



多位点序列和比较基因组学分析对盐单胞菌科菌株 JSM 104105 的系统发育学研究

陈锦华¹, 刘细寒^{1,2}, 邓丽颖^{1,3}, 冯玉周⁴, 刘祝祥^{1*}, 陈义光^{1*}

1 吉首大学生物资源与环境科学学院, 湖南 吉首 416000

2 吉首大学医学院, 湖南 吉首 416000

3 吉首大学张家界学院, 湖南 张家界 427000

4 四川大学华西医院 华西生物医学大数据中心, 四川 成都 610065

陈锦华, 刘细寒, 邓丽颖, 冯玉周, 刘祝祥, 陈义光. 多位点序列和比较基因组学分析对盐单胞菌科菌株 JSM 104105 的系统发育学研究[J]. 微生物学报, 2023, 63(2): 683-699.

CHEN Jinhua, LIU Xihan, DENG Liying, FENG Yuzhou, LIU Zhuxiang, CHEN Yiguang. Phylogenetics of *Halomonadaceae* strain JSM 104105 based on multilocus sequence analysis and comparative genomics analysis[J]. Acta Microbiologica Sinica, 2023, 63(2): 683-699.

摘要:【目的】为了精准地判定产铁载体嗜盐新菌株 JSM 104105 在盐单胞菌科(*Halomonadaceae*) 中的系统发育地位。【方法】采用基于 16S rRNA 基因、DNA 促旋酶 B 亚基基因(DNA gyrase B subunit gene, *gyrB*)和核糖核酸聚合酶 σ 因子 RpoD 基因(RNA polymerase sigma factor RpoD gene, *rpoD*)序列的多位点序列分析方法(multilocus sequence analysis, MLSA), 对菌株 JSM 104105 的系统发育地位进行初步分析; 采用比较基因组学分析(comparative genomics analysis), 分析该菌株与其系统发育关系密切的典型菌株之间的 GC 含量、平均核苷酸一致性(average nucleotide identity, ANI)和数字 DNA-DNA 杂交值(digital DNA-DNA hybridization, dDDH), 并进行基因组系统发育学分析(phylogenomic analysis), 更准确地判定菌株 JSM 104105 在盐单胞菌科中的系统发育地位。【结果】MLSA 分析结果表明, 无论是单基因序列分析, 还是多基因串联序列分析, 对盐单胞菌科各分类单元以及菌株 JSM 104105 的系统发育地位都能提供较为一致的结果。分析表明, 菌株 JSM 104105 归属于盐单胞菌科盐单胞菌属(*Halomonas*), 与该属的 *Halomonas gudaonensis*、*Halomonas azerbaijanica* 和 *Halomonas lysinitropha* 的系统发育关系较密切, 它们在系统进化树上形成稳定亚簇(subcluster), 其中菌株 JSM 104105 占据稳定的独立进化分支。尽管它们之间的 16S rRNA 基因

资助项目: 国家自然科学基金(31460004, 31360062); 湖南省研究生科研创新项目(CX20211062)

This work was supported by the National Natural Science Foundation of China (31460004, 31360062) and the Foundation of Hunan Graduate Research Innovation Project (CX20211062).

*Corresponding authors. E-mail: LIU Zhuxiang, liuzhuxiang@126.com; CHEN Yiguang, mchen@jzu.edu.cn, mchenjzu@aliyun.com

Received: 2022-06-08; Accepted: 2022-07-12; Published online: 2022-07-14

序列相似性较高, 为 97.3%–98.9%, 但菌株 JSM 104105 很难归属于其中任何已知物种。比较基因组学分析结果确认了这一判断。菌株 JSM 104105 与系统发育关系密切的 *H. gudaonensis* CGMCC 1.16133^T、*H. azerbaijanica* TBZ202^T 和 *H. lysinitropha* 3(2)^T 的 ANI 值(78.9%–91.6%)和 dDDH 值(22.1%–43.7%), 均显著低于界定原核生物物种的阈值(ANI, 95%–96%; dDDH≥70%)。基因组系统发育分析结果表明, 菌株 JSM 104105 明确构成盐单胞菌属独立的亚分支(subclade)。

【结论】从分子系统发育学视角精准地判定产铁载体嗜盐细菌 JSM 104105 归属于盐单胞菌科盐单胞菌属, 与该属的已知物种 *H. gudaonensis*、*H. azerbaijanica* 和 *H. lysinitropha* 的系统发育关系最密切, 但不能归属于该属的已知种, 应该代表了一个新的基因种(genospecies)。

关键词: 盐单胞菌科; 多位点序列分析; 比较基因组学分析; 基因组系统发育学; 基因组特征

Phylogenetics of *Halomonadaceae* strain JSM 104105 based on multilocus sequence analysis and comparative genomics analysis

CHEN Jinhua¹, LIU Xihan^{1,2}, DENG Liying^{1,3}, FENG Yuzhou⁴, LIU Zhuxiang^{1*}, CHEN Yiguang^{1*}

1 College of Biology and Environmental Sciences, Jishou University, Jishou 416000, Hunan, China

2 School of Medicine, Jishou University, Jishou 416000, Hunan, China

3 Zhangjiajie College, Jishou University, Zhangjiajie 427000, Hunan, China

4 West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu 610065, Sichuan, China

Abstract: **[Objective]** To investigate the phylogenetic position of the siderophore-producing halophilic bacterium JSM 104105 in the family *Halomonadaceae*. **[Methods]** Molecular phylogenetic analysis of strain JSM 104105 was performed by means of multilocus sequence analysis (MLSA) based on 3 housekeeping genes, i.e., 16S rRNA gene, DNA gyrase B subunit gene (*gyrB*) and RNA polymerase sigma factor RpoD gene (*rpoD*), and then its exact systematic position was investigated by comparative genomics analysis based on whole-genome sequences, including comparisons of G+C content, average nucleotide identity (ANI), and digital DNA–DNA hybridization (dDDH) values, and phylogenomic analysis. **[Results]** The MLSA results showed that the data both from single-gene (16S rRNA, *gyrB*, or *rpoD*) sequence comparisons and concatenated 16S rRNA-*gyrB*-*rpoD* sequence comparisons were consistent, and could indicate the phylogeny of strain JSM 104105 and related taxa of *Halomonadaceae*, and that strain JSM 104105 belonged to the genus *Halomonas*, being in closely genetic relationship with *Halomonas gudaonensis* CGMCC 1.16133^T (16S rRNA gene similarity, 98.9%), *Halomonas azerbaijanica* TBZ202^T (98.6%) and *Halomonas lysinitropha* 3(2)^T (97.3%), and strain JSM 104105 formed a coherent branch in the phylogenetic trees which were generated based on either single-gene sequence or concatenated-gene sequence comparisons. MLSA data also suggested that strain JSM 104105 could represent a different phylogenetic subline separated from recognized *Halomonas* species. The results of comparative genomics analysis supported the conclusion from

MLSA data. Either ANI (78.9%–91.6%) or dDDH values (22.1%–43.7%) were all well below the standard criteria (ANI, 95%–96%; dDDH \geq 70%) for delineation of prokaryotic species. The results of phylogenomic analysis also showed that strain JSM 104105 clearly represented an independent subclade in *Halomonas*. **[Conclusion]** From the perspective of molecular phylogeny, the combination of MLSA and comparative genomics analysis demonstrated definitely that strain JSM 104105 belongs to *Halomonas*, being phylogenetically closely related to *H. gudaonensis*, *H. azerbaijanica* and *H. lysinitropha*, and that strain JSM 104105 cannot be assigned to any recognized species but represents a novel genospecies of the genus *Halomonas*.

Keywords: *Halomonadaceae*; multilocus sequence analysis; comparative genomics analysis; phylogenomics; whole-genome feature

原核生物核糖体小亚基RNA基因(16S rRNA基因), 因为具有长度适中(约1 500 bp)、保守性较强、又有9个高可变区、能用通用引物进行扩增和测序等特点, 其基因序列已经成为目前最为主要的和必需的系统发育分析的分子标记物, 适合于各级分类单元(taxa)的系统发育分析。基于16S rRNA基因序列的系统发育分析结果, 连同基因组DNA-DNA同源性(DNA-DNA relatedness), 一直以来被普遍接受为原核生物物种分子鉴定的金标准^[1]。但是, 16S rRNA基因存在基因异质性^[2], 对原核微生物多样性的估计有偏差, 尤其在极端环境来源的微生物中, 该基因存在着更高的分歧^[3-5]。为减少这种偏差, 以便更准确地进行原核生物系统发育学研究, 综合包括16S rRNA基因在内的多个管家基因(housekeeping gene)序列的多位点序列分析(multilocus sequence analysis, MLSA), 已经受到较广泛的关注和应用^[6-8]。例如23S rRNA基因、DNA促旋酶B亚基基因(DNA gyrase B subunit gene, *gyrB*)、核糖核酸聚合酶 σ 因子RpoD基因(RNA polymerase sigma factor RpoD gene, *rpoD*)和ATP合成酶F1 β 亚基基因(ATP synthase F1 beta subunit gene, *atpD*)等, 被建议用作新的系统发育标记物, 用于MLSA分析, 以期提供更多信息用于确定关系密切的类群(物种)之间的

系统发育关系^[6,9-15]。

菌株间基因组DNA-DNA杂交(DNA-DNA hybridization)是菌株间基因组的两两比较, 能反映整个基因组序列间的遗传相关性(DNA-DNA同源性)。普遍认为同一物种的菌株间DNA-DNA同源性大于或等于70%^[16]。但是, 传统的DNA-DNA杂交试验技术要求高、耗时长, 且不同实验人员和不同实验方法所得结果重现性差^[17]。随着测序技术的进步, 全基因组序列数据快速增加, 基于全基因组序列的比较基因组学分析(comparative genomics analysis), 有助于更系统地理解原核生物类群的发生、演化, 以及不同生境类群之间的相互关系^[18]。基于全基因组序列的菌株间相关性评估(estimation of genomic relatedness using genome sequencing data)等多种比较基因组学新技术得以开发和应用, 例如平均核苷酸一致性(average nucleotide identity, ANI)和数字DNA-DNA杂交(digital DNA-DNA hybridization, dDDH), 其结果已经被采用为原核生物物种间分子鉴定新的金标准^[16,19]。

盐单胞菌科(*Halomonadaceae*)成员之间, 由于其16S rRNA基因序列的高度保守性, 16S rRNA基因序列分析法有时无法清晰地阐明关系密切的类群(物种)之间的系统发育关系。de la

Haba等采用多基因序列(16S rRNA、23S rRNA、*atpA*、*gyrB*、*rpoD*和*secA*)的MLSA分析,对盐单胞菌科52个代表性物种进行了单基因序列分析和多基因串联序列分析,表明其中的16S rRNA、*gyrB*和*rpoD*基因是最为合适的用于盐单胞菌科进行MLSA分析的分子标记^[13]。本实验室在调查非盐环境嗜盐和极端耐盐细菌多样性期间^[20-22],从湖南省吉首市一份健康成人粪便样品中分离到一株产铁载体的嗜盐细菌JSM 104105。基于16S rRNA基因序列的系统发育分析结果表明该菌株属于盐单胞菌科盐单胞菌属(*Halomonas*)。本研究采用基于16S rRNA基因、*gyrB*和*rpoD*基因序列的MLSA方法,对菌株JSM 104105的系统发育学地位进行初步分析,进一步采用比较基因组分析,估算菌株JSM 104105与其系统发育关系密切的典型菌株之间的GC含量、ANI值和dDDH值,并进行基于基因组序列的系统发育分析,以期更精准地判定该菌株在盐单胞菌科中的系统发育地位。

1 材料与方 法

1.1 菌株来源

菌株 JSM 104105 由本实验室采用补充 5% (质量体积分数) NaCl 的 Marine Agar 2216 (Difco) 平板分离自健康成人粪便样本。经划线纯化后,采用斜面、甘油管和牛奶管在本实验室进行了妥善保存,并保藏在中国典型培养物

保藏中心 (China Center for Type Culture Collection; 保藏号: CCTCC AB 2014160)、中国普通微生物菌种保藏中心 (China General Microbiological Culture Collection Center; CGMCC 1.12945) 和韩国典型培养物保藏中心 (Korean Collection for Type Cultures; KCTC 42146)。菌株 JSM 104105 轻度嗜盐 (slightly halophilic), 在含有多种微量盐的半合成培养基中需要添加 0.5%–18% (质量体积分数) NaCl (最适 4%–9%); 铬蓝琼脂平板法检测到该菌株具有产铁载体活性。

1.2 管家基因和全基因组序列的获得

1.2.1 菌株 JSM 104105 管家基因序列测定

菌株 JSM 104105 用补充 5% (质量体积分数) NaCl 的 Marine Agar 2216 (pH 7.5) 平板于 30 °C 培养 5 d, 挑取菌体, 按文献推荐的方法进行基因组 DNA 提取和 16S rRNA 基因的 PCR 扩增^[23]。基因 *gyrB* 和 *rpoD* 基因的 PCR 扩增引物和扩增程序参照 de la Haba 等^[13]的方法进行改进(表 1)。所得 PCR 产物由生工生物工程(上海)股份有限公司用相同引物进行序列测定。所得序列用 DNASTar 软件包中的 SeqMan 进行质量检测 and 序列拼接, 并对首尾端序列进行手工校正和适度切除。所得 16S rRNA 基因序列在 NCBI 网站采用 GenBank 上传工具上载, 所得 *gyrB* 和 *rpoD* 基因序列采用 BankIt 上载(<https://submit.ncbi.nlm.nih.gov/>), 获得 GenBank 基因序列号(accession numbers)。

表 1 菌株 JSM 104105 的 *gyrB* 和 *rpoD* 基因 PCR 扩增引物和程序

Table 1 Primers used for PCR amplification and sequencing and PCR cycling conditions for *gyrB* and *rpoD* of strain JSM 104105

Gene	Primer	Sequence (5'→3')	Position*	PCR cycling
<i>gyrB</i>	gyrB216F	GAR GTB ATC ATG ACS GTG CT	216–235	5 min, 95 °C; 35×(1 min, 94 °C; 1 min, 55 °C; 1 min 30 s, 68 °C); 10 min, 68 °C
	gyrB1419R	GCR TCS GTC ATG ATG ATS AY	1 438–1 419	
<i>rpoD</i>	rpoD88F	ATG ATY AAC GAC ATG GGY AT	88–107	5 min, 95 °C; 35×(1 min, 94 °C; 1 min, 50 °C; 1 min 30 s, 68 °C); 10 min, 68 °C
	rpoD1321R	TTS AKC TTR TTG ATG GTC TC	1 340–1 321	

*: Position relative to the sequences taken from the complete genome of *Escherichia coli* K-12^[24-25].

1.2.2 菌株 JSM 104105 全基因组序列测定

菌株 JSM 104105 用补充 5% (质量体积分数) NaCl 的 Marine Broth 2216 (pH 7.5) 于 30 °C、180 r/min 摇瓶培养 3–4 d。4 000 r/min 离心收集菌体, 用双蒸水洗涤 2 次。用细菌基因组 DNA 提取试剂盒(Solarbio)进行 DNA 分离和纯化。用紫外分光光度法检测 DNA 纯度, 要求 A_{260}/A_{280} 达到 1.7–2.0。全基因组测序由生工生物工程(上海)股份有限公司在 HiSeq XTen 平台 (Illumina) 进行。用 FastQ Screen 进行质量控制^[26], 用 Trimmomatic 0.36 对接头和低质量读长进行整理过滤^[27], 用 SOAPdenovo 进行组装^[28]。

1.2.3 参考菌株的选择和序列下载

用测序所得菌株 JSM 104105 的 16S rRNA 基因序列在 EzBioCloud 在线服务(<http://www.ezbiocloud.net/identify>)^[29]进行典型菌株序列相似性搜索, 把结果中全部 50 个典型菌株中有 *gyrB* 和 *rpoD* 基因序列数据、或有基因组数据的菌株列为主要的参考菌株。在 NCBI 网站搜索下载这些菌株的 16S rRNA、*gyrB* 和 *rpoD* 基因序列和全基因组序列 (<https://www.ncbi.nlm.nih.gov/>)。盐单胞菌属典型种 *Halomonas elongata* 和部分其他盐单胞菌科典型菌株也列为参考菌株, 根据 de la Haba 等^[13]提供的 GenBank/EMBL/DDBJ 序列号从 NCBI 网站下载其相应管家基因序列, 并从 NCBI 网站搜索和下载其基因组数据。无效种(effectively but not validly published) '*Halomonas sambharensis*'^[30]也被列为参考菌株之一, 下载其相应的管家基因和基因组序列。

1.3 管家基因注释

根据基于 16S rRNA 基因序列系统发育分析结果, 与菌株 JSM 104105 系统发育关系较密切的典型菌株中没有现成的 *gyrB* 和 *rpoD* 基因

序列数据、但有基因组数据的, 采用原核基因组注释工具 DFAST^[31]提取 *gyrB* 和 *rpoD* 基因序列, 用于多位点序列分析。同时, 采用 RNAmmer^[32]分别从菌株 JSM 104105 和 '*H. sambharensis* SBS 10' 的基因组中提取其 16S rRNA 基因序列, 用于对比研究。

1.4 基于多位点序列的系统发育分析

采用基于 3 个管家基因序列(16S rRNA、*gyrB* 和 *rpoD*)的多位点序列分析法(MLSA)对菌株 JSM 104105 在盐单胞菌科中的系统发育地位进行分析。菌株 JSM 104105 及盐单胞菌科相关典型菌株的 16S rRNA 基因序列两两比对和相似性计算采用 EzBioCloud 在线服务^[29]。分别构建基于 16S rRNA 基因、*gyrB* 和 *rpoD* 基因的单基因序列集, 使用 CLUSTAL_X 进行序列多重比对^[33], 首尾切除后保存为 .aln 文件, 然后用 BioEdit 转换为 fasta 格式文件。多基因串联分析时, 用 MUSCLE v3.8.31 软件对各单基因序列集进行序列多重比对^[34], 用 PAUP*v.4.0b10 软件进行序列串联(concatenate)^[35]。进化距离(evolutionary distance)采用 Kimura 双参数方法计算^[36], 采用 MEGA 7^[37]中的邻接法(neighbor-joining method)^[38]进行进化距离聚类分析和系统发育树(NJ 树)构建。重复取样 1 000 次进行自举值(bootstrap value)计算来评估进化树的拓扑结构稳定性^[39]。

1.5 比较基因组学分析

1.5.1 基因组特征分析

菌株 JSM 104105 的基因组 DNA G+C 含量从其基因组序列计算获得。采用 Barrnap 0.9 扫描其基因组检测 rRNA 基因(5S, 16S, 23S), 采用 tRNAscan-SE2.0.7 进行 tRNA 基因预测^[40]。依次在单基因簇与非冗余蛋白质序列(RefSeq non-redundant protein sequence, NR)数据库、直系同源簇(clusters of orthologous groups, COG)

数据库、Swiss-Prot 数据库和京都基因和基因组百科全书(Kyoto encyclopedia of genes and genomes, KEGG)数据库中进行蛋白质编码基因注释。

1.5.2 全基因组序列相关性评估

采用 EzBioCloud 在线网络服务(<https://www.ezbiocloud.net/tools/ani>)按 OrthoANI 算法计算菌株 JSM 104105 和系统发育密切的典型菌株基因组间的 ANI 值^[41-42]。菌株基因组间的 dDDH 值利用 DSMZ 网络平台服务(<https://ggdc.dsmz.de/distcalc2.php>)采用 genome-to-genome distance calculator (GGDC) 3.0 基于 formula 2 (identities/HSP length)估算^[43]。对照目前普遍采纳的原核生物有关基于全基因组的分子分类标准^[16,19]，对所得结果进行综合分析。

1.5.3 基因组系统发育学分析

采用 Type Strain Genome Server 在线服务 (TYGS, <https://tygs.dsmz.de>)进行基于全基因组序列的系统发育学分析^[43-44]。为了选择高度相关的典型菌株，采用 MASH 算法把菌株 JSM 104105 基因组与 TYGS 数据库中所有典型菌株基因组进行比较^[45]，并用从菌株 JSM 104105 基因组注释得到的 16S rRNA 基因序列^[32]同 TYGS 数据库中所有典型菌株的 16S rRNA 基因序列进行 BLAST 分析^[46]。采用 Genome BLAST Distance Phylogeny (GBDP) d5 进化距离模型，

分别用‘coverage’和‘trimming’算法计算典型菌株 16S rRNA 基因序列之间以及基因组之间的 GBDP 进化距离(100 次重复)^[47]。物种水平聚类(species clustering)采用 70% dDDH 阈值^[16,44]，亚种水平聚类(subspecies clustering)采用 79% dDDH 阈值^[48]。采用 FASTME 2.1.6.1 构建最小进化距离系统发育树(minimum evolution tree)^[49]。

2 结果与分析

2.1 管家基因和基因组序列的获得

测定了菌株 JSM 104105 的 3 个管家基因序列，即 16S rRNA 基因(登录号：KM199859)、*gyrB* (ON461989)和 *rpoD* (ON461986)，以及全基因组序列(GCA_010993975.1)；从 7 株密切相关参考典型菌株的相应基因组注释得到了其 *gyrB* 和 *rpoD* 基因序列(ON571520–ON571533，表 2)；从菌株 JSM 104105 和‘*H. sambharensis* SBS 10’ (GCA_003973725.1)的基因组序列中注释得到其 16S rRNA 基因序列，序列号分别为 ON556670 和 ON556671；其它参考典型菌株的管家基因序列从 NCBI 网站下载(序列号见系统发育树)。试图从无效种‘*H. sambharensis* SBS 10’基因组中注释其 *gyrB* 和 *rpoD* 基因序列，但所得序列很短，分别只有 749 bp 和 350 bp (未列出)，既不能申请序列号，更影响序列分析的整体质量，没有被用于后续的 MLSA 研究。

表 2 从部分典型菌株的基因组注释得到的 *gyrB* 和 *rpoD* 基因序列

Table 2 Accession numbers of *gyrB* and *rpoD* sequences annotated from their genomes

Strain	Genome	<i>gyrB</i>	<i>rpoD</i>
<i>Halomonas azerbaijanica</i> TBZ202 ^{T[50]}	GCA_004551485.1	ON571520	ON571527
<i>Halomonas lysinitropha</i> 3(2) ^{T[51]}	GCA_902500215.1	ON571521	ON571528
<i>Halomonas cerina</i> CECT7282 ^{T[52]}	GCA_014192215.1	ON571522	ON571529
<i>Halomonas pellis</i> L5 ^{T[53]}	GCA_008297955.1	ON571523	ON571530
<i>Halomonas kenyensis</i> DSM17331 ^{T[54]}	GCA_013697085.1	ON571524	ON571531
<i>Halomonas saliphila</i> LCB169 ^{T[55]}	GCA_002930105.1	ON571525	ON571532
<i>Halomonas organivorans</i> CECT5995 ^{T[56]}	GCA_014192055.1	ON571526	ON571533

2.2 多位点序列分析(MLSA)

采用 EzBioCloud 在线服务, 用菌株 JSM 104105 的 16S rRNA 基因序列进行 BLAST, 进行了序列两两比对和相似性计算。分别构建了菌株 JSM 104105 及与之系统发育关系密切的盐单胞菌科参考典型菌株的 16S rRNA 基因、*gyrB* 和 *rpoD* 的单基因序列集。16S rRNA 基因序列分析也包括无效种‘*H. sambharensis*’。进行了多重序列比对和进化树构建, 分别得到了用邻接法(neighbor-joining method)构建的基于 16S rRNA 基因(图 1A)、*gyrB*(图 1B)、*rpoD* 单基因序列(图 1C)和串联基因序列(图 1D)的系统发育树(NJ 树)。

2.2.1 基于 16S rRNA 基因序列的系统发育分析

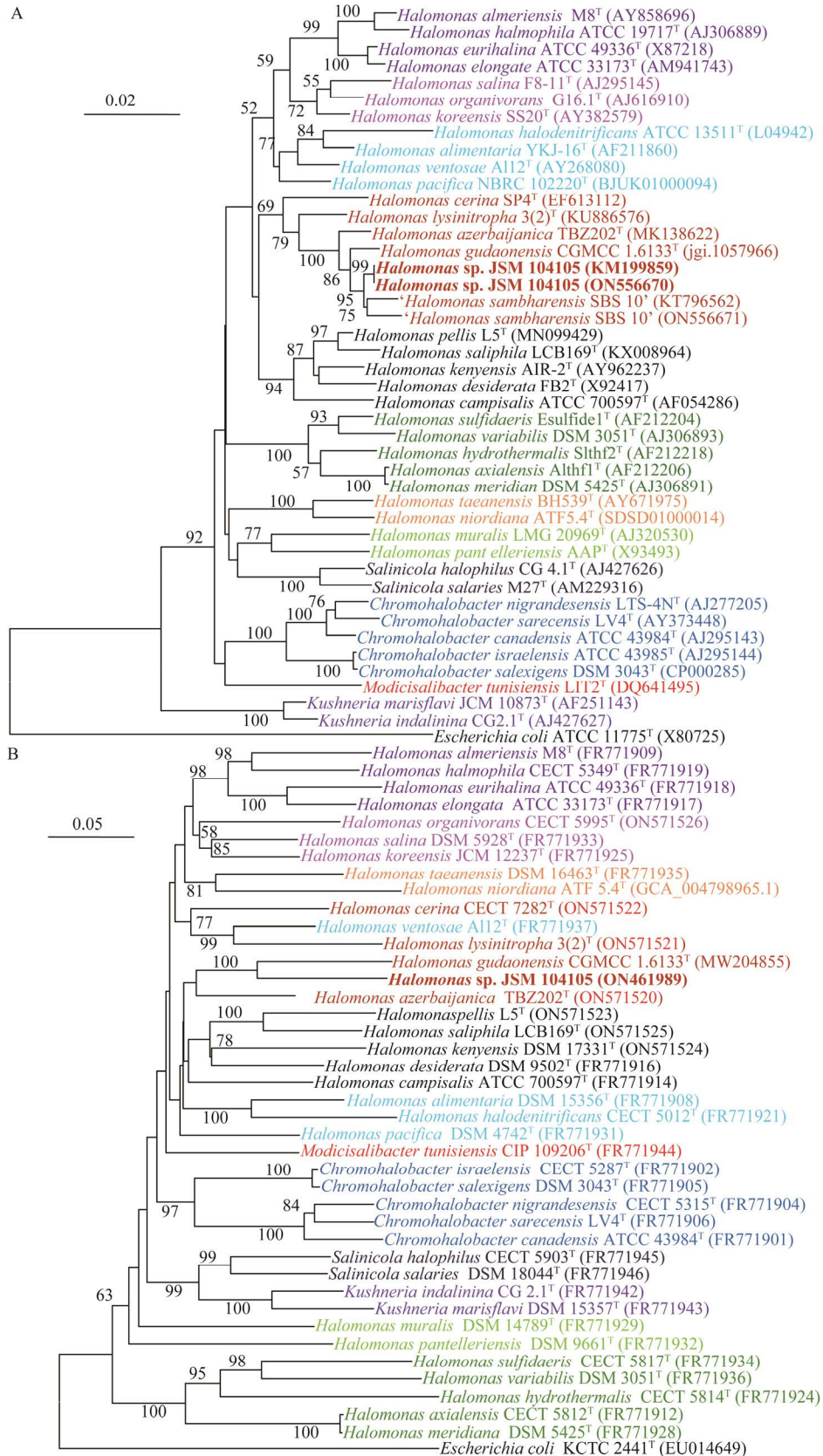
基于 16S rRNA 基因序列的系统发育分析表明, 菌株 JSM 104105 属于盐单胞菌科盐单胞菌属, 与‘*H. sambharensis* SBS 10’的序列相似性最高(99.6%), 其余依次为 *H. gudaonensis* CGMCC 1.6133^{T[57]} (98.9%)、*H. azerbaijanica* TBZ202^T (98.6%)、*H. lysinitropha* 3 (2)^T (97.3%)和 *H. cerina* CECT7282^T (96.5%)。菌株 JSM 104105 与盐单胞菌属其他典型菌株的 16S rRNA 基因序列相似性小于 96.5%。从基于 16S rRNA 基因序列的 NJ 树(图 1A)可以看出, 盐单胞菌科的 *Salinicola*、*Chromohalobacter* 和 *Kushneria* 属分别构成稳定的簇群(cluster); *Modicisalibacter* 属唯一已知种 *M. tunisiensis* 和 *Chromohalobacter* 属松散地聚在一起。盐单胞菌属为盐单胞菌科的主体类群, 其典型菌株形成了 7 个较稳定的簇群(图 1)。菌株 JSM 104105 与盐单胞菌属 4 个有效种(validly published) *H. gudaonensis*、*H. azerbaijanica*、*H. lysinitropha* 和 *H. cerina* 以及无效种‘*H. sambharensis*’在进化树上聚为一簇, 其中与‘*H. sambharensis*’、*H. gudaonensis* 和 *H. azerbaijanica* 聚类较为稳定, 但菌株 JSM

104105 处于一个独立进化分支(bootstrap value, 95%)。

进一步分析发现, 从菌株 JSM 104105 的 PCR 产物测序所得的 16S rRNA 基因序列(KM199859)和从其基因组(GCA_010993975.1)注释所得相应序列(ON556670)的相似性为 100%, 在进化树上也紧密相聚(图 1A), 说明这些数据是一致的。但‘*H. sambharensis* SBS 10’的两条 16S rRNA 基因序列(KT796562, ON556671)之间的相似性却只有 98.9%, 二者在 NJ 树上也处于离散状态(图 1A), 提示‘*H. sambharensis* SBS 10’的 PCR 产物测序所得的 16S rRNA 基因序列(KT796562)或基因组提取数据, 至少有一个严重失真, 不足采信。

2.2.2 基于单基因和基因串联序列的综合分析

与基于 16S rRNA 基因序列的系统发育分析结果对比, 发现盐单胞菌科中各属在基于 *gyrB*(图 1B)或 *rpoD*(图 1C)或 3 基因串联 16S rRNA-*gyrB*-*rpoD*(图 1D)序列的 NJ 树的聚类情况基本相似, 其中 3 基因串联树(图 1D)的聚类情况最为接近, 其次为 *gyrB* 树(图 1B), 聚类相对不稳定的是 *rpoD* 树(图 1C)。最不稳定的有 2 个菌株, 即 *M. tunisiensis* LIT2^T 和 *Halomonas pacifica* NBRC 102220^T。在基于 16S rRNA 基因序列的 NJ 树上, 它们分别松散地聚在其他菌株的外围, 而在其他 3 种树上, 它们彼此关联地离散分布(图 1B, 1D)或松散地与其他菌株相聚(图 1C); 其次是 *Halomonas muralis* LMG 20969^T 和 *Halomonas pantelleriensis* AAP^T。它们彼此关联, 但与其他菌株的聚类关系在每个树上都不一样; 较不稳定的还有 2 个与菌株 JSM 104105 序列相似性较高的菌株 *H. lysinitropha* 3(2)^T 和 *H. cerina* SP4^T, 它们在其他 3 种树中都不再与菌株 JSM 104105 相聚, 其中 *H. lysinitropha* 3(2)^T 常与 *Halomonas ventosae* A112^T 紧密相聚,



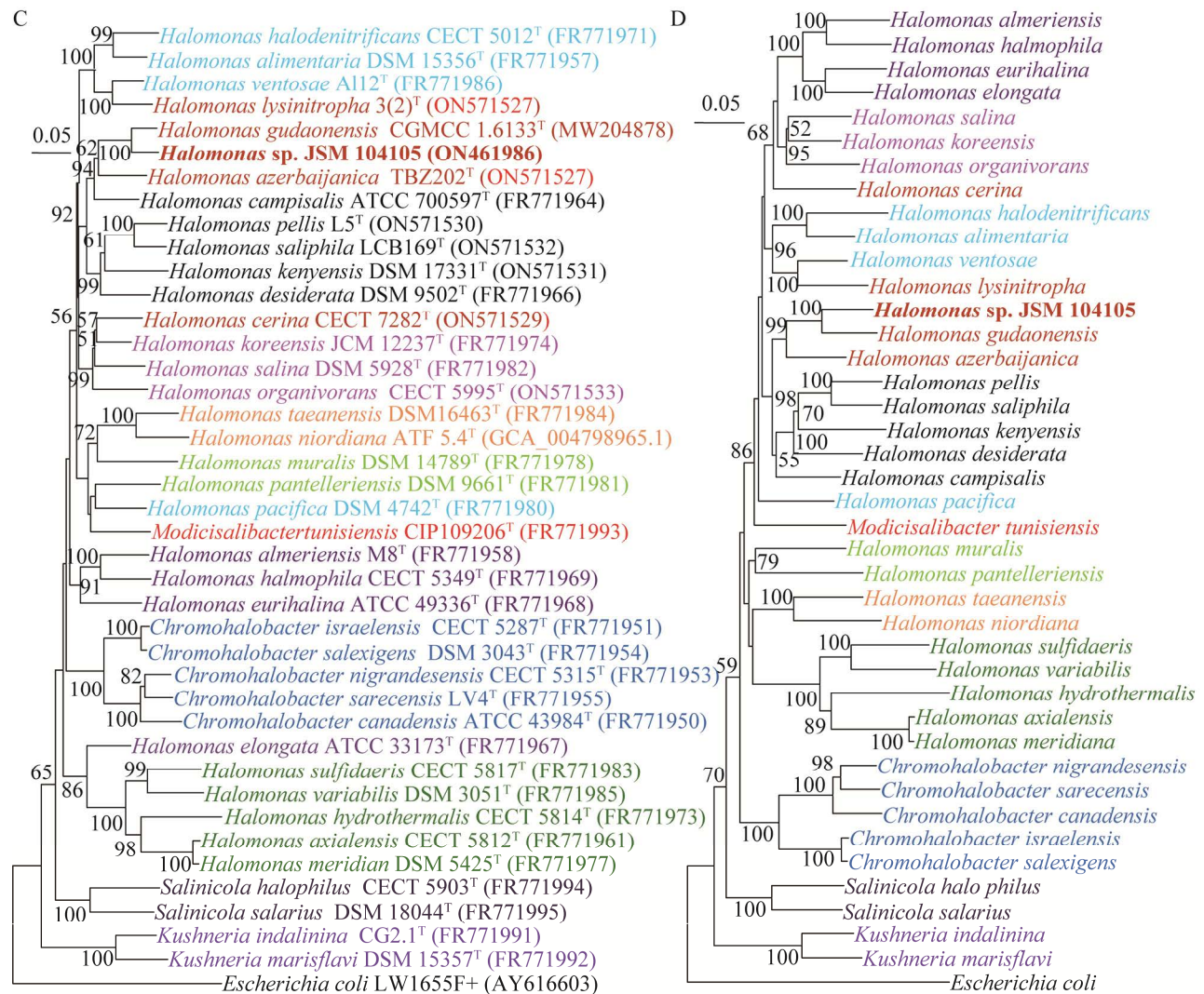


图 1 用邻接法构建的基于单基因序列(16S rRNA, A; *gyrB*, B; *rpoD*, C)和 16S rRNA-*gyrB*-*rpoD* 3 基因串联序列(D)的系统发育树

Figure 1 NJ trees based on the 16S rRNA (A), *gyrB* (B), *rpoD* (C) and concatenated 16S rRNA-*gyrB*-*rpoD* (D) gene sequences showing the phylogenetic relationships between members of the family Halomonadaceae and the phylogenetic position of strain JSM 104105 among related taxa. Bootstrap values (>50%) based on 1 000 resampled datasets are shown at branch nodes. Bar, 0.02 or 0.05 substitutions per nucleotide position.

而 *H. cerina* SP4^T 则在其他 3 种树中都处于离散状态; 盐单胞菌属典型种 *H. elongata* 在 3 种树上一直非常稳定地与 *Halomonas eurihalina* ATCC 49336^T 等 4 个典型菌株聚在一起, 但在 *rpoD* 树上则离开了原来的簇群, 而是在 *Halomonas sulfidaeris* CECT 5817^T 等 5 个菌株形成簇群的外围占据一个独立分支(图 1C)。

特别需要指出的是, 菌株 JSM 104105 在 4 个 MLSA 系统进化树中均与 *H. gudaonensis* CGMCC 1.6133^T 和 *H. azerbaijanica* TBZ202^T 形成稳定的亚簇, 其中菌株 JSM 104105 始终占据一个稳定分支(图 1), 说明这些菌株之间的系统发育关系最为密切, 但菌株 JSM 104105 不能归属于任何已知物种。

2.3 比较基因组学分析

2.3.1 菌株 JSM 104105 基因组基本特征

SOAPdenovo 组装获得菌株 JSM 104105 基因组(draft genome)共 59 contigs (N50: 218, 317 bp), 测序覆盖度为 57 \times 。菌株 JSM 104105 基因组全长 4.2 Mbp, 被上载到 DDBJ/ENA/GenBank 获得序列号为 AACZO000000000。从菌株 JSM 104105 基因组测得其 DNA G+C 含量为 64.5 mol%。经 tRNA-scan-SE 和 Barrnap 鉴定, 该基因组含有 67 个 tRNA 和 7 个 rRNA regions (16S rRNA 1, 5S rRNA 5, 23S rRNA 1), 组成 63 个 tRNA 基因和 3 个 rRNA 基因。Barrnap 注释结果显示, 菌株 JSM 104105 基因组含 1 个 16S rRNA 基因, 注释所得的 2 个 16S rRNA 基因片段分别覆盖该基因左端 1–798 bp 和右端 988–1 467 bp, 它们与从该菌株基因组 DNA 直接扩增所得 16S rRNA 基因序列(KM199859)的相似性为 100%。共从 NR、KEGG、Swissprot 和 COG 数据库注释获得 3 788 个蛋白质编码候选基因(candidate protein-coding sequences, CDSs), 详细情况见图 2。

2.3.2 基于全基因组序列的基因组相关性

比较基因组学分析结果表明, 菌株 JSM 104105 和系统发育关系最密切的盐单胞菌属典型菌株 *H. gudaonensis* CGMCC 1.6133^T 的 ANI

值为 91.6%, 非常明显地低于普遍认可的细菌物种划分的阈值(95%–96%)^[19], 这 2 个菌株之间的 dDDH 估值为 43.7%, 也明显低于原核生物物种区分的标准尺度(70%)^[16,44]。菌株 JSM 104105 与包括盐单胞菌属典型种 *H. elongata* 和无效种 '*H. sambharensis*' 在内的其他系统发育关系较密切菌株之间 ANI 值和 dDDH 值均显著低于以上界定原核生物物种的阈值(表 3)。所以, 基于比

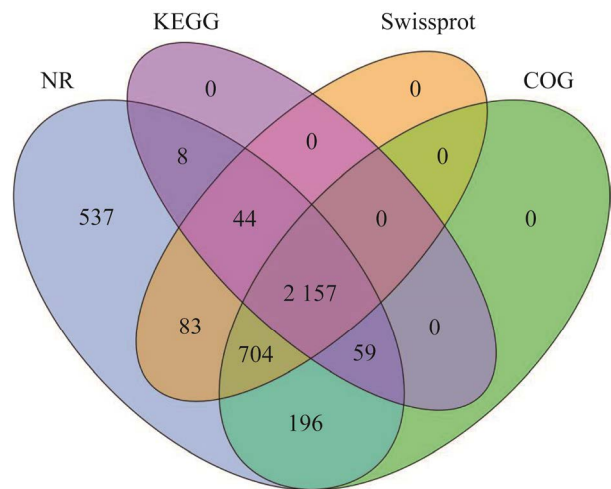


图 2 菌株 JSM 104105 基因组中注释蛋白质编码基因在公共数据中的分布维恩图解

Figure 2 Venn diagram of CDSs annotated from the genome of strain JSM 104105 assigned to NR, KEGG, Swissprot or COG databases.

表3 菌株 JSM 104105 和系统发育关系密切的盐单胞菌属典型菌株的基因组比较

Table 3 Pairwise comparisons of genomes of strain JSM 104105 vs. related type strains of the genus *Halomonas*

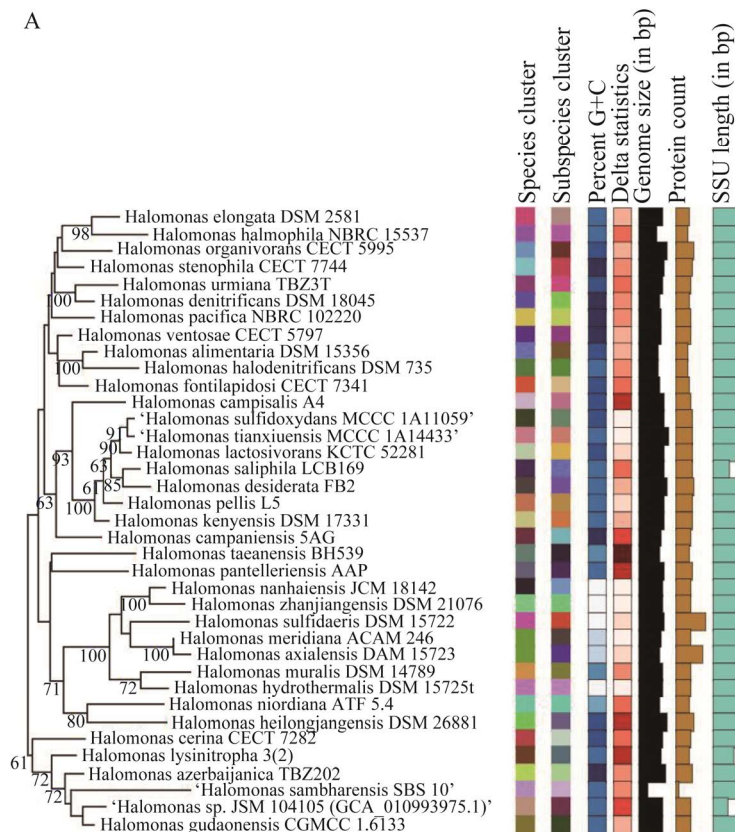
Strains	Accession number	GC content (mol%)	ANI (%)	dDDH (formula 2; %)
<i>Halomonas gudaonensis</i> CGMCC 1.6133 ^T	GCA_900100195.1	64.9	91.6	43.7 (41.2–46.3)
<i>Halomonas azerbaijanica</i> TBZ202 ^T	GCA_004551485.1	67.5	83.7	26.8 (24.4–29.3)
<i>Halomonas lysinitropha</i> 3(2) ^T	GCA_902500215.1	64.8	78.9	22.1 (19.8–24.5)
<i>Halomonas cerina</i> CECT7282 ^T	GCA_014192215.1	65.7	78.6	22.1 (19.8–24.5)
<i>Halomonas ventosae</i> CECT 5797 ^T	GCA_004363555.1	67.5	79.6	22.9 (20.6–25.3)
<i>Halomonas pellis</i> L5 ^T	GCA_008297955.1	63.6	78.3	21.9 (19.6–24.3)
<i>Halomonas alimentaria</i> DSM 15356 ^T	GCA_009902005.1	66.3	79.5	22.6 (20.3–25.0)
<i>Halomonas kenyensis</i> DSM 17331 ^T	GCA_013697085.1	63.8	78.3	22.0 (21.9–24.4)
<i>Halomonas pacifica</i> NBRC 102220 ^T	GCA_007989625.1	67.2	78.1	21.9 (19.6–24.3)
<i>Halomonas saliphila</i> LCB169 ^T	GCA_002930105.1	64.1	78.8	22.1 (19.8–24.5)
' <i>Halomonas sambharensis</i> SBS 10'	GCA_003973725.1	63.6	89.1	38.4 (35.9–40.9)
<i>Halomonas elongate</i> DSM 2581 ^T	GCA_000196875.2	63.6	77.9	22.0 (19.8–24.5)

较基因组学分析结果, 根据普遍认可的原核生物物种评判标准, 菌株 JSM 104105 明显不能归属于盐单胞菌属任何已知物种, 应该代表该属一个物种级别的新的分类单元(基因种, *genospecies*)^[16,19]。

2.3.3 基因组系统发育学分析

在进行基因组系统发育分析时, 共筛选到 35 个盐单胞菌属已知物种典型菌株, 加上一个无效种‘*H. sambharensis* SBS 10’, 连同目标菌株 JSM 104105, 共计 37 个菌株的基因组进行基因组系统发育分析。结果表明, 采用 79% dDDH 阈值^[48]进行亚种水平聚类(*subspecies clustering*), 这 37 个菌株各自分别代表 1 个亚种簇(*subspecies cluster*); 但采用 70% dDDH 阈值^[16,44]进行物种水平聚类(*species clustering*), 这 37 个菌株只产生 36 个物种簇(*species cluster*), 其中 *Halomonas axialensis* Althfl^T

(GCA_007163885.2) 和 *Halomonas meridiana* ACAM 246^T (GCA_900129255.1) 聚为同一个物种簇, 而菌株 JSM 104105 (GCA_010993975.1) 等其他 35 菌株分别代表一个物种簇(图 3)。在采用 FASTME 2.1.6.1 构建的基因组最小进化距离系统发育树(minimum evolution tree)^[49]上, 无论是 SSU 树(图 3A)还是全基因组树(图 3B), 菌株 JSM 104105 都与 *H. gudaonensis* CGMCC 1.6133^T (GCA_900100195.1)、*H. azerbaijanica* TBZ202^T (GCA_004551485.1) 和‘*H. sambharensis* SBS 10’ (GCA_003973725.1) 聚在一起, 但菌株 JSM 104105 始终占据一个独立系统发育分支。因此, 基于基因组的系统发育学分析结果也明确显示, 菌株 JSM 104105 不属于任何一个已知的盐单胞菌属已知物种, 也独立于无效种‘*H. sambharensis*’, 明显代表了该属一个新的亚分支(*subclade*)。



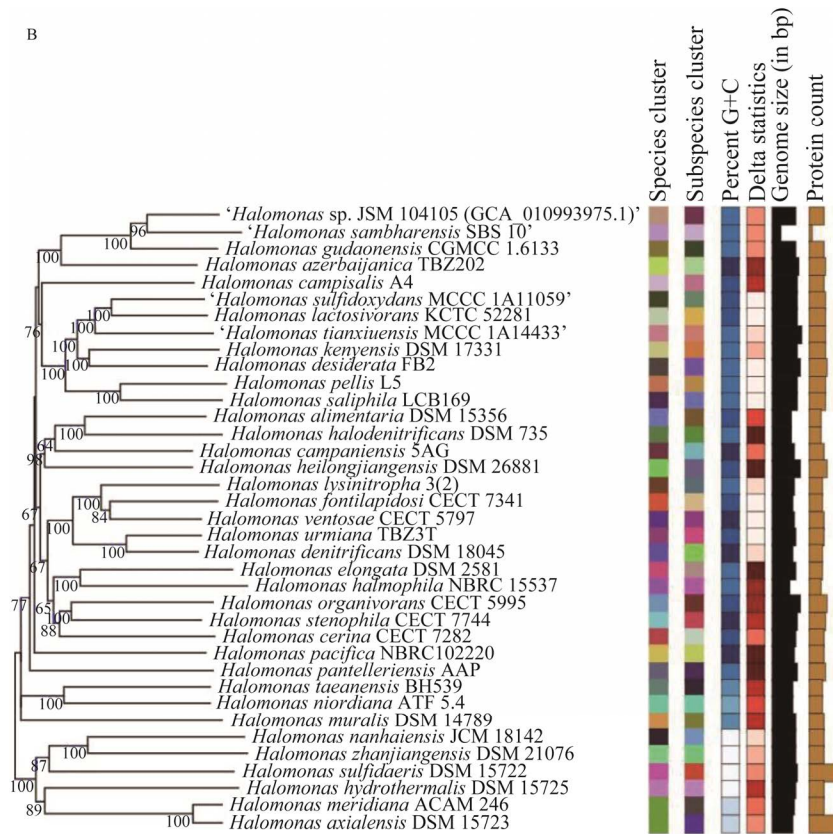


图3 菌株 JSM 104105 和盐单胞属相关菌株基于全基因组构建的 SSU (A)和基因组(B)系统发育树

Figure 3 Whole-genome-based phylogenetic trees inferred with FastME 2.1.6.1 from GBDP distances (formula d5) calculated from 16S rRNA gene sequences (A) and genome sequences (B) showing the phylogenetic relationships between strain JSM 104105 and related members of the genus *Halomonas*. The numbers above branches are GBDP pseudo-bootstrap support values >60% from 100 replications.

3 讨论与结论

本研究结果表明,菌株 JSM 104105 归属于盐单胞菌科盐单胞菌属,与该属的已知物种 *H. gudaonensis*、*H. azerbaijanica* 和 *H. lysinitropha* 的系统发育关系密切。虽然它们之间的 16S rRNA 基因相似性较高,如最高为 98.9% (*H. gudaonensis* CGMCC 1.6133^T),略高于 Kim 等根据大数据分析提出的细菌物种鉴别的 16S rRNA 基因序列相似性阈值(98.7%)^[58],但多位点序列分析(MLSA)和比较基因组学分析结果都揭示,菌株 JSM 104105 不能归属于任何已知物种,应该代表了盐单胞菌属一个新的基因种。

菌株 JSM 104105 系统发育地位的确定,增加了对盐单胞菌科生物多样性的认识,为对该菌株嗜盐机制、产铁载体机能等功能基因组学研究,以及多相分类和资源化利用等后续基础研究和应用前期研究打下了一定基础。

根据 EzBioCloud 在线 BLAST 分析结果,菌株 JSM 104105 与 Kushwaha 等^[30]新近发表 (effectively published)但没有生效(not validly published)的 '*H. sambharensis*' 的代表菌株 SBS10 的 16S rRNA 基因 PCR 产物测序所得序列(KT796562)相似性高达 99.6%,但二者无论在基于 16S rRNA 基因还是基于基因组的系统

发育树上,各自分别占据独立进化分支。比较基因组学分析发现,菌株 JSM 104105 与‘*H. sambharensis* SBS 10’之间的 ANI 值和 dDDH 值分别只有 89.1%和 36.4%,显著低于界定原核生物物种的阈值;采用 70% dDDH 阈值进行的物种水平聚类(species clustering),结果也表明菌株 JSM 104105 与‘*H. sambharensis* SBS 10’分别代表不同的物种(簇)。

盐单胞菌科等类群由于 16S rRNA 基因序列的高度保守性,使得 16S rRNA 基因序列分析法有时无法清晰地阐明有些关系密切的类群(物种)之间的系统发育关系。de la Haba 等基于多个管家基因序列的 MLSA 分析结果,推荐 16S rRNA 基因、*gyrB* 和 *rpoD* 基因列为盐单胞菌科新的系统发育分析分子标记^[13]。本研究结果也佐证了基于 16S rRNA、*gyrB* 和 *rpoD* 基因序列的 MLSA 分析在盐单胞菌科盐单胞菌属系统发育分析和分类学研究中的辅助作用。本研究结果表明,无论是基于 16S rRNA 基因、*gyrB* 和 *rpoD* 单基因序列分析,还是 3 基因串联序列分析,对于盐单胞菌属物种间的系统发育关系,都能提供较为一致的结果。当然,无论是单基因序列分析法,还是 3 基因串联序列分析法,对于 *M. tunisiensis* LIT2^T、*H. pacifica* NBRC 102220^T 和 *H. muralis* LMG 20969^T 等菌株的系统发育地位,都没能提供稳定的结果。

在本研究过程中,发现比较基因组学分析用于原核生物系统分类学研究存在一定局限性,至少在以下 3 个方面值得探讨。首先值得注意的是,部分基因组数据的质量欠佳,难以采信。例如,我们发现无效种‘*H. sambharensis* SBS 10’的基因组数据 GCA_003973725.1 用于系统学分析就存在 4 点不足:(1) 该基因组仅为 1.5 Mbp,只有多数细菌基因组数据的三分之一,信息量明显不足;(2) 从该基因组提取的 16S rRNA 基

因序列(ON556671),与 Kushwaha 等^[30]从同一菌株的 PCR 产物直接测得序列(KT796562)的相似性仅为 98.9%,相互差异严重失常;(3) 在基于基因组数据的 SSU 系统发育树上,‘*H. sambharensis* SBS 10’形成一个独特的长支,处于菌株 JSM 104105 与 *H. gudaonensis* CGMCC 1.16133^T 形成的亚簇之外(图 3A),与基于直接测序所得 16S rRNA 基因序列的进化树(图 1A)的聚类方式明显不同;(4) 我们尝试从‘*H. sambharensis* SBS 10’基因组中注释其 *gyrB* 和 *rpoD* 基因序列用于 MLSA 研究,但所得序列分别只有 749 bp 和 350 bp,无法采信。第 2 个值得注意的问题是,有些物种的多个基因组数据之间,可能由于菌株不同、或测序技术不同、或质量控制或组装的软件不同,部分基因组数据的质量或所包含的信息数量和质量存在较大偏差,造成基于不同基因组数据的系统学分析的部分结果之间有显著差异,难以用于根据普遍接受的标准来准确判定这些菌株间的系统发育关系。例如,*Halomonas aquamarina* 有 2 个基因组数据, GCF_020075495.1 和 GCF_900110265.1,分别来源于其典型菌株 MB598^T 和菌株 558,它们之间的 dDDH 和 ANI 估值分别只有 20.4% (18.2%–22.8%) (GGDC formula 2,下同)和 77.1%,均显著低于同一物种不同菌株间的阈值。再以 *Halomonas meridiana* 基因组数据为例。目前 *H. meridiana* 有 11 个基因组数据(GCA_000943375.1、GCA_006540125.1、GCA_009846525.1、GCA_011398715.1、GCA_011764625.1、GCA_018402055.1、GCA_020782315.1、GCA_021025935.1、GCA_900128865.1、GCA_900129255.1 和 GCA_900142865.1),其中 GCA_009846525.1 (来源于 *H. meridiana* SCSIO 43005)与其他基因组之间的差异巨大,它们之间的 ANI 值为

87.4%–88.1%，dDDH 估值为 30.3%–36.4%，都显著超越了种内菌株判定的相关阈值。以上例子提示，如果不是某些基因组数据质量欠佳，就是目前认同的基于基因组数据的 dDDH 和 ANI 阈值没有足够高的普适性，亦或是 dDDH 和 ANI 估值的算法有待改进。如果否定以上所有推测，那么这些结果则提示，菌株 558 应该不属于 *H. aquamarina*，菌株 SCSIO 43005 也属于一个不同于 *H. meridiana* 的物种(基因种)。比较基因组学分析用于原核生物系统学研究值得探讨的问题之三是，目前基因组系统发育学分析普遍采用的 Type Strain Genome Server 在线服务(TYGS; <https://tygs.dsmz.de>)，我们发现其各种技术中，至少参比菌株(基因组)的选择方面存在明显的局限性。一是每次选择参比菌株的数量上限为 28 个，无法同时进行大类群分析；二是其数据库只收集典型菌株的基因组，无法分析种内菌株间的基因组异质性；我们认为 TYGS 最大的不足是，用于分析的参比菌株主要依靠 TYGS 设定的一些算法自动选择，有时候一些研究者需要特别分析的个体不在被选之列，又无法手工加入，或者由于更新滞后，有些数据根本不在库中。这样造成研究者有时无法采用基因组数据对特定目标的系统发育地位进行全面精准分析。

当然，基于全基因组序列的 ANI 和 dDDH 估算、以及基因组系统发育学分析等比较基因组学分析新技术，用于评估菌株间遗传相关性(同源性)，相对于传统的 16S rRNA 基因序列分析、MLSA 和常规 DNA-DNA 杂交，多数情况下具有准确性更高和能更客观地反映菌株之间的系统进化关系等优势，在现代微生物系统分类学研究中开始展现其越来越重要的科学价值，正逐步被采用为原核生物物种鉴别的新的金标准^[16,19]。越来越多的以基因组系统发育学

(phylogenomics)为指导的基因组测序研究项目正在进行，例如“细菌和古菌基因组百科全书(genomic encyclopedia of bacteria and archaea, GEBA)计划”^[59]。相信随着测序技术的发展，基因组数据质量会不断提高，数据库会更加充实，分析方法也会在实践中得到检验和修正，基因组数据在不久的将来可能会成为原核生物分类学研究的典型材料(type material)^[60-61]，比较基因组学技术会更加广泛和更加有效地应用于系统分类学实践，原核生物系统学有望全面进入比较基因组学研究的新时期。

参考文献

- [1] STACKEBRANDT E, GOEBEL BM. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology[J]. *International Journal of Systematic and Evolutionary Microbiology*, 1994, 44(4): 846-849.
- [2] ACINAS SG, MARCELINO LA, KLEPAC-CERAJ V, POLZ MF. Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons[J]. *Journal of Bacteriology*, 2004, 186(9): 2629-2635.
- [3] CILIA V, LAFAY B, CHRISTEN R. Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level[J]. *Molecular Biology and Evolution*, 1996, 13(3): 451-461.
- [4] ANDERSON AS, WELLINGTON EM. The taxonomy of *Streptomyces* and related Genera[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2001, 51(3): 797-814.
- [5] SUN DL, JIANG X, WU QL, ZHOU NY. Intra-genomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity[J]. *Applied and Environmental Microbiology*, 2013, 79(19): 5962-5969.
- [6] STACKEBRANDT E, FREDERIKSEN W, GARRITY GM, GRIMONT PAD, KÄMPFER P, MAIDEN MCI, NESME X, ROSSELLÓ-MORA R, SWINGS J, TRÜPER HG, VAUTERIN L, WARD AC, WHITMAN WB. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2002, 52(3): 1043-1047.

- [7] DEVULDER G, de MONTCLOS MP, FLANDROIS JP. A multigene approach to phylogenetic analysis using the genus *Mycobacterium* as a model[J]. International Journal of Systematic and Evolutionary Microbiology, 2005, 55(Pt 1): 293-302.
- [8] GLAESER SP, KÄMPFER P. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy[J]. Systematic and Applied Microbiology, 2015, 38(4): 237-245.
- [9] ZEIGLER DR. Gene sequences useful for predicting relatedness of whole genomes in bacteria[J]. International Journal of Systematic and Evolutionary Microbiology, 2003, 53(Pt 6): 1893-1900.
- [10] ARAHAL DR, VENTOSA A. The family *Halomonadaceae*[AJ]//. Dworkin, S M, Falkow E. Rosenberg KH, Schleifer Stackebrandt E. The Prokaryotes[M]. 3rd ed. New York: Springer, 2006, 6: 811-835.
- [11] MARTENS M, DAWYNDT P, COOPMAN R, GILLIS M, de VOS P, WILLEMS A. Advantages of multilocus sequence analysis for taxonomic studies: a case study using 10 housekeeping genes in the genus *Ensifer* (including former *Sinorhizobium*)[J]. International Journal of Systematic and Evolutionary Microbiology, 2008, 58(Pt 1): 200-214.
- [12] TINDALL BJ, ROSSELLÓ-MÓRA R, BUSSE HJ, LUDWIG W, KÄMPFER P. Notes on the characterization of prokaryote strains for taxonomic purposes[J]. International Journal of Systematic and Evolutionary Microbiology, 2010, 60(1): 249-266.
- [13] de la HABA RR, MÁRQUEZ MC, PAPKE RT, VENTOSA A. Multilocus sequence analysis of the family Halomonadaceae[J]. International Journal of Systematic and Evolutionary Microbiology, 2012, 62(Pt 3): 520-538.
- [14] MAIDEN MCJ, van RENSBURG MJJ, BRAY JE, EARLE SG, FORD SA, JOLLEY KA, McCARTHY ND. MLST revisited: the gene-by-gene approach to bacterial genomics[J]. Nature Reviews Microbiology, 2013, 11(10): 728-736.
- [15] FANG YJ, WANG YL, LIU ZD, DAI H, CAI HY, LI ZP, DU ZJ, WANG X, JING HQ, WEI Q, KAN B, WANG DC. Multilocus sequence analysis, a rapid and accurate tool for taxonomic classification, evolutionary relationship determination, and population biology studies of the genus *Shewanella*[J]. Applied and Environmental Microbiology, 2019, 85(11): e03126-e03118.
- [16] WAYNE LG, MOORE WEC, STACKEBRANDT E, KANDLER O, COLWELL RR, KRICHEVSKY MI, TRUPER HG, MURRAY RGE, GRIMONT PAD, BRENNER DJ, STARR MP, MOORE LH. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics[J]. International Journal of Systematic and Evolutionary Microbiology, 1987, 37(4): 463-464.
- [17] ROSSELLÓ-MORA R. DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation[A]//Molecular Identification, Systematics, and Population Structure of Prokaryotes[M]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006: 23-50.
- [18] 刘志恒, 王剑, 张立新. 基因组时代的放线菌系统学及其研究进展[J]. 微生物学报, 2011, 51(2): 141-153. LIU ZH, WANG J, ZHANG LX. Advances in actinobacterial systematics in genomic era[J]. Acta Microbiologica Sinica, 2011, 51(2): 141-153 (in Chinese).
- [19] RICHTER M, ROSSELLÓ-MÓRA R. Shifting the genomic gold standard for the prokaryotic species definition[J]. Proceedings of the National Academy of Sciences of the United States of America, 2009, 106(45): 19126-19131.
- [20] 陈奇辉, 刘祝祥, 彭清忠, 黄奇, 贺建武, 张丽, 李文均, 陈义光. 小溪自然保护区非盐环境土壤中嗜盐和耐盐菌多样性[J]. 微生物学报, 2010, 50(11): 1452-1459. CHEN QH, LIU ZX, PENG QZ, HUANG N, HE JW, ZHANG L, LI WJ, CHEN YG. Diversity of halophilic and halotolerant bacteria isolated from non-saline soil collected from Xiaoxi National Natural Reserve, Hunan Province[J]. Acta Microbiologica Sinica, 2010, 50(11): 1452-1459 (in Chinese).
- [21] CHEN YG, ZHANG YQ, CHEN QH, KLENK HP, HE JW, TANG SK, CUI XL, LI WJ. *Bacillus xiaoxiensis* sp. nov., a slightly halophilic bacterium isolated from non-saline forest soil[J]. International Journal of Systematic and Evolutionary Microbiology, 2011, 61(9): 2095-2100.
- [22] WU QL, CHEN JH, DENG LY, LIU ZX, HE JW, LIU ZX, YANG LL, TANG SK, CHEN YG. *Sediminibacillus terrae* sp. nov., a moderate halophile isolated from non-saline farm soil[J]. International Journal of Systematic and Evolutionary Microbiology, 2020, 70(2): 1139-1144.
- [23] CUI XL, MAO PH, ZENG M, LI WJ, ZHANG LP, XU LH, JIANG CL. *Streptimonospora Salina* Gen. nov., sp. nov., a new member of the family Nocardioseae[J]. International Journal of Systematic and Evolutionary Microbiology, 2001, 51(2): 357-363.

- [24] BLATTNER FR, PLUNKETT G 3rd, BLOCH CA, PERNA NT, BURLAND V, RILEY M, COLLADO-VIDES J, GLASNER JD, RODE CK, MAYHEW GF, GREGOR J, DAVIS NW, KIRKPATRICK HA, GOEDEN MA, ROSE DJ, MAU B, SHAO Y. The complete genome sequence of *Escherichia coli* K-12[J]. *Science*, 1997, 277(5331): 1453-1462.
- [25] RILEY M, ABE T, ARNAUD MB, BERLYN MKB, BLATTNER FR, CHAUDHURI RR, GLASNER JD, HORIUCHI T, KESELER IM, KOSUGE T, MORI H, PERNA NT, PLUNKETT G, RUDD KE, SERRES MH, THOMAS GH, THOMSON NR, WISHART D, WANNER BL. *Escherichia coli* K-12: a cooperatively developed annotation snapshot—2005[J]. *Nucleic Acids Research*, 2006, 34(1): 1-9.
- [26] WINGETT SW, ANDREWS S. FastQ Screen: a tool for multi-genome mapping and quality control[J]. *F1000Research*, 2018, 7: 1338.
- [27] BOLGER AM, LOHSE M, USADEL B. Trimmomatic: a flexible trimmer for Illumina sequence data[J]. *Bioinformatics*, 2014, 30(15): 2114-2120.
- [28] LUO RB, LIU BH, XIE YL, LI ZY, HUANG WH, YUAN JY, HE GZ, CHEN YX, PAN Q, LIU YJ, TANG JB, WU GX, ZHANG H, SHI YJ, LIU Y, YU C, WANG B, LU Y, HAN CL, CHEUNG DW, et al. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler[J]. *GigaScience*, 2012, 1(1): 18.
- [29] YOON SH, HA SM, KWON S, LIM J, KIM Y, SEO H, CHUN J. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2017, 67(5): 1613-1617.
- [30] KUSHWAHA B, JADHAV I, JADHAV K. *Halomonas sambharensis* sp. nov, a moderately halophilic bacterium isolated from the saltern crystallizer ponds of the sambhar salt lake in India[J]. *Current Microbiology*, 2020, 77(6): 1125-1134.
- [31] TANIZAWA Y, FUJISAWA T, NAKAMURA Y. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication[J]. *Bioinformatics*, 2018, 34(6): 1037-1039.
- [32] LAGESEN K, HALLIN P, RØDLAND EA, STÆRFELDT HH, ROGNES T, USSERY DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes[J]. *Nucleic Acids Research*, 2007, 35(9): 3100-3108.
- [33] THOMPSON JD, GIBSON TJ, PLEWNIAK F, JEANMOUGIN F, HIGGINS DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools[J]. *Nucleic Acids Research*, 1997, 25(24): 4876-4882.
- [34] EDGAR RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput[J]. *Nucleic Acids Research*, 2004, 32(5): 1792-1797.
- [35] WILGENBUSCH JC, SWOFFORD D. Inferring evolutionary trees with PAUP[J]. *Current Protocols in Bioinformatics*, 2003, Chapter 6: Unit 6.4.
- [36] KIMURA M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences[J]. *Journal of Molecular Evolution*, 1980, 16(2): 111-120.
- [37] KUMAR S, STECHER G, TAMURA K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets[J]. *Molecular Biology and Evolution*, 2016, 33(7): 1870-1874.
- [38] SAITOU N, NEI M. The neighbor-joining method: a new method for reconstructing phylogenetic trees[J]. *Molecular Biology and Evolution*, 1987, 4(4): 406-425.
- [39] FELSENSTEIN J. Confidence limits on phylogenies: an approach using the bootstrap[J]. *Evolution; International Journal of Organic Evolution*, 1985, 39(4): 783-791.
- [40] CHAN PP, LIN BY, MAK AJ, LOWE TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes[J]. *Nucleic Acids Research*, 2021, 49(16): 9077-9096.
- [41] LEE I, KIM YO, PARK SC, CHUN J. OrthoANI: an improved algorithm and software for calculating average nucleotide identity[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2016, 66(2): 1100-1103.
- [42] YOON SH, HA SM, LIM J, KWON S, CHUN J. A large-scale evaluation of algorithms to calculate average nucleotide identity[J]. *Antonie Van Leeuwenhoek*, 2017, 110(10): 1281-1286.
- [43] MEIER-KOLTHOFF JP, CARBASSE JS, PEINADO-OLARTE RL, GÖKER M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes[J]. *Nucleic Acids Research*, 2022, 50(D1): D801-D807.
- [44] MEIER-KOLTHOFF JP, GÖKER M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy[J]. *Nature Communications*, 2019, 10: 2182.
- [45] ONDOV BD, TREANGEN TJ, MELSTED P, MALLONEE AB, BERGMAN NH, KOREN S,

- PHILLIPPY AM. Mash: fast genome and metagenome distance estimation using MinHash[J]. *Genome Biology*, 2016, 17(1): 132.
- [46] CAMACHO C, COULOURIS G, AVAGYAN V, MA N, PAPADOPOULOS J, BEALER K, MADDEN TL. BLAST+: architecture and applications[J]. *BMC Bioinformatics*, 2009, 10: 421.
- [47] MEIER-KOLTHOFF JP, AUCH AF, KLENK HP, GÖKER M. Genome sequence-based species delimitation with confidence intervals and improved distance functions[J]. *BMC Bioinformatics*, 2013, 14: 60.
- [48] MEIER-KOLTHOFF JP, HAHNKE RL, PETERSEN J, SCHEUNER C, MICHAEL V, FIEBIG A, ROHDE C, ROHDE M, FARTMANN B, GOODWIN LA, CHERTKOV O, REDDY T, PATI A, IVANOVA NN, MARKOWITZ V, KYRPIDES NC, WOYKE T, GÖKER M, KLENK HP. Complete genome sequence of DSM 30083^T, the type strain (U5/41^T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy[J]. *Standards in Genomic Sciences*, 2014, 9(1): 1-19.
- [49] LEFORT V, DESPER R, GASCUEL O. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program[J]. *Molecular Biology and Evolution*, 2015, 32(10): 2798-2800.
- [50] KAZEMI E, TARHRIZ V, AMOOZEGAR MA, HEJAZI MS. *Halomonas azerbaijanica* sp. nov., a halophilic bacterium isolated from Urmia Lake after the 2015 drought[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2021, 71(1): 004578.
- [51] RAMEZANI M, POURMOHYADINI M, NIKOU MM, MAKZUM S, SCHUMANN P, CLERMONT D, CRISCUOLO A, AMOOZEGAR MA, KÄMPFER P, SPRÖER C. *Halomonas lysinitropha* sp. nov., a novel halophilic bacterium isolated from a hypersaline wetland[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2020, 70(12): 6098-6105.
- [52] GONZALEZ-DOMENECH CM, MARTINEZ-CHECA F, QUESADA E, BEJAR V. *Halomonas cerina* sp. nov., a moderately halophilic, denitrifying, exopolysaccharide-producing bacterium[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2008, 58(4): 803-809.
- [53] LI XG, GAN LZ, HU MY, WANG ST, TIAN YQ, SHI B. *Halomonas pellis* sp. nov., a moderately halophilic bacterium isolated from wetsalted hides[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2020, 70(10): 5417-5424.
- [54] BOLTYANSKAYA YV, KEVBRIN VV, LYSENKO AM, KOLGANOVA TV, TOUROVA TP, OSIPOV GA, ZHILINA TN. *Halomonas mongoliensis* sp. nov. and *Halomonas kenyensis* sp. nov, new haloalkaliphilic denitrifiers capable of N₂O reduction, isolated from soda lakes[J]. *Microbiology*, 2007, 76(6): 739-747.
- [55] GAN LZ, LONG XF, ZHANG HM, HOU YY, TIAN JW, ZHANG YQ, TIAN YQ. *Halomonas saliphila* sp. nov., a moderately halophilic bacterium isolated from a saline soil[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2018, 68(4): 1153-1159.
- [56] GARCÍA MT, MELLADO E, OSTOS JC, VENTOSA A. *Halomonas organivorans* sp. nov., a moderate halophile able to degrade aromatic compounds[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2004, 54(5): 1723-1728.
- [57] WANG YN, CAI H, YU SL, WANG ZY, LIU J, WU XL. *Halomonas gudaonensis* sp. nov., isolated from a saline soil contaminated by crude oil[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2007, 57(5): 911-915.
- [58] KIM M, OH HS, PARK SC, CHUN J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2014, 64(Pt 2): 346-351.
- [59] WU DY, HUGENHOLTZ P, MAVROMATIS K, PUKALL R, DALIN E, IVANOVA NN, KUNIN V, GOODWIN L, WU M, TINDALL BJ, HOOPER SD, PATI A, LYKIDIS A, SPRING S, ANDERSON IJ, D'HAESELEER P, ZEMLA A, SINGER M, LAPIDUS A, NOLAN M, et al. A phylogeny-driven genomic encyclopaedia of bacteria and Archaea[J]. *Nature*, 2009, 462(7276): 1056-1060.
- [60] WHITMAN WB. Genome sequences as the type material for taxonomic descriptions of prokaryotes[J]. *Systematic and Applied Microbiology*, 2015, 38(4): 217-222.
- [61] HUGENHOLTZ P, CHUVOCHINA M, OREN A, PARKS DH, SOO RM. Prokaryotic taxonomy and nomenclature in the age of big sequence data[J]. *The ISME Journal*, 2021, 15(7): 1879-1892.