staphylococcus xylosus*.

Keywords: *Staphylococcus xylosus*; Multilocus Sequence Typing; Household Genes; Group Structure; Recombination Analysis

基金项目: 贵州省科技厅青年引导项目 (黔科合基础-[2024]青年 054);
黔南民族师范学院支持高层次人才引进研究专项 (qnsyrc202312);
黔南民族师范学院大学生创新创业训练计划项目 (qnsycxcy202574).

*通信作者: 张振东, zg1415@126.com

收稿日期: 2025-11-17; 接受日期: 2025-12-15; 在线出版日期: 2025-12-24

<http://www.lifescitech.org>

1 引言

木糖葡萄球菌 (*Staphylococcus xylosus*) 是一种革兰氏阳性菌，其细胞形态为球形，大多数以聚集方式存在，少数呈散在分布，成对或排列成四联状，广泛分布在自然界中，但主要存在于人和动物的皮肤、消化道和呼吸道中[1]。该种细菌没有荚膜、鞭毛和芽孢，属于兼性厌氧菌，在平板上生长可形成直径为 2mm 左右的圆形、表面湿润光滑但不透明的菌落[2]。有研究表明木糖葡萄球菌在食品添加过程中扮演着重要角色，被广泛用于发酵和制作奶酪等产品中，例如在腊肉、香肠和酸鱼的发酵等食品制备过程中有利于提升肉制品的风味，和改善其品质[3-5]。

当前国内外学者对发酵制品的研究已取得了一定的成果，并广泛应用于发酵制品的工业化生产[6]。同时，越来越多的发酵制品中添加了木糖葡萄球菌，这种趋势使得人们对木糖葡萄球菌的安全性产生了一定的疑虑。因此，对木糖葡萄球菌的遗传特性和变异进行深入研究，成为确保食品安全和消费者健康的一项紧迫任务。而国家为加强对可用于食品添加的菌种的管理，进一步保障消费者权益和规范市场对食品添加的微生物菌种的使用，卫生部发布了《可用于食品的菌种名单》，名单中明确了木糖葡萄球菌可以用于食品[7]。

但在发酵制品中添加微生物非致病性是应用的前提，与之亲缘关系相近、遗传背景相似的腐生葡萄球菌与金黄色葡萄球菌等则是条件致病菌[8]，而木糖葡萄球菌通常被认为是一种非致病菌，但在特定环境下也会引起人类和动物感染，如引起免疫缺陷患者或新生儿不同形式的感染、导致病鼠腿部溃烂和奶牛乳腺炎等[9-11]。因此，建立准确、快速的木糖葡萄球菌鉴定体系对于规范食品行业菌种的管理非常重要。

多位点序列分析 (Multilocus Sequence Typing, MLST) 是一种基于核酸序列测定的细菌分型方法，是从基因分子水平研究菌株的系统发育关系、遗传进化以及种群结构等。通过对所研究菌株的几个持家基因序列进行比较分析，可以建立较为准确可信的物种间关系，因此多位点序列分析可用来评估菌株种群的亲缘远近及遗传进化关系，也可以对微生物进行遗传进化研究[12, 13]。多位点序列分型与传统的表型分型方法相比具有更高的分辨率和重现性，且操作简单快速，研究结果可实现信息共享，是菌株进行种内分型的常用工具[14]。

目前尚无木糖葡萄球菌多位点序列分型的方法，因此，本文拟通过对木糖葡萄球菌基因组的分析，建立一套可用于木糖葡萄球菌多位点序列分型的方法，包括持家基因的选定、引物的设计等，并基于建立的方法对木糖葡萄球菌的遗传进化关系进行分析，以期为木糖葡萄球菌菌株的快速检测和鉴定提供基础。

2 材料与方法

2.1 实验材料

从 NCBI 的 GenBank 数据库直接下载获得已完成全基因组测序的所有木糖葡萄球菌的全部基因组序列和基因组注释信息，去掉重复的菌株后，剩下 53 株菌的基因组信息，详细菌株信息见表 1。

表 1 研究所用的木糖葡萄球菌菌株信息

Table 1 Genome information used for the study

序列号	GCA 编号	来源
NJ	GCA_000338275.1	人
DMB3-Bh1	GCA_000467225.1	鼠
C2a	GCA_000953575.1	未知
NS341	GCA_001476985.1	水稻
22B	GCA_001747725.1	奶酪
CJH_74	GCA_001747735.1	牛奶
BC10	GCA_001747745.1	奶酪
733A	GCA_001748025.1	奶酪
S170	GCA_002078255.1	叶菜
CCM2738	GCA_002732165.1	人
INIFAP009-16	GCA_002836825.1	蜚
SNUC4938	GCA_003041255.1	牛
SNUC5345	GCA_003041265.1	牛
SNUC233	GCA_003041315.1	牛
SNUC4476	GCA_003041485.1	牛
SNUC5423	GCA_003041535.1	牛
SNUC27	GCA_003041555.1	牛
SNUC1397	GCA_003042795.1	牛
SNUC1267	GCA_003043105.1	牛
SNUC196	GCA_003043125.1	牛
SNUC966	GCA_003043145.1	牛
SNUC47	GCA_003043175.1	牛
SNUC5355	GCA_003578725.1	牛
SNUC5173	GCA_003578765.1	牛
SNUC5003	GCA_003578785.1	牛
SNUC3358	GCA_003578815.1	牛
SNUC2618	GCA_003578835.1	牛
SNUC1349	GCA_003578865.1	牛

序列号	GCA 编号	来源
SNUC102	GCA_003578925.1	牛
SNUC5740	GCA_003579145.1	牛
UGA5	GCA_003850005.1	人
WT6	GCA_004571015.1	口腔拭子
FAM20833	GCA_005864455.1	火腿
DE0198	GCA_007677965.1	环境
DMSX03	GCA_014267365.1	发酵豆制品
SIG190	GCA_015059615.1	山羊
J70	GCA_015731745.1	生乳罐
SMG24	GCA_018968045.1	奶牛
2.1523	GCA_020229695.1	香肠
N19	GCA_021189335.1	牛
Ani-LG-156	GCA_021366615.1	奶牛
TMW2.1693	GCA_021403045.1	牛
ABMPRH_121_0	GCA_022791095.1	鼠
MSL3003	GCA_023715825.1	洁净室
TMW2.1324	GCA_024172485.1	香肠
TMW2.1602	GCA_024172565.1	香肠
DSM28566	GCA_024621885.1	鼠
ITEM18282	GCA_029105225.1	人
47-83	GCA_900098615.1	牛奶
MGYG-HGUT-00011	GCA_902362245.1	人肠道
MGYG-HGUT-00006	GCA_902362295.1	人肠道
A8_8.SX.CDM	GCA_910574145.1	鼠
SRR18439529_bin.27	GCA_947425635.1	鼠

2.2 实验方法

2.2.1 基因组信息注释

由于 NCBI 数据库上的基因组注释信息，可能使用了不同的软件，为消除差异，本论文中统一对下载的基因组信息使用 prokav1.14 [15] (<https://github.com/tseemann/prokka>) 重新注释，找出所下载基因组的基因编码区域及其注释的功能信息。

2.2.2 持家基因序列提取

用 TBtoolsv1.6 软件按照基因组注释文件中的基因位置信息，从下载的 72 株木糖葡萄糖球菌基因组中提取所有持家基因序列，根据这些持家基因在基因组上的分布情况，选择其中适合用来进行多位点序列分型的 8 个基因，要求：①所下载的绝大多数基因组中含有这些持家基因；②这些持家基因在基因组中基本上均匀分布；③选择的持家基因含有保守区、适用于引物设计，且扩增区域符合 MLST 要求。

根据以上标准，选取的 8 个持家基因 *arcC*、*bglA*、*dnaJ*、*groL*、*purA*、*pyrD*、*recA* 和 *tpiA* 作为 MLST 分型的研究基因（见图 1）。

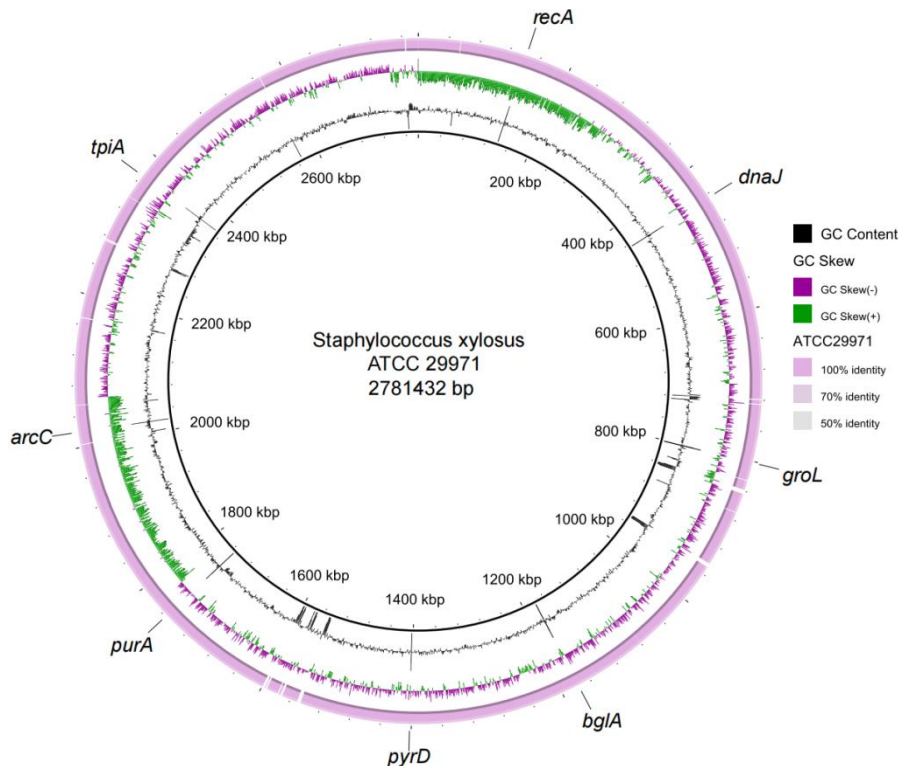


图 1 基于木糖葡萄糖球菌的 BRIG 分析

Figure 1 BRIG analysis based on *Staphylococcus xylosus*

2.2.3 引物设计

将提取的 72 个木糖葡萄糖球菌基因组中的 8 个持家基因使用 mafft7.311 [16] (<https://mafft.cbrc.jp/alignment/server/>) 比对, 获得这些基因的保守区, 使用 Primerv5.0 对持家基因进行引物设计, 引物的详细信息见表 2。

表 2 关于木葡萄球菌用于 MLST 的基因和引物的信息

Table 2 The information on the genes and primers used for MLST

基因	蛋白质	PCR 引物	引物序列(5'-3')	产物大小(bp)	退火温度(℃)
arcC	氨基甲酸酯激酶 1	arcCF	CAACCWAAGCAAGAAGC	802	58
		arcCR	GCAAACCTGAATTGCYGC		
bglA	芳基磷酸-β-D-葡萄糖苷酶 BglA	bglAF	TTYTTCTTCCCTGATGT	533	54
		bglAR	CGTTTYTTCATTTACACC		
dnaJ	伴侣蛋白 DnaJ	dnaJF	GTGGCCAAAAGAGAT	735	56
		dnaJR	CACCTTCACCAGCTA		
groL	60kDa 伴侣蛋白	groLF	GCACGTCAATCWATG	823	54
		groLR	GCTTACGACGATCAC		
purA	腺苷酸盐合成酶	purAF	CAATGGGGAGACGAA	904	53
		purAR	GAATCGAACCAACC		
pyrD	二氢酯脱氢酶	pyrDF	ATRCGCGTTATGAAG	474	52
		pyrDR	GAAGCMCCACTACGTA		
recA	重组酶 A	recAF	CTTTACCTTGYCCCA	750	54
		recAR	GTAGGTGGTTACCCTAA		
tpiA	三磷酸异构酶	tpiAF	TTAAGYGATGCGCCRC	706	60
		tpiAR	ATCGCAGGTAAC TGGA		

2.2.4 持家基因序列多样性分析

利用 DnaSPv6.0 [17]计算 8 个木糖葡萄球菌持家基因的等位基因数量、多态性位点、G+C 含量、核苷酸多样性 (π), 并通过判断 Tajima'sD 值以检验木糖葡萄球菌的持家基因在进化过程中是否符合中性理论模型。

2.2.5 多位点序列分型分析和构建最小生成树

使用 BioNumericsv8.0 软件对木糖葡萄球菌的 8 个持家基因序列进行多位点序列分型分析, 以此确定 53 株木糖葡萄球菌的 ST 型。为了确保结果的准确性和可靠性, 采用了 UPGMA (Unweighted Pair Group Method with Arithmetic Mean) 法以及分类相似系数对等位基因序列进行聚类分析, 为了更直观地展示聚类结果, 利用 Prim's 算法构建了最小生成树(minimum spanning tree, MSTree), 并基于算法构建最小生成树获得各 ST 型与不同分离来源间的关系[18]。

2.2.6 正选择分析

使用 PAMLv4.9 [19]软件中的 codeml 工具, 运用枝位点特异模型(branch-sitemodel)来检测该基因家族

在该物种中是否受到正选择, 即使用 PAML 软件中两种假设的似然比来检验是否存在正选择[20]。

通过计算木糖葡萄球菌 8 个持家基因的同义替换率 (K_s) 与非同义替换率 (K_a), 并根据 K_a/K_s 比值分析木糖葡萄球菌基因的选择压力, 当 $K_a/K_s=1$ 表示中性选择, 表明基因基本上没有受到选择的作用; 当 $K_a/K_s>1$ 表示正向选择, 表明基因受到积极的适应性选择; 当 $K_a/K_s<1$ 表示纯化选择, 表明基因受到保守选择的约束。

2.2.7 系统发育分析

串联 MLST 木糖葡萄球菌的 8 个持家基因: arcC、bglA、dnaJ、groL、purA、pyrD、recA 和 tpiA。利用 MEGA6.0 软件, 采用邻接法 (neighbor-joining, NJ) 构建系统进化树, bootstrap 设置为 1000, 进行系统发育分析[21]。

2.2.8 持家基因重组分裂分析

采用 SplitsTreev4.0 软件对木糖葡萄球菌 8 个持家基因的串联序列构建重组分裂分析图(split decomposition analysis), 并用该软件附带的 phi-test 检测等位基因内是否存在显著的重组现象。

2.2.9 群体遗传结构分析和主成分分析

使用软件 PGDspiderv2.1.1.5 将持家基因串联产生的序列文件进行 SNP 分析，产生 VCF 文件。利用 Admixturev1.3.0 软件基于生成的 vcf 文件进行群体遗传结构分析[22]。同样的，使用软件 plinkv1.9 基于生成的 vcf 文件进行主成分分析（principal component analysis, PCA）[23]。

2.2.10 重组分析

将 72 株木糖葡萄球菌的 8 个持家基因经多序列比对过串联起来的序列文件和 2.2.7 得到进化树文件，作为重组分析的输入文件，使用软件 Gubbins

（<https://github.com/nickjcroucher/gubbins>）进行基因组分析参数使用默认值，并使用 Splits Treev4.0 软件的 phi-test 检测等位基因内是否存在重组现象。

3 结果与分析

3.1 持家基因多样性分析

使用 DnaSPv6.0 软件对分离到的 53 株木糖葡萄球菌的 8 个持家基因(*arcC*、*bglA*、*dnaJ*、*groL*、*purA*、*pyrD*、*recA* 和 *tpiA*)进行计算，得到其基因的大小、等位基因的数量、多态性位点、G+C 含量百分比、Tajima'sD 值、 π （核苷酸多样性）和 Ka/Ks 等的计算结果见表 3。

表 3 木糖葡萄球菌的持家基因序列多样性分析

Table 3 Diversity analysis of housekeeping genes sequences in *Staphylococcus xyloso*

基因	大小/bp	等位基因数量	多态性位点	G+C 含量%	Tajima'sD	Phitest (P-value)	π (persite)	Ka/Ks
<i>arcC</i>	805	25	78	37.20	-1.09	0.11 (0.01)	0.01	0.06
<i>bglA</i>	534	23	70	37.50	-0.70	0.07 (0.96)	0.02	0.10
<i>dnaJ</i>	734	24	77	38.90	1.02	0.13 (0.72)	0.03	0.07
<i>groL</i>	824	20	85	35.80	-1.04	0.12 (0.29)	0.02	0.01
<i>purA</i>	905	24	101	38.20	-1.46	0.16 (0.16)	0.01	0.04
<i>pyrD</i>	475	23	59	33.10	-0.26	0.07 (0.48)	0.03	0.11
<i>recA</i>	751	18	64	40.10	-0.18	0.10 (0.01)	0.02	0.03
<i>tpiA</i>	707	29	54	37.80	-0.82	0.11 (0.59)	0.01	0.07
总	5735	53	567	37.50	-0.56	0.15 (0.00)	0.02	-

注：Tajima'sD 统计是检测和鉴定基因在进化过程中是否符合中性理论模型，即是否受到选择压力作用。如果 Tajima'sD 值小于 0，表明检测的基因受到定向选择作用，而该值大于 0，表明受到多样化选择作用[24]。 π 值指的是单个核苷酸变异产生的基因水平差异。Ka/Ks 值是判断生物在进化过程中基因是否受到环境因素的选择压力作用。

Note: Tajima's *D* statistic is employed to test for departures from the neutral evolution model in genes, thereby assessing the action of selective pressures during evolution. A Tajima's *D* value significantly less than zero indicates that the gene has been subjected to directional selection, whereas a value significantly greater than zero suggests diversifying selection [24]. The π value (pi) represents the nucleotide diversity within a population, quantifying the average pairwise nucleotide differences per site. The Ka/Ks ratio (or dN/dS ratio) is utilized to infer the presence and type of natural selection acting on protein-coding genes during evolution; a ratio significantly deviating from 1 indicates the action of selective pressure.

对测序结果的分析表明，用作 MLST 分析的 8 个持家基因的长度为 475 (*purD*) ~ 905bp (*purA*)，合并后序列总长度为 5735bp。等位基因数量的多少可反映基因多样性的强弱，等位基因数量越多，则表示基因的多样性越丰富。木糖葡萄球菌 8 个持家基因的等位基因数量为 18~29 个，共有 53 个等位基因。其中 *recA* 基因的等位基因数量最少，为 18 个，说明该基因片段的多样性最低。*tpiA* 基因的等位基因数量最多，为 29 个，说明该基因片段的多样性最高。木糖葡萄球菌的多态性位点为 54~101 个，其中 *tpiA* 基因的多态性位点最少为 54 个，*purA* 基因的多态性位点最多为 101 个，

而 8 个持家基因共有 567 个多态性位点。G+C 含量是细菌遗传进化研究中的重要指标，其中木糖葡萄球菌的 *purD* 基因 G+C 含量最低，为 33.10%，基因 *recA* 的 G+C 含量最高，为 40.1%，8 个持家基因串联的总 G+C 含量为 37.5%。

在本次研究中有 7 个持家基因(*arcC*、*bglA*、*groL*、*purA*、*pyrD*、*recA* 和 *tpiA*)的 Tajima'sD 值小于 0，表明它们受到定向选择作用，而只有基因 *dnaJ* 大于 0，表明它们受到多样化选择作用。说明木糖葡萄球菌基因序列在进化过程中大部分受到定向选择作用，极少部分受到多样化选择的作用。8 个持家基因的 π 值在

0.01 (*arcC*、*purA* 和 *tpiA*) 到 0.03 (*dnaJ* 和 *recA*) 之间, 说明木糖葡萄球菌的单个核苷酸变异产生的基因水平差异较小。木糖葡萄球菌 8 个持家基因的 Ka/Ks 值在 0.0098 (*groL*) ~ 0.1115 (*purD*)之间, 均小于 1, 说明木糖葡萄球菌的基因基本上未受到环境因素的选择压力作用。

3.2 多位点序列分型分析

依据等位基因位点差异将 8 个木糖葡萄球菌菌株的等位基因按 *recA*、*dnaJ*、*groL*、*bglA*、*pyrD*、*purA*、*arcC* 和 *tpiA* 的顺序进行组合, 得到等位基因图谱, 再对图谱编号, 作为菌株的 ST 型, 木糖葡萄球菌菌株的

分型结果见表 4。由表 4 可知, 72 株木糖葡萄球菌共分成 53 个 ST 型, 其中 ST-10 是优势序列型, 含有 8 株菌, 分别分离来源自牛、火腿、奶酪、香肠、蜚和洁净室等不同的生存环境; ST-7 和 ST-32 各包含有 3 株菌, 其中 ST-7 分离来源自香肠和奶酪, ST-32 分离来源自人和鼠; ST-4、ST-6、ST-13、ST-15、ST-23、ST-28、ST-34 和 ST-39, 各包含 2 株菌株; 剩余的 42 个 ST 型都只含有单一菌株。

以上结果表明, 统一 ST 型的木糖葡萄球菌往往来源于不同的环境, 仅此木糖葡萄球菌的多位点序列分型并不具有环境特异性。

表 4 基于 8 个持家基因的木糖葡萄球菌 ST 型及等位基因图谱

Table 4 STs and allelic profiles of *Staphylococcus* based on eight house-keeping genes

菌株编号	序列号	ST 型	克隆复合体	分离来源	等位基因编号							
					<i>arcC</i>	<i>bglA</i>	<i>dnaJ</i>	<i>groL</i>	<i>purA</i>	<i>pyrD</i>	<i>recA</i>	<i>tpiA</i>
NJ	GCA_000338275.1	1	singleton	Human	1	1	1	1	1	1	1	1
DMB3-Bh1	GCA_000467225.1	2	Cluster1	mouse	2	2	2	2	2	2	2	2
C2a	GCA_000953575.1	3	singleton	UN	3	3	3	3	3	3	3	3
NS341	GCA_001476985.1	4	Cluster1	plant	2	2	4	4	2	4	2	2
22B	GCA_001747725.1	5	Cluster1	Cheese	2	4	5	2	2	4	2	2
CJH_74	GCA_001747735.1	6	singleton	milk	4	5	6	3	3	3	3	3
BC10	GCA_001747745.1	7	singleton	Cheese	5	6	6	5	3	3	3	3
733A	GCA_001748025.1	8	singleton	Cheese	6	5	6	6	4	5	4	3
S170	GCA_002078255.1	9	Cluster1	plant	2	2	2	2	2	4	2	2
CCM2738	GCA_002732165.1	10	Cluster1	Human	2	2	5	2	2	4	2	2
INIFAP009-16	GCA_002836825.1	11	singleton	microplus	7	3	7	7	5	6	5	4
SNUC4938	GCA_003041255.1	12	singleton	bovine	8	7	8	8	6	7	3	5
SNUC5345	GCA_003041265.1	13	singleton	bovine	9	8	9	3	3	8	3	6
SNUC3232	GCA_003041275.1	6	singleton	bovine	4	5	6	3	3	3	3	3
SNUC233	GCA_003041315.1	14	singleton	bovine	10	9	7	9	7	9	6	7
SNUC4476	GCA_003041485.1	15	singleton	bovine	10	3	7	9	7	9	6	8
SNUC2570	GCA_003041515.1	15	singleton	bovine	10	3	7	9	7	9	6	8
SNUC5423	GCA_003041535.1	16	singleton	bovine	11	10	10	10	8	10	7	9
SNUC27	GCA_003041555.1	17	singleton	bovine	9	11	9	8	9	11	8	10
SNUC1397	GCA_003042795.1	18	singleton	bovine	12	3	7	9	10	9	6	8
SNUC1267	GCA_003043105.1	19	singleton	bovine	13	12	11	11	11	12	9	11
SNUC196	GCA_003043125.1	20	singleton	bovine	14	13	12	12	12	13	10	12
SNUC966	GCA_003043145.1	21	singleton	bovine	4	5	6	3	3	14	3	3
SNUC47	GCA_003043175.1	22	singleton	bovine	15	14	1	13	13	1	1	13
SNUC5355	GCA_003578725.1	23	singleton	bovine	4	5	6	7	4	3	3	3
SNUC5173	GCA_003578765.1	24	singleton	bovine	16	8	3	8	14	15	11	5
SNUC5003	GCA_003578785.1	25	singleton	bovine	17	15	13	14	15	16	12	14
SNUC3358	GCA_003578815.1	26	singleton	bovine	6	5	6	7	3	17	3	3
SNUC2618	GCA_003578835.1	27	singleton	bovine	9	11	14	15	14	5	3	15
SNUC1349	GCA_003578865.1	28	singleton	bovine	18	3	7	9	16	6	6	8
SNUC237	GCA_003578895.1	10	Cluster1	bovine	2	2	5	2	2	4	2	2
SNUC102	GCA_003578925.1	29	singleton	bovine	15	14	15	13	17	1	1	16
SNUC4735	GCA_003579115.1	28	singleton	bovine	18	3	7	9	16	6	6	8

菌株编号	序列号	ST 型	克隆复合体	分离来源	等位基因编号							
					<i>arcC</i>	<i>bglA</i>	<i>dnaJ</i>	<i>groL</i>	<i>purA</i>	<i>pyrD</i>	<i>recA</i>	<i>tpiA</i>
SNUC5740	GCA_003579145.1	30	singleton	bovine	9	8	9	16	4	8	3	17
UGA5	GCA_003850005.1	31	singleton	human	14	16	12	12	12	13	10	18
WT6	GCA_004571015.1	32	Cluster1	human	2	2	16	2	2	4	2	2
FAM20833	GCA_005864455.1	33	singleton	ham	9	8	9	15	14	8	3	15
DE0198	GCA_007677965.1	34	singleton	UN	10	3	7	9	18	18	6	19
2	GCA_007814625.1	13	singleton	cow	9	8	9	3	3	8	3	6
DMSX03	GCA_014267365.1	35	Cluster2	Meju	2	17	17	2	2	4	2	2
SIG190	GCA_015059615.1	36	singleton	Goat	9	8	3	8	14	15	13	5
19428wA1_WT6	GCA_015235475.1	32	Cluster1	mouse	2	2	16	2	2	4	2	2
J70	GCA_015731745.1	37	singleton	milk	19	5	6	7	3	3	3	3
RIT806	GCA_016892565.1	23	singleton	plant	4	5	6	7	4	3	3	3
G19-4C	GCA_017583125.1	10	Cluster1	environment	2	2	5	2	2	4	2	2
SMG24	GCA_018968045.1	38	singleton	cow	20	18	18	2	2	19	14	6
K19	GCA_019148995.1	10	Cluster1	bovine	2	2	5	2	2	4	2	2
K46	GCA_019334165.1	10	Cluster1	bovine	2	2	5	2	2	4	2	2
2.1523	GCA_020229695.1	39	singleton	sausage	21	19	6	7	19	20	6	8
N24	GCA_021189235.1	4	Cluster1	bovine	2	2	4	4	2	4	2	2
N19	GCA_021189335.1	40	Cluster2	bovine	2	2	17	2	2	2	2	2
Ani-LG-156	GCA_021366615.1	41	singleton	cow	9	20	9	3	6	11	11	20
TMW2.1704	GCA_021402985.1	34	singleton	bovine	10	3	7	9	18	18	6	19
TMW2.1780	GCA_021403025.1	7	singleton	sausage	5	6	6	5	3	3	3	3
TMW2.1693	GCA_021403045.1	42	singleton	bovine	9	8	9	7	20	8	3	21
ABMPRH_121_0	GCA_022791095.1	43	singleton	mouse	22	21	19	17	21	21	15	22
MSL3003	GCA_023715825.1	44	singleton	environment	10	3	6	7	19	6	6	3
TCD16	GCA_023716225.1	32	Cluster1	mouse	2	2	16	2	2	4	2	2
TMW2.1324	GCA_024172485.1	45	singleton	sausage	6	3	20	6	3	3	3	23
TMW2.1521	GCA_024172505.1	7	singleton	sausage	5	6	6	5	3	3	3	3
TMW2.1602	GCA_024172565.1	46	Cluster1	sausage	2	2	21	2	2	4	2	24
TMW2.1523	GCA_024442095.1	39	singleton	sausage	21	19	6	7	19	20	6	8
DSM28566	GCA_024621885.1	47	Cluster1	mouse	2	2	5	2	2	4	2	25
ITEM18282	GCA_029105225.1	48	singleton	human	2	2	22	2	2	4	2	2
47-83	GCA_900098615.1	49	singleton	milk	23	22	23	18	22	22	16	26
ATCC29971	GCA_900240095.1	10	Cluster1	sausage	2	2	5	2	2	4	2	2
NCTC11043	GCA_900458755.1	10	Cluster1	UN	2	2	5	2	2	4	2	2
MGYG-HGUT-00011	GCA_902362245.1	50	singleton	human	24	23	24	19	23	1	17	1
MGYG-HGUT-00006	GCA_902362295.1	51	singleton	human	25	8	9	8	14	15	13	27
A8_8.SX.CDM	GCA_910574145.1	52	singleton	mouse	22	21	19	17	24	21	15	28
IM1419	GCA_925279085.1	10	Cluster1	UN	2	2	5	2	2	4	2	2
SRR18439529_bin.27	GCA_947425635.1	53	singleton	mouse	22	21	19	20	24	23	18	29

3.3 最小生成树

为了进一步分析已发布的 72 株木糖葡萄球菌的 ST 型的相互关系, 通过最小生成树可以清晰地看出各个聚类群组之间的亲缘关系和进化路径。从图上整体看, 72 株木糖葡萄球菌的最小生成树可分为 A、B 两个分支, 其菌株的主要分离来源有牛、鼠和人。其中

分离来源自牛的菌株分布最为广泛, A 分支和 B 分支都有它的存在; 而分离来源自鼠的菌株主要集中在 A 分支, B 分支中只存在分离来源自牛和人的菌株, 但在 A 分支中同样也存在少量分离来自人的菌株。由此表明, 木糖葡萄球菌的 ST 型不具有宿主特异性。

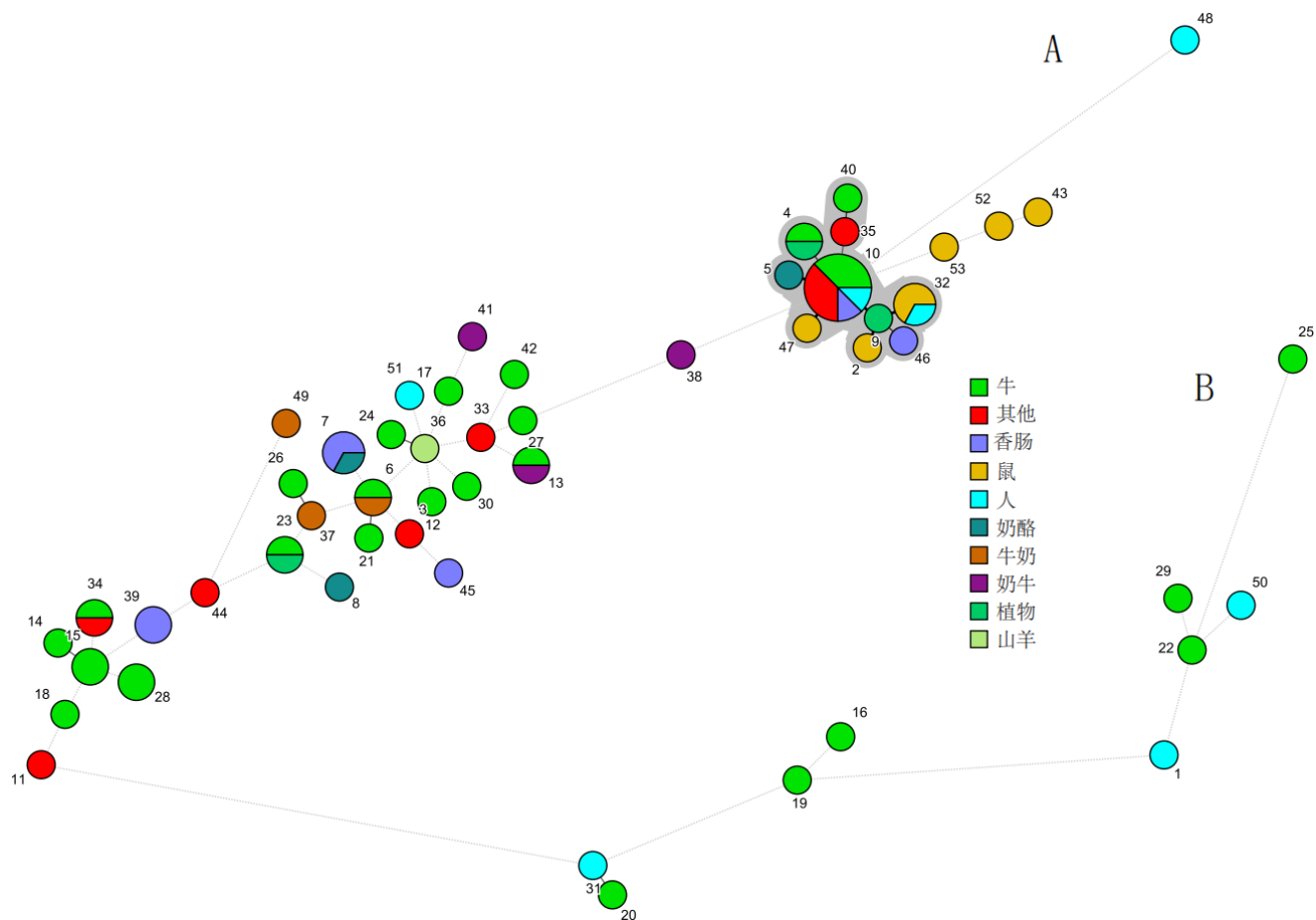


图 2 木糖葡萄球菌最小生成树

Figure 2 Minimum spanning tree of *Staphylococcus xylosus*

注：ST 型节点的颜色表示木糖葡萄球菌的不同分离来源，圆圈的大小则表示该 ST 型所包含的菌株数量。

Note: Node color denotes the isolation source of *Staphylococcus xylosus* strains belonging to different Sequence Types (STs), while node size is scaled proportionally to the number of strains comprising each ST.

3.4 系统发育分析

本研究基于 8 个持家基因串联序列，利用邻接法 (neighbor-joining, NJ) 绘制系统发育树，得到图 3 中的进化树。如图 3 所示，木糖葡萄球菌的 53 个 ST 型可通过系统进化树分成三个分支：ClusterI、ClusterII 和

ClusterIII，其中 ClusterI 含有的 ST 型最多，ST 型的分离来源最多样，多数菌株分离来自牛。在 ClusterII 中的 ST43、ST52 和 ST53 都分离来源于鼠。而在 ClusterIII 中的所有 ST 型都分离来源自牛和人。

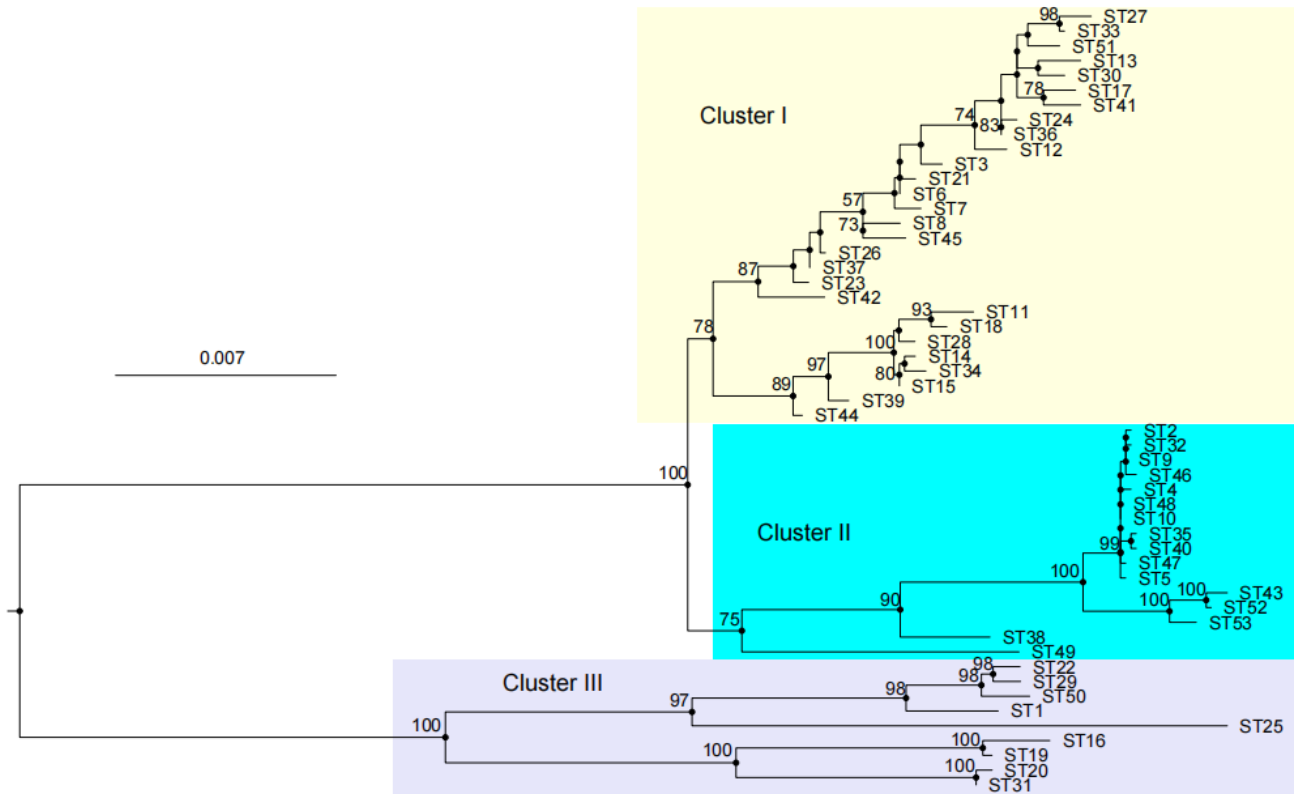


图 3 基于木糖葡萄球菌 8 个持家基因串联序列的系统发育树

Figure 3 NJtreebasedontandemsequencesof8housekeepinggenesinStaphylococcusxylosus

注：邻接法是一种常用的分子系统发生学方法，它通过衡量物种之间的遗传差异来构建进化树，以揭示它们之间的亲缘关系。步长值（bootstrap）设置为 1000 意味着在构建进化树时，会进行 1000 次随机重抽样来评估树节点的可靠性，从而提高树的准确性和可信度。

Note: The Neighbor-Joining method is a commonly used phylogenetic inference method in molecular systematics. It reconstructs evolutionary trees (phylogenies) by measuring the genetic divergence between species, thereby elucidating their evolutionary relationships. A bootstrap value of 1000 indicates that 1000 replicates of random resampling were performed during tree reconstruction to assess the topological robustness (reliability) of tree nodes, thereby enhancing the accuracy and confidence of the inferred phylogeny.

3.5 持家基因重组分裂分析

在细菌中，高重组率导致了基因组多样化的加速，进一步影响了菌株间的基因组差异。同源重组是造成序列数据中核苷酸位点差异的原因。通过分裂重组分析可以识别到这些差异，并利用网络图来表示它们之间的关系。

使用 SplitTreev4.0 软件构建了 8 个木糖葡萄球菌持家基因的串联序列的重组分析图，若图中出明显的平行四边形结构，则说明该些基因在进化过程中发生了明显的基因重组，反之无四边行结构说明在进化过

程中没有经历基因重组或重组现象不显著[25, 26]。如图 4 所示，图中出现了明显的平行四边形结构，说明在遗传进化过程中木糖葡萄球菌发生了重组与进化。运行 phi-test 进行计算，发现 *arcC* 和 *recA* 基因重组事件发生的 *P* 值（表 3）小于 0.05，说明木糖葡萄球菌这两个基因在遗传进化过程中显著的发生了基因重组事件。另外，从图 4 可知，在 8 个持家基因的分裂重组图中，53 个 ST 型可以划分成 3 个谱系，与图 3 结果一致。

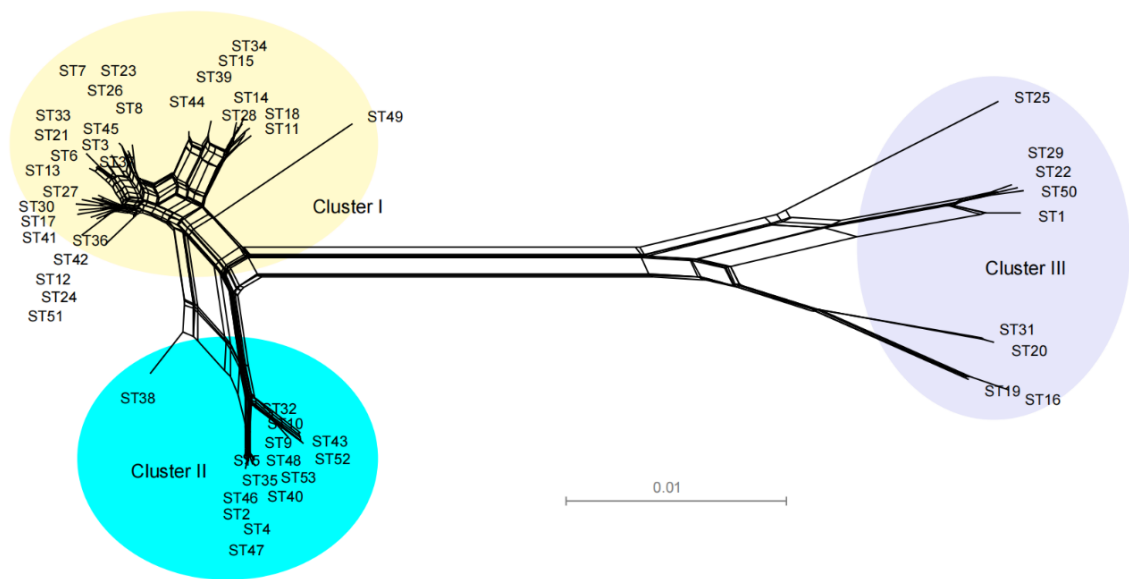


图 4 基于木糖葡萄糖菌 8 个持家基因串联序列的重组分裂图

Figure 4 Splitting and recombination diagram based on the tandem sequences of 8 housekeeping genes in *Staphylococcus xylosus*

3.6 群体遗传结构分析

群体遗传结构是指遗传变异在物种或种群中的一种非随机分布。对于同一亚群来说不同个体的亲缘关系相对较高，而亚群与亚群之间的亲缘关系相对较远 [27]。如图所示，通过颜色可以明显判断 53 个样本可

划分为 3 个亚群体，其中 Lineage1 含有 28 个样本，Lineage2 含有 16 个样本以及 Lineage3 含有 9 个样本。当 K=2 时，将较大的类群划分为两个独立的亚群。当 K=3 时，结果与 8 个持家基因序列串联的系统进化树以及重组分裂图的结果一致。其中，

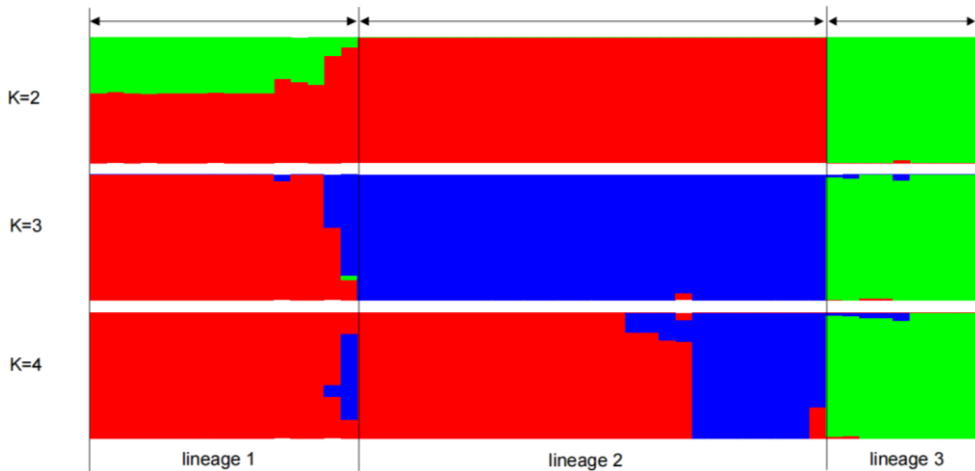


图 5 基于木糖葡萄糖菌 8 个持家基因串联序列的群体结构分析图

Figure 5 Population structure analysis based on tandem sequences of 8 house-keeping genes in *Staphylococcus xylosus*

注：在群体结构分析图中每个直方柱子分别代表 1 个样本(共 72 个菌株，53 个 ST 型样本)，柱子的颜色以及颜色比例则表示这个样本所属的亚群构成比例。

Note: In the population structure analysis plot, each vertical bar represents an individual sample (comprising a total of 72 strains, among which 53 distinct Sequence Types (STs) were identified). The color composition of each bar visually depicts the proportional contribution of genetic ancestry components (subpopulations) assigned to that sample.

3.7 主成分分析

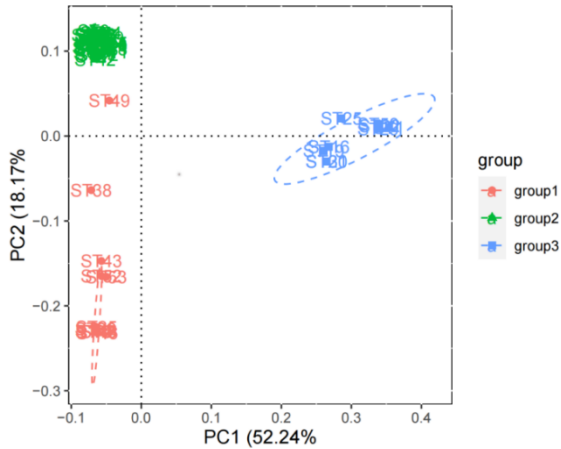


图 6 基于木糖葡萄球菌 8 个持家基因串联序列的 PCA 分析图

Figure 6 PCA analysis based on the tandem sequences of 8 house-keeping genes in *Staphylococcus xylosus*

3.8 重组分析

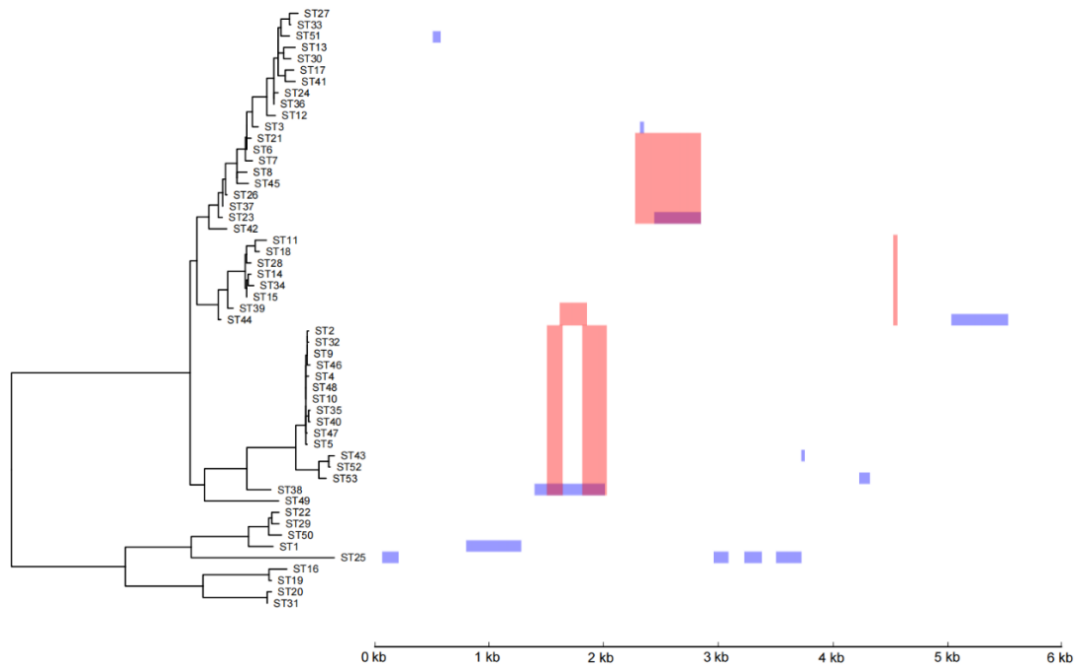


图 7 基于木糖葡萄球菌 8 个持家基因的重组分析

Figure 7 Recombinant analysis of eight house-keeping genes in *Staphylococcus xylosus*

注：每行代表系统发育中的一个 ST 型。红色方块表示进化树内部分支上发生的预测重组，说明是由多个分离物共同享有，蓝色方块代表在终端分支上发生的重组，说明这个是单个分离物所独有的。

Note: Each row in the figure corresponds to a distinct Sequence Type (ST) within the phylogeny. Red squares depict predicted recombination events occurring on internal branches of the phylogenetic tree, indicating ancestry shared by multiple isolates. Blue squares denote recombination events located on terminal branches, representing genetic variation unique to a single isolate.

Gubbins 是一种基于假定点突变的算法,用于构建系统发育,迭代地识别具有高密度碱基替换的位点,并将其标记为重组[28]。结合图 7 和表 2 可以明显发现,在木糖葡萄球菌的 8 个持家基因的串联序列片段上,共检测到了 17 次重组事件。Gubbins 结果显示,进化树 ClusterI 和 Cluster2 都具有多个红色块和多个蓝色块,而 ClusterIII 只存在蓝色块,表明预测到有多个重组事件发生。另外,在 8 个持家基因的串联序列片段上,检测到了 2 个频繁发生基因重组的区域,对应了基因 *arcC* 和 *recA*,该结果与 phitest 结果一致,更加证实了木糖葡萄球菌的持家基因 *arcC* 和 *recA* 在进化过程中发生了基因重组。

4 讨论与结论

木糖葡萄球菌是一种革兰氏阳性菌,无荚膜和鞭毛,不产芽孢,广泛存在于腊肉、香肠和酸鱼的发酵等食品[3-5],并被广泛应用于发酵制品的生产[6]。为区分木糖葡萄球菌与亲缘关系较近,而不能用于食品的葡萄球菌属菌种,建立了一种多位点分型方法,并基于该方法对木糖葡萄球菌进行了多位点分型与遗传进化分析。综合来看,建立的 MLST 分型方法能用于木糖葡萄球菌的分型分析,具有一定的分辨力。

本文结论如下:

- (1) 通过对以公布基因组的 72 株木糖葡萄球菌进行多位点分型分析,发现它们可以分为 53 个 ST 型。
- (2) 基于 MLST 的群体结构分析与系统发育分析能将这 72 株木糖葡萄球菌都可以形成 3 个亚群;
- (3) 纳入 MLST 方法的 8 个持家基因中, *arcC* 基因和 *recA* 基因显著发生了基因重组。

参考文献

- [1] 贾士芳,陈美玲,钟瑾,等.肉制品发酵剂木糖葡萄球菌 I2 的研究 [J].食品与发酵工业,2006(08): 72-74.
- [2] 周宝青.致病葡萄球菌分子靶标挖掘平台构建及基于 CRISPR/Cas 快速检测新技术研究 [D].广东:华南理工大学,2022.
- [3] 文瑜,张驰,杨思艺等.木糖葡萄球菌对川式腊肉产品特性的影响 [J].食品研究与开发,2023,44(13): 62-69.

- [4] 慕婷婷,吴建军,荣良燕等.副干酪乳杆菌和木糖葡萄球菌对发酵香肠品质和风味的影响 [J].食品与发酵工业,2023,49(15): 129-135.
- [5] 孙海鑫,陈力,石文琪等.自然发酵酸鱼中微生物群落多样性与理化性质分析 [J].食品研究与开发,2023,44(12): 187-195.
- [6] 邢巍,刘贯勇,牛保坤,等.微生物对肉制品发酵的影响研究进展 [J].肉类研究,2023,37(04): 61-68.
- [7] 姚粟,于学健,白飞荣,等.中国传统发酵食品用微生物菌种名单的研究 [J].食品与发酵工业,2017,43(09): 238-258.
- [8] 叶圣雨,吴佳佳,屈炯,等.凝固酶阴性葡萄球菌在发酵食品中的应用 [J].中国食品学报,1-11.
- [9] 李虹瑾,唐玉娇,刘嘉琳,等.犬源皮氏葡萄球菌的分离鉴定和耐药性分析 [J].中国兽医杂志,2022,58(11): 49-53.
- [10] 汪斌,王倩.利用实验室病原微生物抽检控制实验动物质量 [J].中南农业科技,2023,44(09): 113-115.
- [11] 卢艳华,张忠斌,陈金清等.大环内酯类药物对木糖葡萄球菌敏感菌株及其耐药菌株生物被膜形成的影响 [J].黑龙江畜牧兽医,2023(11): 1-7.
- [12] 王中强,邱少富,王勇等.多位点序列分型技术及其研究进展 [J].军事医学科学院院刊,2010,34(01): 76-79.
- [13] El Mazini S, Barhoumi M, Mhaidi I, et al. Genetic Diversity and Population Structure of Leishmania infantum in Morocco as Revealed by Multilocus Sequence Typing (MLST) Approach [J]. Pathogens, 2023, 12(6): 785.
- [14] 李明雨.传统发酵乳制品中植物乳杆菌基因多样性研究 [D].黑龙江:东北农业大学,2018.
- [15] Seemann T. Prokka: rapid prokaryotic genome annotation [J]. Bioinformatics, 2014, 30(14): 2068-2069.
- [16] Kazutaka K, M D S. MAFFT: iterative refinement and additional methods. [J]. Methods in molecular biology (Clifton, N.J.), 2014, 1079: 131-46.
- [17] Rozas J, Ferrer-Mata A, Sánchez-DelBarrio J C, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets [J]. Molecular biology and evolution, 2017, 34(12): 3299-3302.
- [18] 金之烜,王平,张雪玲等.基于多位点序列分型对新疆伊宁学龄儿童肠道两歧双歧杆菌 (*Bifidobacterium bifidum*) 群体遗传差异解析 [J].微生物学通报,2023,50(12): 5487-5504.
- [19] Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood [J]. Computer applications in the biosciences, 1997, 13(5): 555-556.

- [20] 许吾琴. 枫香树和缺萼枫香树的物种形成与适应性进化 [D]. 浙江: 浙江大学, 2023.
- [21] 杜佳庚, 周瑾, 李晓燕, 等. 芹菜 APX 基因家族鉴定及其表达分析 [J]. 干旱地区农业研究, 2023, 41(06): 54-63.
- [22] 王爻. 霸王复合体 (蒺藜科) 的群体遗传学和物种界定研究 [D]. 陕西: 西北农林科技大学, 2023.
- [23] 黎巷汝. 基于简化基因组 SNP 标记的茶树气候适应性分化研究 [D]. 福建: 福建农林大学, 2023.
- [24] 张振东, 王玉荣, 向凡舒等. 老年人肠道乳酸菌的分离与鉴定及其来源发酵乳杆菌的遗传进化分析 [J]. 食品科学, 2021, 42(22): 105-112.
- [25] 余中节, 赵洁, 宋宇琴等. 罗伊氏乳杆菌肠道分离株的遗传多样性研究 [J]. 微生物学报, 2019, 59(09): 1786-1797.
- [26] 王玉荣. 酸粥细菌多样性及其风味品质形成机制研究 [D]. 内蒙古: 内蒙古农业大学, 2021.
- [27] 闫锋. 无芒雀麦种质资源遗传多样性研究 [D]. 新疆: 新疆农业大学, 2021.
- [28] Croucher N J, Page A J, Connor T R, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins [J]. Nucleic acids research, 2015, 43(3): e15-e15.