



短杆菌属海洋环境适应机制的泛基因组学

裴盛祥, 黄小云, 牛四文, 谢富全, 张改云*

自然资源部第三海洋研究所海洋生物遗传资源重点实验室, 福建 厦门 361005

裴盛祥, 黄小云, 牛四文, 谢富全, 张改云. 短杆菌属海洋环境适应机制的泛基因组学. 微生物学报, 2022, 62(1): 145–159.

Pei Shengxiang, Huang Xiaoyun, Niu Siwen, Xie Fuquan, Zhang Gaiyun. Adaptation mechanism of *Brevibacterium* species to the marine environment based on pan-genomics analysis. *Acta Microbiologica Sinica*, 2022, 62(1): 145–159.

摘要: 【目的】为了探究短杆菌属对海洋环境的适应机制。【方法】本研究通过对 6 株分离自不同洋区、属于不同分类单元的短杆菌菌株进行测序、拼接和注释, 结合 23 株从美国国家生物技术信息中心(NCBI)下载的短杆菌属模式菌株及非模式菌株的基因组数据, 进行泛基因组学分析和物种进化分析。【结果】泛基因组学分析表明短杆菌属具有开放型泛基因组, 这与该属生存环境多样性特征相符合。海洋来源的短杆菌与其他生境来源的短杆菌在基因组水平上表现出明显的差异性, 主要是进化过程中的基因家族扩增收缩、转运蛋白家族、代谢通路和 CRISPR 等方面的差异。【结论】通过基因组水平上的差异初步揭示了海洋来源短杆菌对海洋环境的适应性, 为深入了解短杆菌菌株的环境适应机制奠定了基础。

关键词: 短杆菌属; 泛基因组; 比较基因组; 海洋环境适应性

Adaptation mechanism of *Brevibacterium* species to the marine environment based on pan-genomics analysis

PEI Shengxiang, HUANG Xiaoyun, NIU Siwen, XIE Fuquan, ZHANG Gaiyun*

Key Laboratory of Marine Biogenetic Resources, Third Institute of Oceanography, Ministry of Natural Resources, Xiamen 361005, Fujian, China

Abstract: [Objective] Our main goal was to clarify the adaptation mechanism of *Brevibacterium*

基金项目: 自然资源部第三海洋研究所基本科研业务费(2019011)

Supported by the Scientific Research Foundation of the Third Institute of Oceanography, MNR (2019011)

*Corresponding author. Tel/Fax: +86-592-2195833; E-mail: zhgyun@tio.org.cn

Received: 12 March 2021; Revised: 28 June 2021; Published online: 6 July 2021

species to the marine environment. **[Methods]** In this paper, six *Brevibacterium* strains isolated from different sea regions and belonging to different operational taxonomic units were sequenced, assembled and annotated. Combined with the genomic data of 23 typical and atypical *Brevibacterium* strains downloaded from NCBI, pan genomics analysis and species evolution analysis were carried out. **[Results]** Pan genomics analysis showed that the *Brevibacterium* strains had an open pan genome, which was consistent with the diversity of its living environment. There are obvious differences between marine *Brevibacterium* strains and other habitat *Brevibacterium* strains at the genomic level, mainly in the aspects of gene family expansion and contraction, transporter family, metabolic pathway and CRISPR. **[Conclusion]** These differences at the genomic level preliminarily revealed the adaptability of marine *Brevibacterium* strains to marine environment, which laid the foundation for a deep-going understanding of the environmental adaptation mechanism of *Brevibacterium* strains.

Keywords: *Brevibacterium*; pan-genomics; comparative genome; marine environment adaptation

短杆菌属(*Brevibacterium*)属于放线菌门,是短杆菌科中有效种数目最多的一个属。1953年Breed首次以*Brevibacterium linens*为模式种建立了该属^[1]。目前该属有效发表的菌种有35个(<https://lpsn.dsmz.de/genus/brevibacterium>),菌株分离自多种环境,例如大米^[2]、奶酪^[3]、人类皮肤^[4]、受损的壁画^[5],部分菌株是从发酵海鲜^[6]、腐烂的褐藻体^[7]和深海沉积物^[8-9]等高盐度的海洋环境中分离获得。短杆菌属菌株在工业上具有很多应用,研究发现短杆菌属可以利用细胞外蛋白酶和脂肪酶分解脂类和蛋白质(如酪蛋白)^[10-11],同时,许多短杆菌属细菌也具有修饰含硫氨基酸以产生挥发性硫化物的能力,从而有利于挥发性风味物质的产生^[12-14]。短杆菌属菌株常在许多不同类型的奶酪中用作表面成熟培养物,来缩短奶酪的成熟时间,减少变质,更好地控制奶酪香气并提高食品安全性^[15]。除此之外,研究表明部分短杆菌还具有使合成染料脱色的功能^[16]。在农业中,菌株*B. linens* RS16可通过调节抗氧化剂防御和H⁺ATPase活性赋予水稻基因耐盐性^[17]。*Brevibacterium antiquum* DSM 21545^T被用于通

过在混合培养的废水流中生长来回收富含磷的产品^[18]。

由于短杆菌在工业奶酪制造、生物降解和污水处理等生物技术方面应用成果显著,越来越多的短杆菌完成了基因组测序,为了解短杆菌菌株的生物学特征提供了有力支撑。Tettelin等^[19]于2005年第一次提出了泛基因组(pan-genome)的概念。泛基因组分析能够获得同一物种内的许多个体的基因组,用以完善该物种的遗传信息,能够捕获只存在于单个或部分个体及种群基因组上的特异性序列和功能基因,为深入研究该物种的系统进化地位以及形态差异相关的功能生物学奠定基础。本文以实验室前期分离鉴定的6株来自不同海域的短杆菌属菌株作为主要研究对象,结合从NCBI数据库下载的23株短杆菌属模式菌株或非模式菌株的基因组数据,采用泛基因组学方法,分析比较不同菌株之间的基因组特征。利用核心基因簇,研究菌株之间的演化关系。此外,结合基因组功能分析,探讨海洋短杆菌与其他生境中的短杆菌在基因组上的差异,探究短杆菌属对海洋环境的适应机制。

1 材料与amp;方法

1.1 菌株来源及选择标准

本研究中所使用的短杆菌基因组数据主要是从 NCBI 下载, 有模式菌株基因组数据的种优先下载该种的模式菌株基因组数据, 没有模式菌株基因组数据的种, 下载该种非模式菌株的相关基因组数据。目前, 短杆菌属共有 35 个有效发表的种, 在 NCBI 数据库中, 已有 14 个种的模式菌株基因组数据被公开, 另有 9 个种虽没有该种的模式菌株基因组数据, 但该种中有非模式性菌株的相关基因组数据。因此, 我们从 NCBI 下载了这 14 个种的模式菌株及另外 9 个种中的 9 个非模式菌株的基因组数据, 共计 23 个种的相关菌株基因组数据进行后续研究。此外, 本实验室也从不同洋区获得了大量的短杆菌属菌株, 以 16S rRNA 基因序列相似性 99% 为标准共划分为 6 个操作分类单元 (operational taxonomic units, OTU), 从每个 OTU 中各选取了 1 株菌株 (YB235、Wo024、WW007、O1、o2、W7.2) 作为代表性菌株进行了全基因组测序, 其中菌株 O1 已完成新种鉴定并公布了有效名, YB235、Wo024、W7.2 和 o2 为潜在的新种。29 株短杆菌的具体相关信息见表 1, 其中有 9 株菌来源于海洋, 称为海洋来源菌株, 其他 20 个菌株来自多个生境 (人、土壤、奶酪和肉汤等), 称为多重来源菌株。

1.2 基因组测序、组装和注释

本实验所分离的 6 株海洋来源的短杆菌完整基因组由天津生物芯片公司采用第三代测序技术 PacBio Sequel 系统测序完成。利用 HGAP 4^[20] 软件进行 *De novo* 组装获得。为了避免由于各项目所采用的基因预测及注释方法不同而造成的偏差, 我们对 29 株菌采用

Prokka1.13 进行了重预测。同样, 利用 BLAST 分别对 nr 库和 COG 直系同源蛋白数据库进行基因注释和 COG 注释, 利用软件 Canoco5 对基因组大小、CDS 和 GC 含量进行主成分分析。利用 CRISPRCasFinder^[21], 在线预测各个菌株中的 CRISPR 元件, 统计每个菌株中 CRISPR 元件的数量。

1.3 泛基因组学分析

1.3.1 泛基因组特征分析

利用原核生物泛基因组自动化分析软件 (pan-genomics analysis pipeline, PGAP)^[22] 对 29 株短杆菌进行泛基因组分析。首先, 根据流程中的格式要求, 编写 python 程序, 依据 Prokka 的注释结果, 准备核酸 (nuc)、蛋白 (pep) 和功能 (function) 文件。功能聚类选择 GF 方法进行直系同源基因鉴定, 将聚类结果 Orthologs_Cluster.txt 输入 PanGP 绘制泛基因组特征曲线和核心基因组特征曲线^[23]。我们从 PGAP 功能基因的聚类结果中, 统计了各个菌株的核心基因、非必需基因以及特异性基因的分布等信息。

1.3.2 核心基因组建树、分化时间鉴定及 ANI (average nucleotide identity) 分析

为更深入了解 29 株短杆菌的进化关系, 我们挑选与短杆菌属同科不同属的 *Spelaieicoccus albus* D3-40^T (accession: JACBZP000000000) 菌株作为外源。对这 30 个菌株进行 PGAP 分析, 得到单拷贝核心基因; 利用 MAFFT v7.471 进行多序列比对^[24], 并用 TrimAL 切割多序列比对结果; 利用软件 IQ-TREE1.6.12 筛选碱基替换模型并构建系统进化树^[25]。

从 TIMETREE (<http://www.timetree.org/>) 上获取不同物种的分化时间, 然后利用 r8s 软件在系统进化树的基础上对菌株系统发育的分化时间进行鉴定^[26]。

表 1 29 株短杆菌基因组信息

Table 1 Genomic information of 29 *Brevibacterium* strains

Strains	Size/bp	CDS No.	GC	CRISPER	Level	Sources	Accession number
<i>B. profundi</i> O1 ^T	4 167 127	3 654	0.65	5	Complete	Ocean	CP051626
<i>Brevibacterium</i> sp. o2	4 253 182	3 746	0.65	4	Complete	Ocean	CP050154
<i>Brevibacterium</i> sp. Wo024	4 156 794	3 617	0.65	4	Complete	Ocean	CP050152
<i>Brevibacterium</i> sp. YB235	3 913 130	3 513	0.65	2	Complete	Ocean	CP050153
<i>Brevibacterium</i> sp. W7.2	3 861 567	3 390	0.68	8	Complete	Ocean	CP073089
<i>B. oceani</i> WW007	4 304 299	3 741	0.65	2	Complete	Ocean	CP072846
<i>B. sediminis</i> FXJ8.269 ^T	4 173 147	3 664	0.65	6	Draft	Ocean	BMJG00000000
<i>B. marinum</i> DSM 18964 ^T	4 221 031	3 756	0.64	6	Draft	Ocean	JAATJN00000000
<i>B. oceani</i> BBH7 ^T	4 481 272	4 006	0.65	2	Draft	Ocean	JABKDE00000000
<i>B. ravenstergense</i> CCUG56047 ^T	2 297 397	2 020	0.62	3	Draft	Human	PKGO00000000
<i>B. ihuae</i> CV3 ^T	3 096 817	2 736	0.71	3	Draft	Human	FXWZ00000000
<i>B. senegalense</i> JC43 ^T	3 427 329	3 042	0.69	6	Draft	Human	CABKRC00000000
<i>B. mcbrellneri</i> ATCC 49030 ^T	2 561 804	2 368	0.58	1	Draft	Human	ADNU00000000
<i>B. epidermidis</i> NBRC 14811 ^T	3 703 261	3 258	0.64	4	Draft	Human	PDHL00000000
<i>B. album</i> DSM 18261 ^T	4 094 970	3 541	0.71	10	Complete	Soil	KE384526
<i>B. permense</i> VKM Ac 2280 ^T	3 702 760	3 308	0.65	6	Draft	Soil	JABSSX00000000
<i>B. sanguinis</i> 2b TX	3 763 581	3 332	0.68	6	Draft	Fresh water	QNRZ00000000
<i>B. siliguriense</i> DSM 23676 ^T	4 023 083	3 594	0.64	1	Complete	River water	LT629766
<i>B. jeotgali</i> DSM 29217 ^T	3 330 439	2 909	0.68	2	Draft	Seafood	VIUB00000000
<i>B. sandarakinum</i> DSM 22082 ^T	4 441 057	4 041	0.63	2	Complete	Wall	LT629739
<i>B. iodium</i> NCTC 9742	3 614 741	3 269	0.64	6	Draft	Meat	UFTQ00000000
<i>B. aurantiacum</i> SMQ 1419	4 038 634	3 709	0.63	1	Draft	Cheese	VLTK00000000
<i>B. antiquum</i> DSM 21545 ^T	4 254 817	3 813	0.63	4	Draft	Soil	FXZE00000000
<i>B. linens</i> ATCC 19391	3 814 300	3 404	0.65	6	Complete	Unknown	CP030797
<i>B. paucivorans</i> UMB 1301	2 999 495	2 636	0.56	1	Draft	Human	PNHK00000000
<i>B. celere</i> 3b TX	3 766 273	3 336	0.68	7	Draft	Unknown	QNSB00000000
<i>B. luteolum</i> NEB 1784	3 179 618	2 885	0.67	2	Complete	Unknown	CP035810
<i>B. yomogidense</i> B Co 03.11	3 286 027	2 846	0.68	2	Draft	Unknown	FWFF00000000
<i>B. casei</i> FDAARGOS 902	3 960 094	3 546	0.68	5	Draft	Unknown	VEKE00000000

利用 perl 脚本 ANI.pl (<https://github.com/jhbadger/scripts/blob/master/ANI.pl>)计算 29 个短杆菌相互之间的 ANI, 并通过 Heml 软件绘制热图。

1.3.3 单拷贝核心基因一致性分析

将 29 株短杆菌分成海洋来源和多重来源

两大分支, 比较 2 个分支内部以及 2 个分支之间短杆菌基因序列的一致性, 利用 blastn 对物种中单拷贝核心基因序列进行 all to all 比对。分别提取出两个分支内序列一致性大于 70% 的比对结果, 用 python 绘制序列一致性的密度分布曲线来观察两个分支内部基因序列的一致性差异。

1.3.4 转运系统

为研究不同环境下短杆菌与环境的相互作用,我们比较了海洋来源和多重来源两个分支短杆菌中转运蛋白系统的差异。提取注释为转运蛋白(transporter protein)和假定蛋白(hypothetical protein)的基因作为转运蛋白的候选集,编写 python 脚本,下载 TransportDB 数据库中所有的 fasta 序列,通过 BLAST 比对(identity 大于 50%, evalue 值 $1e-5$, 如果一个基因比对上多个功能基因,选取 identity 最大的一个),进一步注释各个候选集中的转运蛋白。然后根据 TransportDB 中的转运蛋白家族功能分类对各个菌株中注释后的转运蛋白进行分类并统计数量,绘制热图比较分析具有差异的转运蛋白家族。

1.3.5 KEGG 代谢通路分析

为研究短杆菌属在不同环境下的信号通路,利用 KEGG 注释网站(<https://www.genome.jp/kegg/kaas/>)对各个菌株进行了在线的 KEGG 代谢通路分析,然后编写 python 脚本下载菌株所有通路的结果。比较细菌的信号系统、双组分系统以及磷酸转移酶系统 3 个与环境适应相关的通路^[27]。

1.3.6 基因家族扩增和收缩

为研究短杆菌属在不同环境下的基因家族的变化,利用 Orthofinder2.3.3 分析短杆菌属所有菌株的基因组^[28],获取统计基因家族中基因数量的文件。据此,用 CAFÉ 软件计算分析短杆菌属基因家族的扩增和收缩^[29]。

2 结果与分析

2.1 基因组组装和注释

对 29 株短杆菌的注释结果如表 1,本实验室所测 6 株菌株均组装成 1 条完整的 contig,说明组装质量比较好,基因组大小在 4 Mb 左右,

GC 含量等于或大于 0.65,编码 DNA 序列(coding DNA sequences, CDS)的数量为 2 020–4 041。

2.2 基因组特征比较

为研究不同环境下短杆菌基因组基本特征的差异,我们对 29 株短杆菌的基因组大小、CDS 数量和 GC 含量进行比较(图 1A),发现海洋来源的短杆菌普遍具有较大的基因组,对以上三者进行主成分分析(图 1B),发现 CDS 和基因组大小可以比较明显的将海洋短杆菌和部分其他生境来源的短杆菌分开。

规律间隔成簇短回文重复序列(clustered regularly interspaced short palindromic repeat, CRISPR)是一种特异性的防御机制,能够有效防御质粒或噬菌体对细菌本身的水平基因转移,保持自身遗传信息的完整性。短杆菌属 CRISPR 分析发现,所有短杆菌均含有 CRISPR 元件。海洋来源短杆菌和多重来源短杆菌含有的 CRISPR 数量的均值都为 4。在 9 株海洋来源短杆菌菌株中,有 2/3 的菌株含有的 CRISPR 元件数量大于等于平均值;而 20 株其他生境来源的短杆菌有 1/2 菌株 CRISPR 元件数量大于等于平均值,说明大部分海洋来源短杆菌含有更多的 CRISPR 元件。这可能是因为海洋环境中含有更多的噬菌体,因此海洋来源的短杆菌需要进化出更多的 CRISPR 元件,以抵御外源 DNA 的侵害,降低水平基因转移。

2.3 泛基因组学分析

2.3.1 泛基因组大小及其特征曲线

为了直观展示 29 株短杆菌泛基因组特征,根据 PGAP 聚类结果,利用 PanGP 绘制短杆菌的泛基因组特征曲线(图 2),蓝色表示泛基因组特征曲线,绿色则表示核心基因组特征曲线。可以看出,随着菌株数量增多,泛基因组呈现明显增大趋势,因此短杆菌具有开放型泛基因组,同时随着短杆菌数量增加核心基因组明显

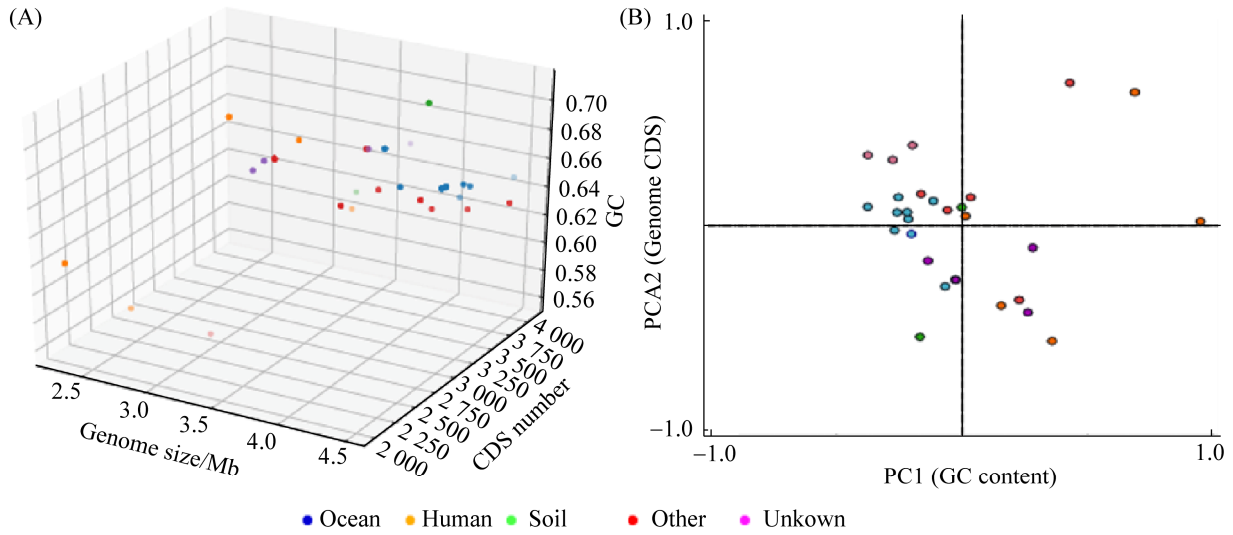


图 1 基因组特征比较

Figure 1 Genomic characteristics comparison. A: three-dimensional plots of the genome size, CDS number and GC content. B: the principal component analysis of the genome size, CDS number and GC content.

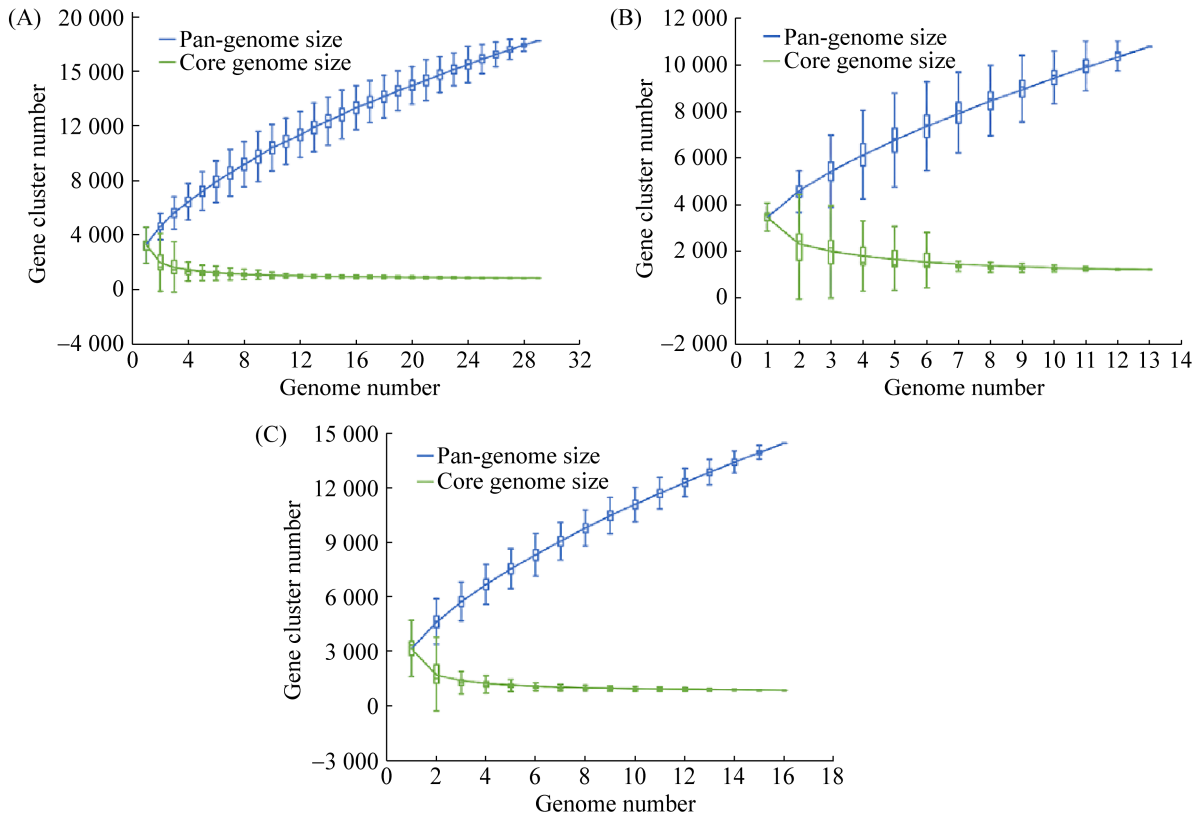


图 2 泛基因组特征曲线

Figure 2 Pan-genome characteristic curve. A: pan-genome characteristic curves of 29 strains; B: pan-genome characteristic curves of 9 marine-derived strains; C: pan-genome characteristic curves of 20 multiple source strains. All results are expressed as the mean±standard deviation (SD).

减小。这 29 株短杆菌的泛基因组大小为 17 445。泛基因组的特征方程为 $y=2\,561.89x^{0.55}+696.79$ ，同时 $R^2=0.999\,946$ 、平均每增加一个基因组就增加 500 多个基因。通常认为细菌需要改变自身的基因组，适应不同的环境，因此在多种不同的生境中生存的细菌需要一个更大的泛基因组，由于短杆菌生境的多样性，短杆菌属开放型的泛基因组与其生境多样性一致。

对海洋来源分支和多重来源分支分别进行泛基因组分析，可以看出两分支内也具有开放型泛基因组(图 2B、C)，泛基因组的特征方程

分别为 $y=1\,823.19x^{0.53}+1\,762.53$ 且 $R^2=0.999\,901$ 和 $y=3\,034.61x^{0.53}+43.88$ 且 $R^2=0.999\,957$ 。

2.3.2 核心基因及菌株特异性基因组分布

为研究短杆菌属的基因组的差异，我们分析了各个菌株中核心基因和菌株特异基因的分布。PAGP 功能聚类结果显示，29 株菌株共有的核心基因为 838 个，其中 734 个是单拷贝核心基因(图 3A)。分析两个分支内部核心基因和特有基因，海洋分支来源短杆菌共有的功能基因有 2 042 个，其中 1 912 个是单拷贝核心基因(图 3B)。多重分支来源短杆菌共有的功能基因

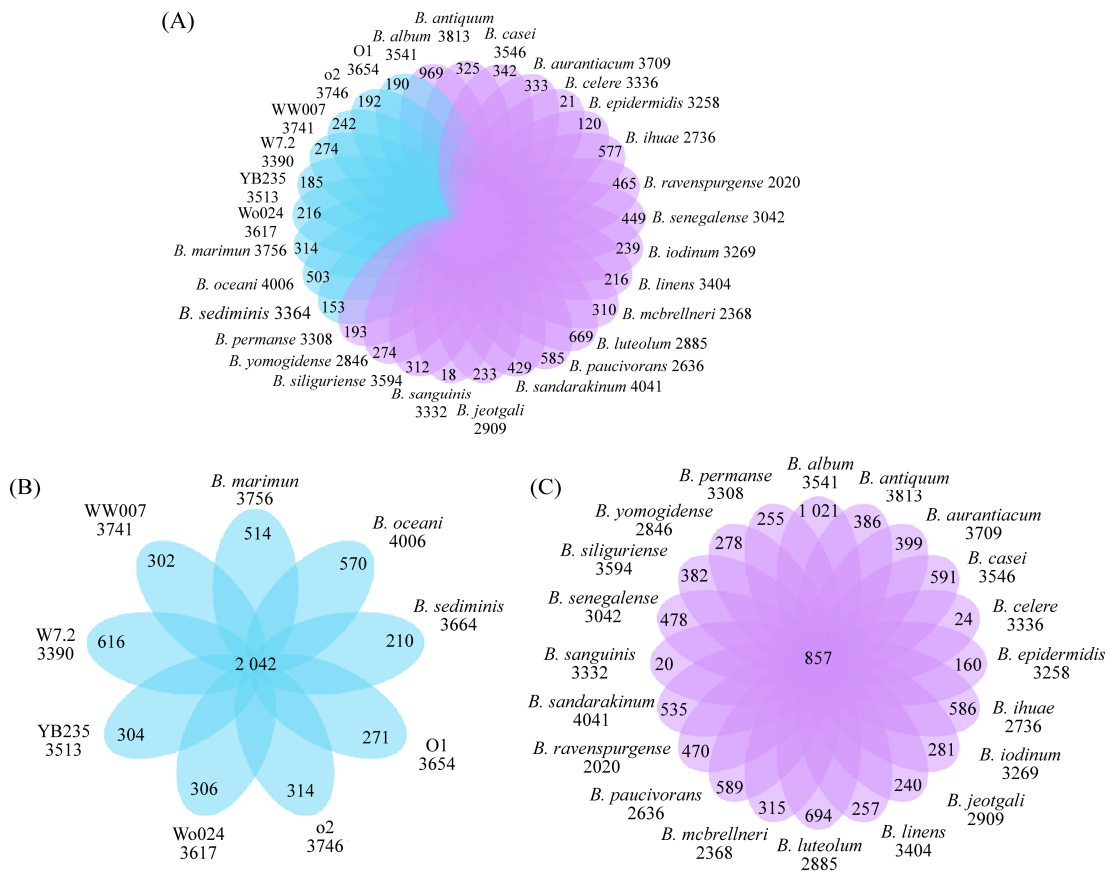


图 3 核心基因和菌株特异性基因花瓣图

Figure 3 Flower plots of core genes and strain-specific genes. A: flower plots showing the core gene number (in the center) and strain-specific gene number (in the petals) in 29 *Brevibacterium* strains; B: flower plots showing the core gene number (in the center) and strain-specific gene number (in the petals) in the marine-derived subgroup; C: flower plots showing the core gene number (in the center) and strain-specific gene number (in the petals) in the multiple source's subgroup. The numbers under the strain names represent the total number of coding proteins. Blue indicates the marine-derived subgroup and purple indicates the multiple sources subgroup.

有 857 个,其中 757 个是单拷贝核心基因(图 3C)。将 29 株短杆菌特异性基因的分布进行比较,两个分支内各个菌株中特异性基因的数量较多,这是由于随着菌株数量的增加,一部分菌株特异性基因变成非必需基因。

2.3.3 进化分析

为了深入研究短杆菌属进化关系,以菌株 *Spelaicoccus albus* D3-40^T 作为外群,以 30 个物种共有的 640 个单拷贝核心基因构建系统进化树。借助软件 IQ-TREE 测试了 748 个碱基替换模型,最优模型是 Blosum62+F+R4,之后利用最大似然法,进行 1 000 次自展,成功构建了核心基因组系统进化树(图 4A)。同时构建了 30 个菌株的 16S rRNA 基因序列进化树(图 4B)。二者的拓扑结构并不一致,在核心基因组树中,基本所有分支的 bootstrap 值均为 100,而 16S rRNA 基因树只有 3 个,说明核心基因组树具有更大的可靠性。

在核心基因组树的基础上,探索短杆菌属的分化时间,发现短杆菌属最近共同祖先最早出现在大约 88 百万年前(图 5)。菌株 *B. siliguriense* DSM 23676^T 和 *Brevibacterium* sp. o2 分化时间最近,在 100 万年前。

平均核苷酸相似度(average nucleotide identity, ANI)是在核苷酸水平上比较两个基因组亲缘关系的指标,在近源物种之间具有较高的相似性,我们将 29 个短杆菌进行相互比较,计算 ANI。结果如图 6 所示,可以看出 *B. iodinum*、*B. sediminis*、*B. permense* 的 ANI 聚类在一起,说明它们之间的遗传距离比较接近。

2.3.4 单拷贝核心基因序列一致性

为了研究不同生境下短杆菌基因序列的差异,我们提取短杆菌核心基因,分成海洋和多重来源两大分支。在两大分支内部,不同物种的核心基因序列进行相互比较,统计序列一致性。结果如图 7 所示,海洋来源的短杆菌核心

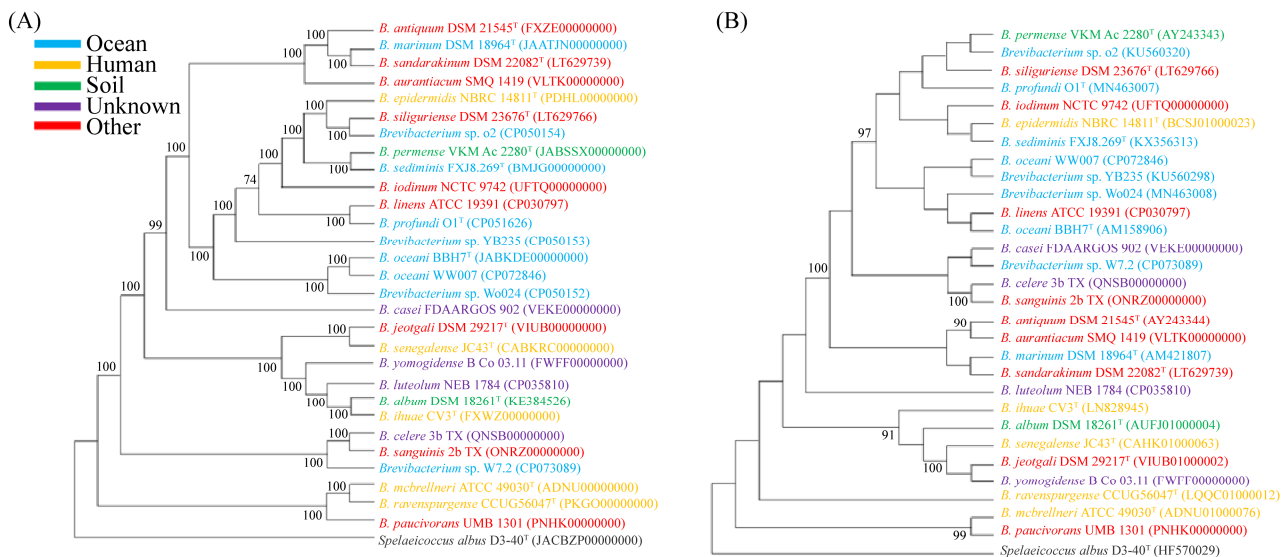


图 4 物种进化树

Figure 4 The evolutionary tree of species. A: the evolutionary tree based on single copy genes. B: the phylogenetic tree based on 16S rRNA gene sequences. The GenBank accession numbers for genome sequences of the sequenced strains are shown in parentheses. Source: ocean (blue), human (yellow), soil (green), unknow (purple), other (red).

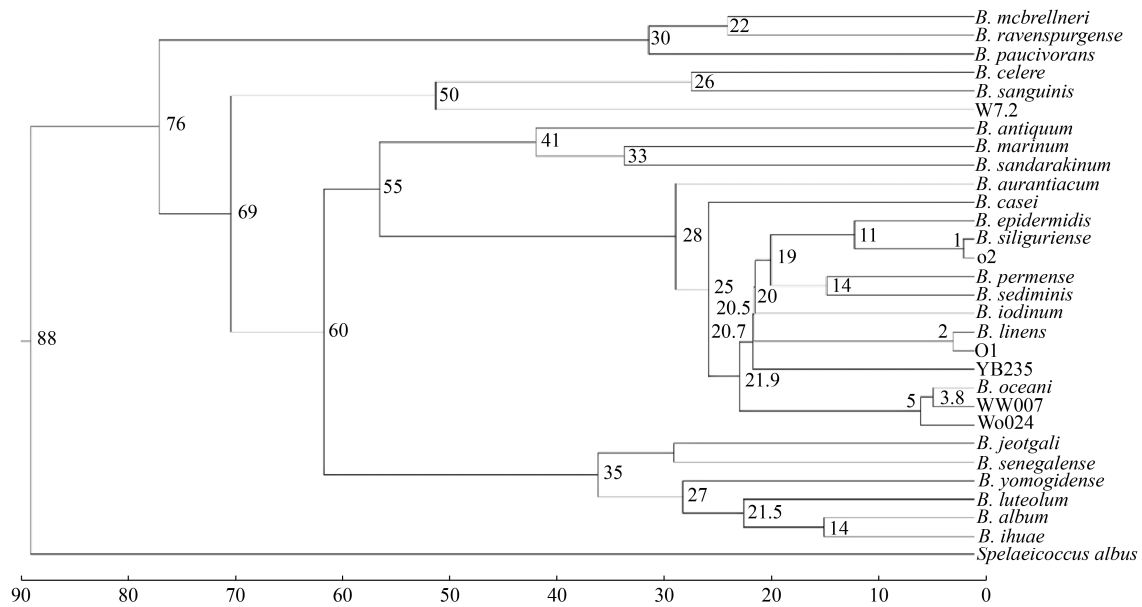


图 5 29 株短杆菌的分化时间

Figure 5 The differentiation time of 29 *Brevibacterium* strains. The bootstrap value represents the branch differentiation time, the unit is million years.

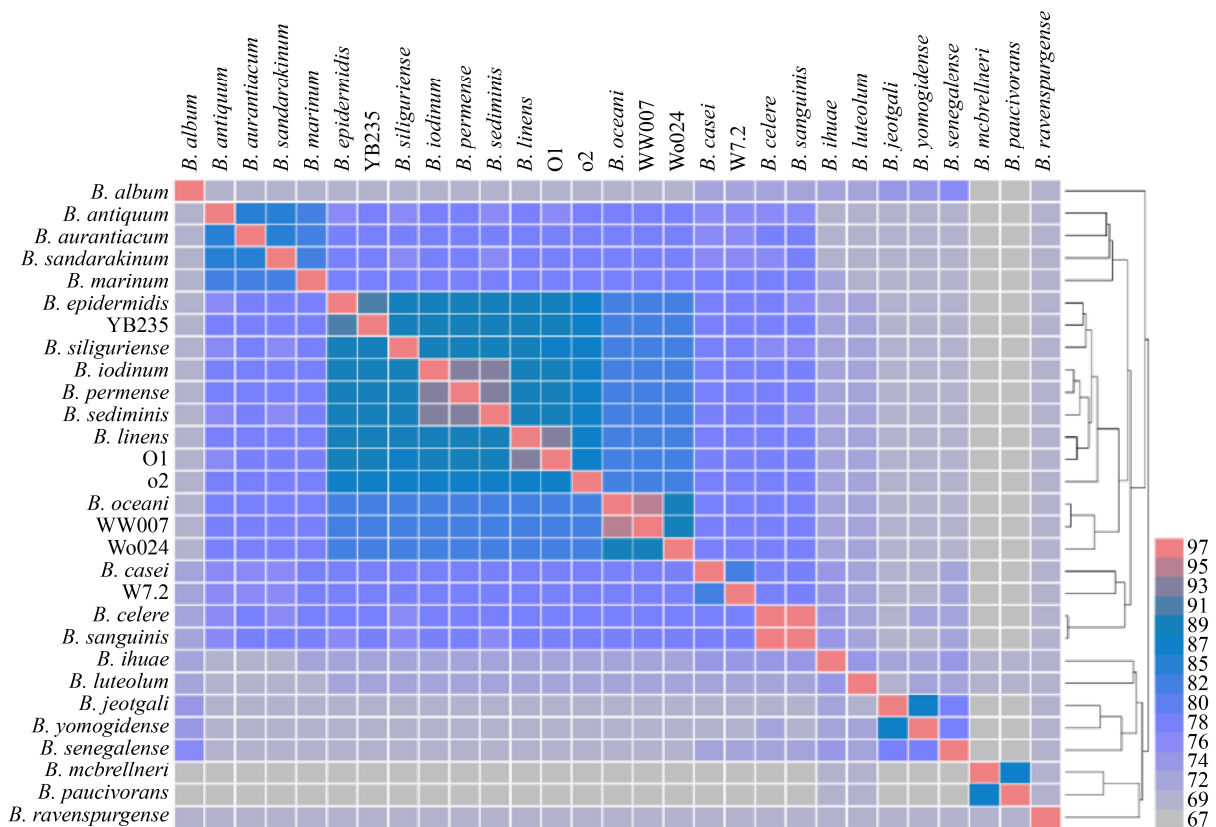


图 6 29 株短杆菌 ANI 比较分析

Figure 6 ANI comparative analysis of 29 *Brevibacterium* strains.

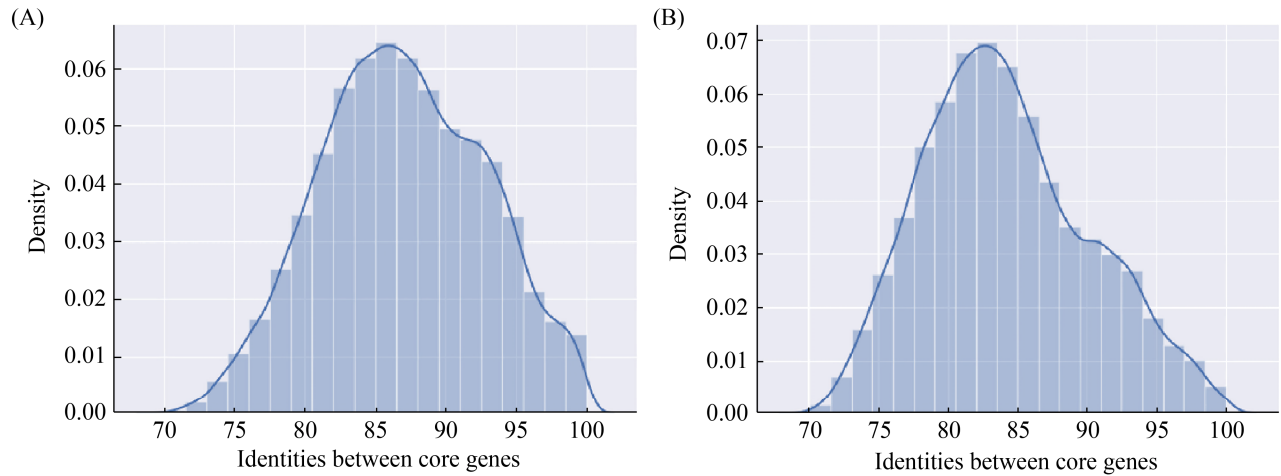


图 7 基因序列一致性比较

Figure 7 Comparison of gene sequence identity. A: density map of marine-derived strains; B: density map of multiple source strains.

基因一致性密度分布曲线的峰值约为 87%，而多重来源分支一致性密度曲线峰值约为 83%。海洋来源分支核心基因更保守，而多重来源分支核心基因具有更高的多样性以适应不同的生存环境。

2.3.5 转运蛋白家族比较

为了研究短杆菌和环境之间的相互作用，我们分析比较了海洋来源和多重来源跨膜运输系统的差异。我们根据 TransportDB 数据库中转运蛋白家族的分类，对两个分支富集的转运蛋白进行了比较(图 8)。海洋来源均含有的转运蛋白有 ABC (The ATP-binding cassette superfamily)、BCCT (The betaine/carnitine/choline transporter (BCCT) Family)、MFS (The major facilitator superfamily)、SSS (The solute: sodium symporter family)、Nramp (The metal ion transporter family)、MgtE (The Mg^{2+} transporter-E family)、CitMHS (The citrate- Mg^{2+} : H^+ citrate- Ca^{2+} : H^+ symporter family)、NSS (The neurotransmitter: sodium symporter family)、NCS2 (The nucleobase: cation symporter-2

family)、PiT (The inorganic phosphate transporter family)、BASS (The bile acid: Na^+ symporter family)，基本含有的是 Sulp (The sulfate permease family)、TRAP-T (The tripartite ATP-independent periplasmic transporter family)、DASS (The divalent anion: Na^+ symporter family)。海洋来源富集更多 ABC、MFS、BCCT、SSS、Nramp、MgtE、CitMHS、DASS 等转运蛋白。其中 BCCT 有助于协助相容性溶质甜菜碱，肉毒碱和胆碱的积累。这不仅有利于维持细胞渗透平衡，而且还充当蛋白质和细胞成分的稳定剂，防止变性，抵制高离子强度的影响。海洋来源菌株含有更多与营养吸收相关的转运蛋白：SSS、NCS2。ABC 转运蛋白可能与其需要转运和代谢各种有毒物质和几丁质、支链淀粉、纤维素和淀粉能力相关。MFS 促进糖、药物分子、肽、三羧酸循环代谢产物、有机阴离子和无机阴离子等溶质在电化学梯度下进行跨膜运输^[30]。

2.3.6 KEGG 代谢通路

为研究短杆菌如何针对不同环境做出适应性反应，针对各个短杆菌的 KEGG 代谢信号通

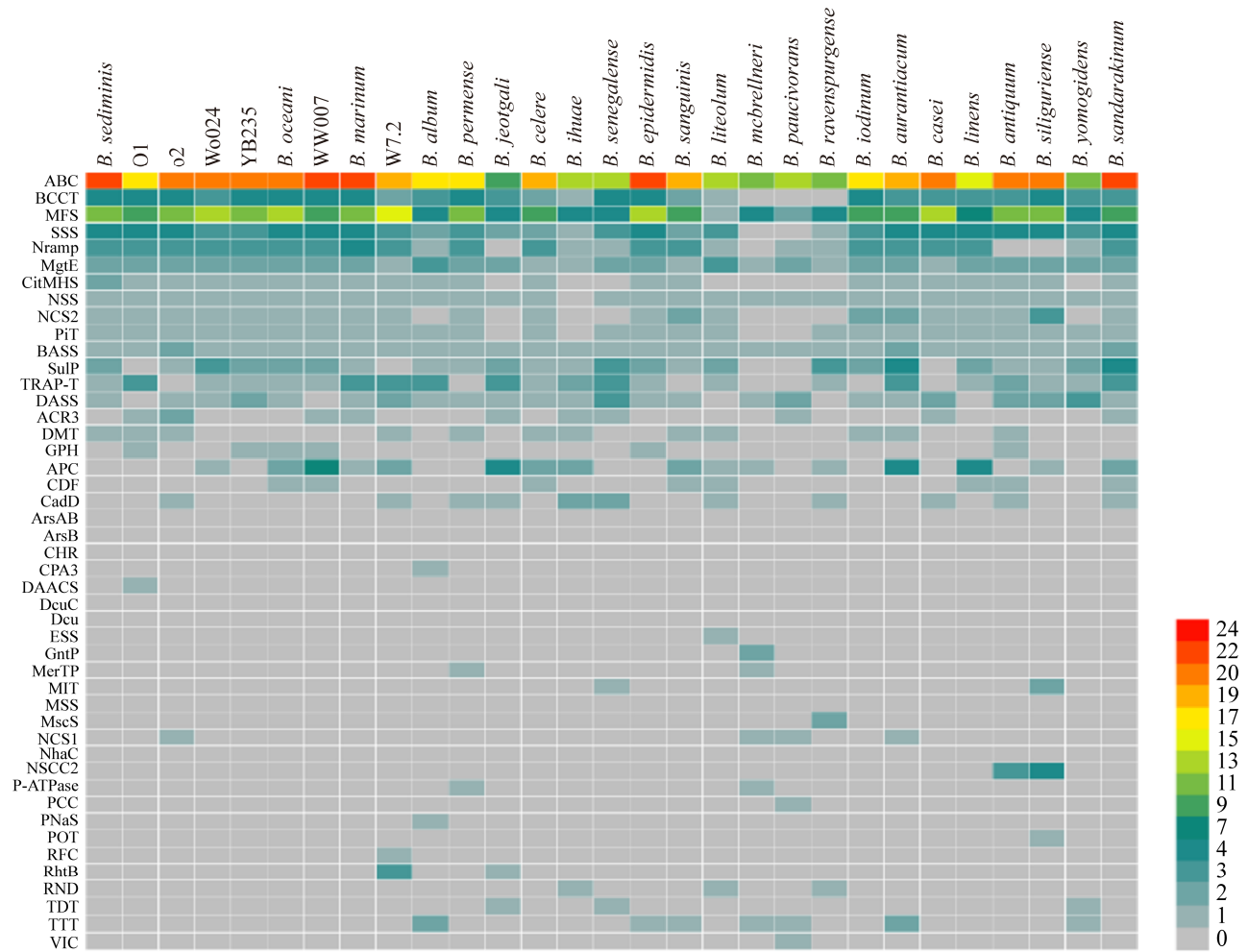


图 8 转运蛋白家族比较
Figure 8 Comparison of transporter family.

路进行了分析。主要比较了两个分支中的双组分系统(two-component system)、细菌分泌系统(bacterial secretion system)以及磷酸转移酶系统(phosphotransferase system)。结果显示,在细菌分泌系统中,所有短杆菌均含有 SecDF、SecY、YidC、SecA、FstY、ffh、TatC。部分海洋来源短杆菌(O1、o2、Wo024、YB235、FXJ8.269)含有 SecE 基因,只有少部分多重来源短杆菌含有 SecE 基因(NBRC 14811、NCTC9742、ATCC 19391s、VKM Ac-2280、DSM 23676)。Sec 途径是细菌中第一个发现的跨细胞质膜向细菌周

质和外膜输出蛋白质的分泌途径^[31]。SecE 基因是一个与细胞胞外运输相关的基因^[32],这可能是因为大部分海洋来源短杆菌向周围环境释放一类大分子的蛋白杂多糖胞外聚合物,防止冰晶形成,以适应低温环境。此外,在双组分系统中,只有 O1 同时含有与盐压力相关的 DegS 基因,与低温相关的 DesK 基因,细胞壁压力刺激物 VraS 基因,细胞壁活性物 LiaS 基因。这可能是因为短杆菌 O1 生活在 7 000 m 下的海底,这些基因可以帮助其适应高压、低温的环境。磷酸烯醇式丙酮酸依赖的糖磷酸转

移酶系统(PTS)是一个大的碳水化合物活性转运系统,可以催化底物葡萄糖苷和葡萄糖、海藻糖磷酸化。我们发现含有 SecE 基因的菌株(O1、o2、YB235、DSM 21545、FXJ8.269、FDAARGOS_902、NBRC 14811、VKM Ac-2280、DSM 22082、DSM 23676)基本含有与葡萄糖苷和葡萄糖、海藻糖磷酸化相关基因,可能是这

些菌株为适应生存而在进化过程中获得的一种自身调节机制,可以更高效地利用碳源。

2.3.7 基因家族扩增和收缩

为研究短杆菌在进化过程中基因家族的变化,我们做了基因家族的扩张和收缩分析(图9)。最近共同祖先含有的基因家族有5966个,在短杆菌进化过程中大部分基因家族都在收

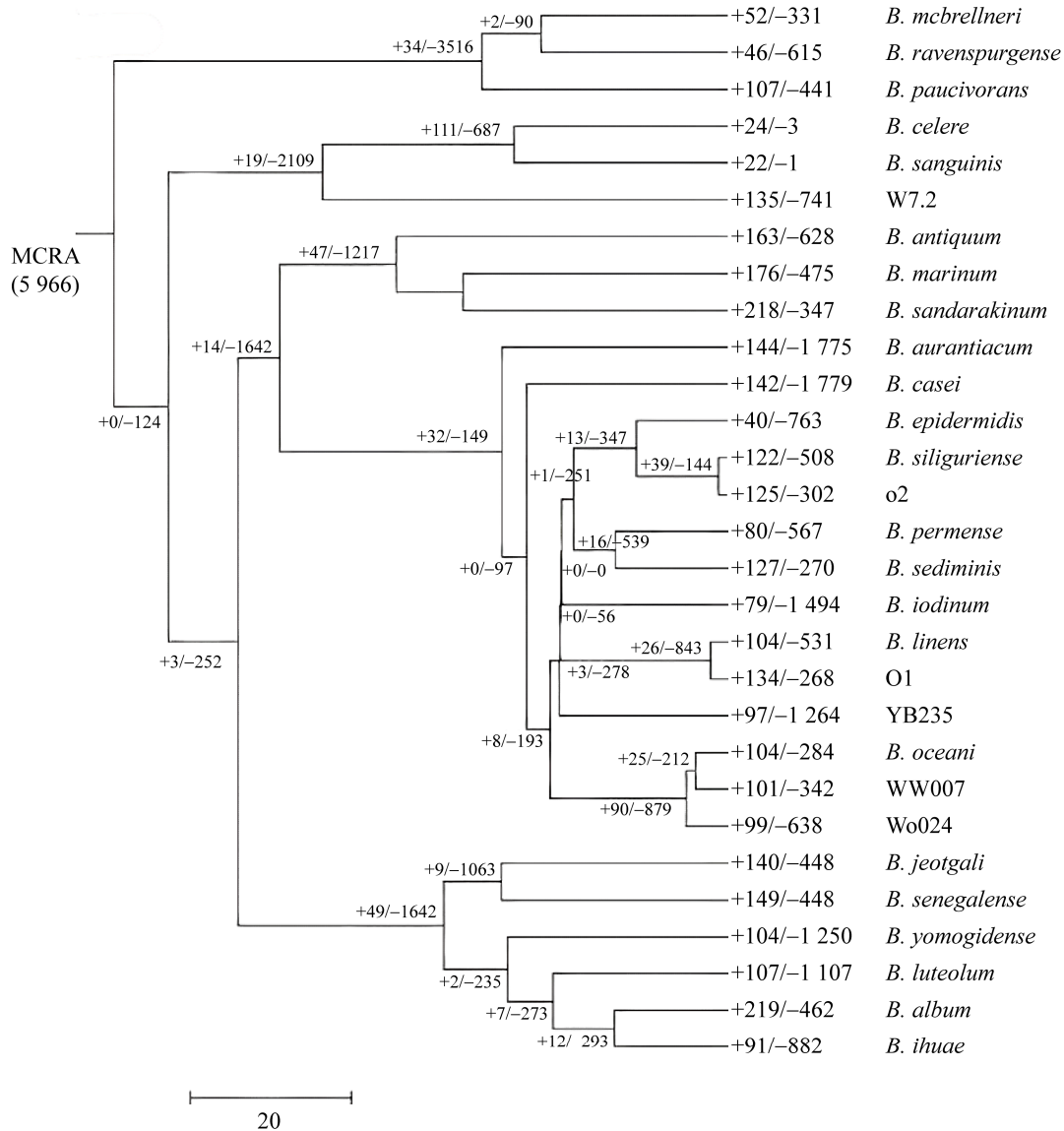


图9 基因家族的扩张和收缩

Figure 9 Expansion and contraction of gene families. +: gene gain; -: gene loss; the numbers of genes within each strain are listed in the branch.

缩, 在 55 百万年、50 百万年、41 百万年、35 百万年、30 百万年分支节点的时候, 基因家族收缩得最剧烈, 从而导致大量的基因丢失, 而海洋来源的短杆菌基本不在这些分支上(图 5, 图 9), 且大部分海洋来源短杆菌近期也没有经历大规模基因家族的收缩, 这可能就是海洋来源短杆菌基因组偏大的原因。

3 总结

通过对短杆菌属不同生境菌株基因组进行比较分析, 短杆菌基因组大小变化范围是 2.4–4.9 Mb。与多重来源短杆菌基因组相比, 海洋来源短杆菌基因组偏大。这可能与多重来源短杆菌经历更多的基因家族收缩相关。比较两个分支的核心基因的一致性密度曲线, 发现海洋来源短杆菌具有更高的一致性, 而多重来源短杆菌具有更高的多样性, 这可能是短杆菌属菌株为适应不同的生存环境, 其核心基因的大小随环境的变化而调整导致的。

在转运蛋白家族上, 海洋来源短杆菌富集更多的 ABC、MFS、BCCT、SSS、Nramp、MgtE、CitMHS、DASS 及 BCCT 转运蛋白家族成员, 这些物质有助于协助相容性溶质甜菜碱、肉碱和胆碱的积累。这些兼容的溶质可维持细胞渗透平衡, 充当蛋白质和细胞成分的稳定剂。分析 KEGG 代谢路径, 发现部分短杆菌菌株为适应海洋环境进化出了自己独特的基因以适应高压、寒冷的环境, 同时为了更加高效地利用碳源获得了与 PTS 代谢相关的基因。海洋来源短杆菌也具有更多的 CRISPR 元件以应对海洋中大量噬菌体的威胁。

综上, 本研究通过比较海洋来源与其他生境来源的短杆菌在基因组水平上的差异性, 主

要是进化过程中的基因家族扩增收缩、转运蛋白家族、代谢通路和 CRISPR 等方面的差异揭示了短杆菌的海洋环境适应性机制。下一步可对特定基因进行相关实验验证, 深入分析短杆菌的海洋环境适应性机制。

参考文献

- [1] Breed RS. The families developed from Bacteriaceae Cohn with a description of the family Brevibacteriaceae. 1953.
- [2] Komagata K, Iizuka H. New species of *Brevibacterium* isolated from rice. *Journal of the Agricultural Chemical Society of Japan*, 1964, 38(10): 496–502.
- [3] Collins MD, Farrow JAE, Goodfellow M, Minnikin DE. *Brevibacterium casei* sp. nov. and *Brevibacterium epidermidis* sp. nov.. *Systematic and Applied Microbiology*, 1983, 4(3): 388–395.
- [4] Roux V, Raoult D. *Brevibacterium massiliense* sp. nov., isolated from a human ankle discharge. *International Journal of Systematic and Evolutionary Microbiology*, 2009, 59(8): 1960–1964.
- [5] Heyrman J, Verbeeren J, Schumann P, Devos J, Swings J, de Vos P. *Brevibacterium picturae* sp. nov., isolated from a damaged mural painting at the Saint-Catherine chapel (Castle Herberstein, Austria). *International Journal of Systematic and Evolutionary Microbiology*, 2004, 54(5): 1537–1541.
- [6] Choi EJ, Lee SH, Jung JY, Jeon CO. *Brevibacterium jeotgali* sp. nov., isolated from jeotgal, a traditional Korean fermented seafood. *International Journal of Systematic and Evolutionary Microbiology*, 2013, 63(Pt_9): 3430–3436.
- [7] Ivanova EP, Christen R, Alexeeva YV, Zhukova NV, Gorshkova NM, Lysenko AM, Mikhailov VV, Nicolau DV. *Brevibacterium celere* sp. nov., isolated from degraded thallus of a brown alga. *International Journal of Systematic and Evolutionary Microbiology*, 2004, 54(6): 2107–2111.
- [8] Chen P, Zhang LM, Wang J, Ruan JS, Han XQ, Huang Y. *Brevibacterium sediminis* sp. nov., isolated from deep-sea sediments from the Carlsberg and Southwest

- Indian Ridges. *International Journal of Systematic and Evolutionary Microbiology*, 2016, 66(12): 5268–5274.
- [9] Bhadra B, Raghukumar C, Pindi PK, Shivaji S. *Brevibacterium oceani* sp. nov., isolated from deep-sea sediment of the Chagos Trench, Indian Ocean. *International Journal of Systematic and Evolutionary Microbiology*, 2008, 58(1): 57–60.
- [10] Ozturkoglu-Budak S, Wiebenga A, Bron PA, de Vries RP. Protease and lipase activities of fungal and bacterial strains derived from an artisanal raw ewe's milk cheese. *International Journal of Food Microbiology*, 2016, 237: 17–27.
- [11] Rattray FP, Fox PF. Aspects of enzymology and biochemical properties of *Brevibacterium linens* relevant to cheese ripening: a review. *Journal of Dairy Science*, 1999, 82(5): 891–909.
- [12] Amarita F, Yvon M, Nardi M, Chambellon E, Delettre J, Bonnarme P. Identification and functional analysis of the gene encoding methionine-gamma-lyase in *Brevibacterium linens*. *Applied and Environmental Microbiology*, 2004, 70(12): 7348–7354.
- [13] Bonnarme P, Psoni L, Spinnler HE. Diversity of L-methionine catabolism pathways in cheese-ripening bacteria. *Applied and Environmental Microbiology*, 2000, 66(12): 5514–5517.
- [14] Yvon M, Chambellon E, Bolotin A, Roudot-Algaron F. Characterization and role of the branched-chain aminotransferase (BcaT) isolated from *Lactococcus lactis* subsp. *cremoris* NCDO 763. *Applied and Environmental Microbiology*, 2000, 66(2): 571–577.
- [15] Anast JM, Dzieciol M, Schultz DL, Wagner M, Mann E, Schmitz-Esser S. *Brevibacterium* from Austrian hard cheese harbor a putative histamine catabolism pathway and a plasmid for adaptation to the cheese environment. *Scientific Reports*, 2019, 9: 6164.
- [16] Díez-Méndez A, García-Fraile P, Solano F, Rivas R. The ant *Lasius niger* is a new source of bacterial enzymes with biotechnological potential for bleaching dye. *Scientific Reports*, 2019, 9: 15217.
- [17] Chatterjee P, Samaddar S, Niinemets Ü, Sa TM. *Brevibacterium linens* RS16 confers salt tolerance to *Oryza sativa* genotypes by regulating antioxidant defense and H⁺ ATPase activity. *Microbiological Research*, 2018, 215: 89–101.
- [18] Simoes F, Vale P, Stephenson T, Soares A. Understanding the growth of the bio-struvite production *Brevibacterium antiquum* in sludge liquors. *Environmental Technology*, 2018, 39(17): 2278–2287.
- [19] Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou LW, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJB, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(39): 13950–13955.
- [20] Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*, 2013, 10(6): 563–569.
- [21] Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Research*, 2018, 46(W1): W246–W251.
- [22] Zhao YB, Wu JY, Yang JH, Sun SX, Xiao JF, Yu J. PGAP: pan-genomes analysis pipeline. *Bioinformatics: Oxford, England*, 2012, 28(3): 416–418.
- [23] Zhao YB, Jia XM, Yang JH, Ling YC, Zhang Z, Yu J, Wu JY, Xiao JF. PanGP: a tool for quickly analyzing bacterial pan-genome profile. *Bioinformatics*, 2014, 30(9): 1297–1299.
- [24] Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in

- performance and usability. *Molecular Biology and Evolution*, 2013, 30(4): 772–780.
- [25] Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 2015, 32(1): 268–274.
- [26] Sanderson MJ. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics: Oxford, England*, 2003, 19(2): 301–302.
- [27] Tian XP, Zhang ZW, Yang TT, Chen ML, Li J, Chen F, Yang J, Li WJ, Zhang B, Zhang Z, Wu JY, Zhang CS, Long LJ, Xiao JF. Comparative genomics analysis of *Streptomyces* species reveals their adaptation to the marine environment and their diversity at the genomic level. *Frontiers in Microbiology*, 2016, 7: 998.
- [28] Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, 2019, 20(1): 238.
- [29] De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics: Oxford, England*, 2006, 22(10): 1269–1271.
- [30] 李纯, 孙春玉, 陈静, 林彦萍, 王义, 张美萍. 主要协同转运蛋白超家族的研究进展. *生物技术通报*, 2018, 34(8): 43–49.
Li C, Sun CY, Chen J, Lin YP, Wang Y, Zhang MP. Research advances in the major facilitator superfamily. *Biotechnology Bulletin*, 2018, 34(8): 43–49. (in Chinese)
- [31] Schatz PJ, Riggs PD, Jacq A, Fath MJ, Beckwith J. The *secE* gene encodes an integral membrane protein required for protein export in *Escherichia coli*. *Genes & Development*, 1989, 3(7): 1035–1044.
- [32] Lycklama a Nijeholt JA, de Keyzer J, Prabudiansyah I, Driessen AJM. Characterization of the supporting role of SecE in protein translocation. *FEBS Letters*, 2013, 587(18): 3083–3088.

(本文责编 张晓丽)