

研究报告

来自南极洲类诺卡氏菌(*Nocardioides* sp.) InS609-2 的全基因组序列分析张琪琪^{1,2}, 李会荣², 张雨欣², 杨帆^{*1}, 俞勇^{*2}

1 上海交通大学医学院附属仁济医院药学部, 上海 200127

2 中国极地研究中心, 上海 200136

张琪琪, 李会荣, 张雨欣, 杨帆, 俞勇. 来自南极洲类诺卡氏菌(*Nocardioides* sp.) InS609-2 的全基因组序列分析[J]. 微生物学通报, 2024, 51(1): 77-95.

ZHANG Qiqi, LI Huirong, ZHANG Yuxin, YANG Fan, YU Yong. Whole-genome sequence analysis of *Nocardioides* sp. InS609-2 isolated from Antarctica[J]. Microbiology China, 2024, 51(1): 77-95.

摘要: 【背景】类诺卡氏菌(*Nocardioides* sp.) InS609-2 是一株分离自南极洲罗斯海特拉诺瓦湾的恩克斯堡岛土壤中潜在的极地放线菌新种。尚无研究报道 *Nocardioides* sp. InS609-2 的全基因组序列, 缺少对其功能基因、代谢产物合成途径及比较基因组学等的研究。【目的】解析 *Nocardioides* sp. InS609-2 的基因组序列信息, 以深入挖掘次级代谢产物基因资源。【方法】利用 Illumina HiSeq 高通量测序平台对菌株 InS609-2 进行全基因组完成图测序, 使用相关软件对测序数据进行基因组组装、基因预测和功能注释、预测次级代谢产物合成基因簇和共线性分析等。【结果】基因组最后得到的总长度为 4 524 052 bp, G+C 含量为 69.42%, 预测到 4 656 个基因、56 个 tRNA 和 6 个 rRNA。根据 *Nocardioides* sp. InS609-2 的全基因组测序结果, 分别有 3 761、3 052、1 767、4 134 和 2 725 个基因在 COG、GO、KEGG、NR 和 Swiss-Prot 数据库中提取到注释信息。同时, 还预测得到 19 个次级代谢产物合成基因簇。基因组测序数据提交至 NCBI 获得 GenBank 登录号为 CP060034。*Nocardioides* sp. InS609-2 与 *N. dokdonensis* CP015079、*N. yefusunii* CP034929、*N. euryhalodurans* CP038267、*N. seonyuensis* CP038436、*N. daphniae* CP038462 和 *N. okcheonensis* CP087710 这 6 株基因组同源性比较高的类诺卡氏菌进行共线性分析和蛋白聚类分类分析, 得到的结果是 7 个基因组间既有保守性又各自有独特性。七个基因组共有 44 个蛋白聚类簇。最后进行 16S rRNA 基因系统发育树、泛基因组、core 基因组和基因组进化树分析, 进一步挖掘了 *Nocardioides* sp. InS609-2 的基因组信息。【结论】从基因组层面上预测了 *Nocardioides* sp. InS609-2 的次级代谢产物的生物合成基因簇, 为 InS609-2 的后续相关研究提供了参考信息, 具有重要意义。

资助项目: 国家重点研发计划(2022YFC2804300, 2022YFC2807501); 国家自然科学基金(U2106227)

This work was supported by the National Key Research and Development Program of China (2022YFC2804300, 2022YFC2807501) and the National Natural Science Foundation of China (U2106227).

*Corresponding authors. E-mail: YANG Fan, yang-fan@sjtu.edu.cn; YU Yong, yuyong@pric.org.cn

Received: 2023-04-20; Accepted: 2023-10-05; Published online: 2023-11-09

关键词: 类诺卡氏菌; 全基因组测序; 次级代谢产物; 比较基因组分析

Whole-genome sequence analysis of *Nocardioides* sp. InS609-2 isolated from Antarctica

ZHANG Qiqi^{1,2}, LI Huirong², ZHANG Yuxin², YANG Fan^{*1}, YU Yong^{*2}

1 Department of Pharmacy, Renji Hospital Affiliated to School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, China

2 Polar Research Institute of China, Shanghai 200136, China

Abstract: [Background] *Nocardioides* sp. InS609-2 is a potential new strain of polar actinomycete isolated from the soil sample collected from Inexpressible Island in Terra Nova Bay of the Ross Sea, Antarctica. Little is known about the full genome sequence, functional genes, metabolite synthesis pathways, and comparative genomics of *Nocardioides* sp. InS609-2. [Objective] To explore the genes associated with secondary metabolites from the genome sequence of *Nocardioides* sp. InS609-2. [Methods] The complete genome map of InS609-2 was sequenced on the Illumina HiSeq platform, followed by read assembly, gene prediction, gene annotation, prediction of gene clusters for secondary metabolite biosynthesis, and synteny analysis. [Results] The full length of the genome was 4 524 052 bp, with the G+C content of 69.42%. A total of 4 656 genes, 56 tRNAs, and 6 rRNAs were predicted. Among them, 3 761, 3 052, 1 767, 4 134, and 2 725 genes were annotated in COG, GO, KEGG, NR, and Swiss-Prot, respectively. Meanwhile, 19 gene clusters for secondary metabolite synthesis were predicted. Genome sequencing data was submitted to NCBI, with the GenBank accession number CP060034. The synteny analysis and protein cluster analysis were performed for *Nocardioides* sp. InS609-2 together with *N. dokdonensis* CP015079, *N. yefusunii* CP034929, *N. euryhalodurans* CP038267, *N. seonyuensis* CP038436, *N. daphniae* CP038462, and *N. okcheonensis* CP087710. The results showed both conserved and unique regions among the seven genomes. The seven genomes shared 44 protein clusters. Finally, the 16S rRNA gene-based phylogenetic analysis, pan genome analysis, core genome analysis, and genome-based phylogenetic analysis were carried out to further mine the genomic information of *Nocardioides* sp. InS609-2. [Conclusion] This study provides reference information for predicting the gene clusters involved in secondary metabolite biosynthesis of *Nocardioides* sp. InS609-2 at the genomic level, which is essential for the subsequent studies of InS609-2.

Keywords: *Nocardioides*; whole-genome sequencing; secondary metabolite; comparative genomic analysis

放线菌是革兰氏阳性细菌,是具有高 G+C 含量的细菌门类之一,它们广泛分布在水生和陆地生态系统中。三分之二的天然抗生素和一系列

抗癌、抗蠕虫、抗真菌和免疫抑制药物来源于放线菌^[1]。临床部分重要抗感染化合物,包括链霉素、四环素、氯霉素、新霉素、红霉素、万古霉

素、卡那霉素、利福霉素、庆大霉素、达托霉素、铂霉素和 PM181104, 以及抗癌化合物蒽环类、博莱霉素、mitosanes 和 enediynes^[2]均由放线菌提取获得。

类诺卡氏菌属(*Nocardioide*s)作为放线菌门的成员之一, 来源广泛。目前, 已经从罗马的地下墓穴^[3]、沙漠土壤^[4]、济州岛的瀑布^[5]、高岭石^[6]、淡水泥浆^[7]、西藏羚羊的粪便^[8]、对虾的肠胃道^[9]、玉米茎^[10]和南极湖泊^[11]等分离得到。本研究的类诺卡氏菌(*Nocardioide*s sp.) InS609-2 分离自南极洲罗斯海特拉诺瓦湾的恩克斯堡岛的土壤样品。

极地、沙漠、深海和火山等地方属于极端环境^[12]。极地这一特殊的生态环境使得大量极地微生物在其基因组成、酶学特点和生物合成的代谢调控等方面具有特异性。研究者从极地的微生物中分离到抗生素 frigocyclinone^[13]、生物碱 microbiaeratin^[14] 和多环内酰胺 cyclamenol^[15]等化合物。从极地来源的菌株中发掘出更多活性优异、结构新颖的次级代谢产物的可能性仍然很大。

测序技术的进步能够更加方便地对微生物基因组进行全基因组测序, 加速了对微生物产生的次级代谢产物的研究。最初, 基因组挖掘是通过检测和开发次级代谢产物产生的生物合成途径的中心酶来完成的。这些酶包括聚酮合成酶(polyketide synthases, PKS)、非核糖体肽合成酶(non-ribosomal peptide synthases, NRPs)^[16]。后来, antiSMASH 和 PRISM 等各种软件被开发出来, 可以识别部分或完整的生物合成基因簇^[15]。这些工具有效地将用户提交的基因组与现有的功能基因簇数据库进行比较, 并提供了次级代谢产物的结构预测。新的方法比如用基因组筛选次级代谢产物生物合成基因簇、进行克隆或者激活沉默状态的基因簇并且在异源宿主高表

达等不断发展, 用来进一步鉴定基因簇^[16]。因此, 基因组学指导下的次级代谢产物的发现是目前次级代谢产物发现的重要手段之一^[17]。

截至 2023 年 4 月, NCBI 的数据显示与类诺卡氏菌有关的生物项目(bioproject)有 138 个, 其中能够检索到的完成全基因组测序的类诺卡氏菌有 7 株^[11,18-23], 包括 1 株来自南极半对流高盐湖泊的 *Nocardioide*s *aquaticus* KCTC 9944^{T[11]}。随着全基因组测序成本的降低及有效的数据分析方法的改进, 全基因组测序将会越来越普遍地被采用, 从而获得分离源更加广泛的类诺卡氏菌株的基因组数据。本研究对极地菌 *Nocardioide*s sp. InS609-2 进行了全基因组测序及分析, 以期探究其次级代谢产物基因簇等基因信息。

1 材料与方法

1.1 材料

1.1.1 样品

类诺卡氏菌(*Nocardioide*s sp.) InS609-2 由中国极地研究中心(Polar Research Institute of China)生态中心实验室分离并保存。该菌株保藏在极地微生物种质资源库(Polar Microorganism Culture Collection, PMCC), 保藏编号为 PMCC 100130。

1.1.2 培养基、主要试剂和仪器

R2A 液体培养基^[24], BD 公司。Qiagen Genomic-Tip20/G 试剂盒, Qiagen 公司。摇床, 上海智诚分析仪器制造有限公司; 透射电子显微镜, 日立公司。

1.2 形态学鉴定

将菌株 *Nocardioide*s sp. InS609-2 接种于 R2A 固体培养基上, 20 °C 培养 3 d, 观察菌落形态特征; 制备菌株 *Nocardioide*s sp. InS609-2 的透射电子显微镜样品, 用染液(pH 7.0, 3%磷钨

酸)染色,观察该菌的细胞形态。

1.3 *Nocardioide*s sp. InS609-2 的培养

在 R2A 固体平板上划线,得到 *Nocardioide*s sp. InS609-2 的单菌落,接种至 4 mL R2A 液体培养基中,20 °C、120 r/min 培养 2 d,取 2 mL 菌悬液接种至 200 mL R2A 液体培养基中,20 °C、120 r/min 培养 2 d,4 °C、7 000 r/min 离心 10 min,收集菌体送至派森诺生物公司进行全基因组测序。

1.4 基因组测序和组装

使用 QiagenGenomic-Tip20/G 试剂盒高效纯化 *Nocardioide*s sp. InS609-2 的基因组 DNA,将其随机剪切成 600 bp 至 10 kb 长度的片段,分别构建序列大小为 20 kb 和 400 bp 的文库,使用 Illumina HiSeq 和 PacBio Sequel 平台测序。使用 HGAP 和 CANU 软件对经由 PacBio 获得的下机数据进行拼装,得到 contigs 序列。通过 PacBio Sequel 平台获取的 contigs 与通过 Illumina HiSeq 平台获取的 contigs 中的间隙用 GapFiller 弥合。再将二代的高质量数据通过 PILON 软件对三代 contigs 结果进行校正,拼接得到完整的序列。最后,对数据进行统计后采用 FastQC 对数据进行质量控制。

1.5 基因组组分分析

分别对 *Nocardioide*s sp. InS609-2 菌株的蛋白编码基因和非编码 RNA 进行结构分析。分别采用 GeneMarks 软件^[25]、tRNAscan-SE 软件^[26]和 Barrnap 软件预测全基因组序列、tRNA 区域和 tRNA 的二级结构以及 rRNA。另外,非编码区 RNA 通过与 Rfam 数据库比较进行预测。

1.6 基因通用功能注释

将基因的氨基酸序列分别与基因本体(gene ontology, GO)数据库、京都基因与基因组百科全书(Kyoto encyclopedia of genes and genomes, KEGG)、蛋白质直系同源簇(cluster of

orthologous groups of proteins, COG)数据库、欧洲蛋白质序列(Swiss-Prot protein sequence, Swiss-Prot)数据库和非冗余蛋白质(non-redundant protein, NR)数据库进行比对,完成基因的一系列功能注释。

1.7 次级代谢产物生物合成基因簇的预测

antiSMASH 是一款用于细菌和真菌基因组序列中次级代谢产物生物合成基因簇的快速鉴定、注释和分析的软件。相关数据可在 <http://antismash.secondarymetabolites.org> 上获得。运用 antiSMASH 软件预测 *Nocardioide*s sp. InS609-2 中次级代谢产物生物合成基因簇及其可能合成的化合物类型。

1.8 比较基因组学分析

根据参考菌 *N. dokdonensis* CP015079.1、*N. yefusunii* CP034929.1、*N. euryhalodurans* CP038267.1、*N. seonyuensis* CP038436.1、*N. daphnia* CP038462.1 和 *N. okcheonensis* CP087710.1 的基因组序列将 *Nocardioide*s sp. InS609-2 的基因组序列基于 Mummer 比对结果进行排序,然后根据 2 个基因组的总长度将序列按一定比例缩小后建立二维共线性图中的 X、Y 坐标轴,以及线性共线性图中的上、下坐标轴。做 BLAST 比对,根据位置信息将比对上的核酸序列区域按相同比例缩小后标到图上。

根据近缘菌株名称从 NCBI RefSeq 数据库获取其基因组信息和基因序列,利用 PGAP 得到测序菌株与近缘菌株的共有基因与特有基因,并进行深入分析。PGAP 首先会将所有菌株的蛋白序列放在一起进行 BLAST 比对,根据 BLAST 比对结果,确定出不同蛋白间相似程度。利用 BLASTClust 对 *Nocardioide*s sp. InS609-2 和参考基因组进行蛋白聚类分析。

1.9 基于 16S rRNA 基因序列的系统发育树构建

用 MAFFT 软件对包括 *Nocardioides* sp. InS609-2 在内的 20 株菌的 16S rRNA 基因序列进行比对, 采用 FastTree 软件构建系统发育树。

1.10 核心基因组分析

基于物种的蛋白序列利用 orthofinder 软件 (v2.5.4) 对包括 *Nocardioides* sp. InS609-2 在内的 31 株菌株进行基因家族分析, 其中使用 diamond 进行 all-VS-all blastp 分析, 寻找潜在的同源基因。

1.11 泛基因组分析

采用 PGAP 软件构建包括 *Nocardioides* sp. InS609-2 在内的 31 株菌株的 pan-genome 和 core-genome, 分析得到稀释曲线结果。

1.12 基于基因组序列的系统发育树构建

采用 FastTree 软件中的 ML 算法构建基于基因组序列的系统发育树。

1.13 基因组组装及评价结果

采用 SOAP denovo 2.04 对基因进行组装, 通过 *k*-mer 分析及 GC-depth 分析评价组装结果。

2 结果与分析

2.1 形态学鉴定结果

在 R2A 固体培养基上, *Nocardioides* sp. InS609-2 菌株的菌落形态呈圆形, 直径 1 mm 左右, 有隆起, 不透明, 边缘光滑完整, 见图 1; 透射电子显微镜下, 菌株呈较规则的短杆状, 表面不光滑, 菌体有些部位轻微突出, 有些部位轻微凹陷, 见图 2。

2.2 基因组组装结果和基因组组分分析

Nocardioides sp. InS609-2 基因组总长度 4 524 052 bp, 全基因组序列包含一条 4 524 052 bp 的染色体, G+C 含量为 69.42%, N50 和 N90 的长度分别为 8 874 和 5 199。菌株的基因组包含 4 656 个蛋白编码基因, tRNA 基因 56 个, rRNA

基因 6 个。*Nocardioides* sp. InS609-2 基因组与部分已经完成基因组测序的诺卡氏菌比较(表 1), 与 *N. dokdonensis* CP015079.1、*N. yefusunii* CP034929.1、*N. euryhalodurans* CP038267.1、*N. seonyuensis* CP038436.1、*N. daphnia* CP038462.1 和 *N. okcheonensis* CP087710.1 基因组大小相近, G+C 含量也和表 1 所列的 6 种类诺卡氏菌相近, 符合类诺卡氏菌基因组的基本特征。*Nocardioides* sp. InS609-2 全基因组序列已提交 NCBI, GenBank 登录号为 CP060034.1。

Nocardioides sp. InS609-2 基因组圈图见图 3, 有 3 761 个基因成功获得 COG 注释(图 4)。

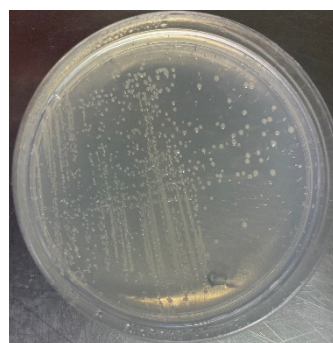


图 1 *Nocardioides* sp. InS609-2 的菌落形态
Figure 1 Colony morphology of *Nocardioides* sp. InS609-2.

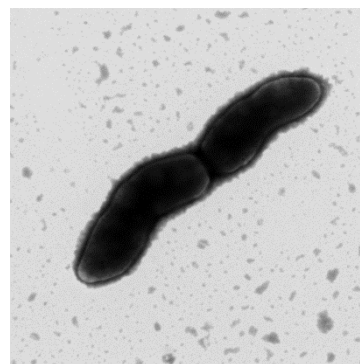


图 2 *Nocardioides* sp. InS609-2 的透射电子显微镜图
Figure 2 Transmission electron microscope image of *Nocardioides* sp. InS609-2.

表 1 *Nocardioides* sp. InS609-2 与其他部分类诺卡氏菌基因组比较
Table 1 Genome characteristics comparison of *Nocardioides* sp. InS609-2 with other *Nocardioides*

菌株 Strain	基因组总长度 Total length of genome (Mb)	Scaffolds 数量 Scaffolds number	G+C 含量 G+C content (%)	预测基因数 Predicted gene number	GenBank accession No.
<i>Nocardioides</i> sp. InS609-2	4.52	1	69.42	4 656	CP060034.1
<i>Nocardioides dokdonensis</i>	4.38	1	72.30	4 173	CP015079.1
<i>Nocardioides yefusunii</i>	3.29	1	69.10	3 060	CP034929.1
<i>Nocardioides euryhalodurans</i>	4.02	1	72.20	3 962	CP038267.1
<i>Nocardioides seonyuensis</i>	3.97	1	70.40	3 916	CP038436.1
<i>Nocardioides daphniae</i>	3.55	1	71.10	3 466	CP038462.1
<i>Nocardioides okcheonensis</i>	4.47	1	72.90	4 363	CP087710.1

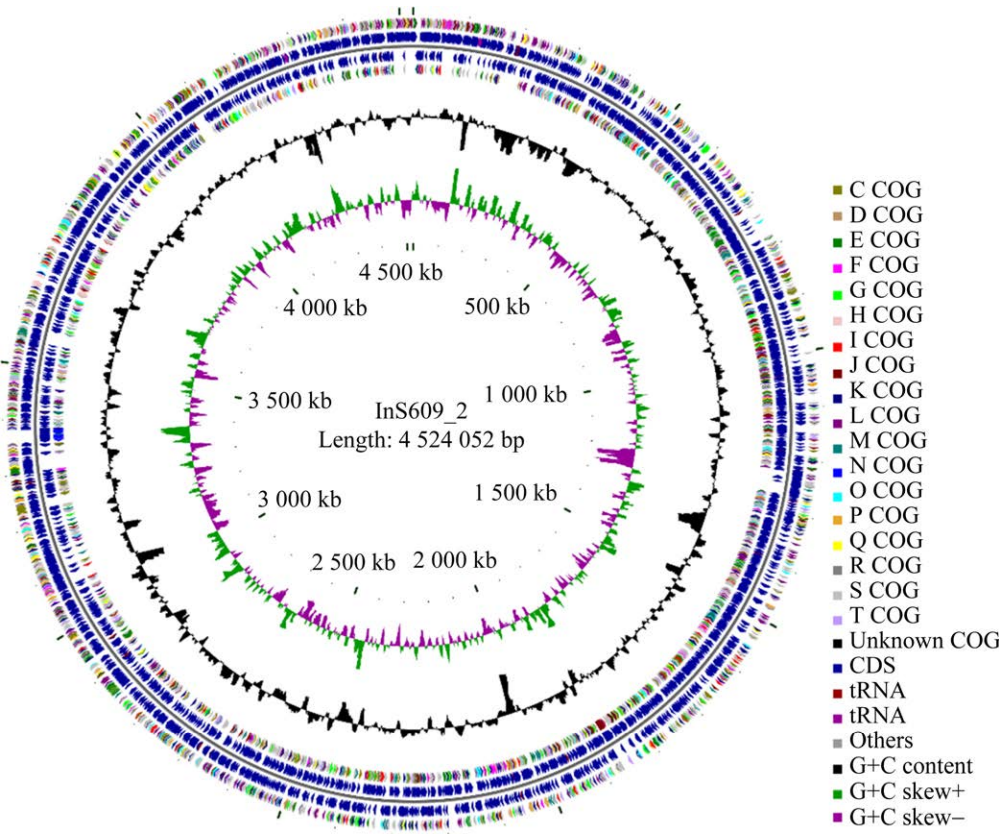


图 3 *Nocardioides* sp. InS609-2 基因组圈图 从内到外，第 1 圈代表刻度；第 2 圈代表 G+C skew；第 3 圈代表 G+C 含量；第 4 圈和第 7 圈代表每一个 CDS 所属的 COG；第 5 圈和第 6 圈代表 CDS、tRNA 及 rRNA 在基因组上的位置

Figure 3 Genomic map of *Nocardioides* sp. InS609-2. From inside to outside, the first circle represents the scale; The second circle represents G+C skew; The third circle represents G+C content; The fourth and seventh circles represent the COG to which each CDS belongs; The fifth and sixth circles represent the positions of CDS, tRNA, and rRNA on the genome.

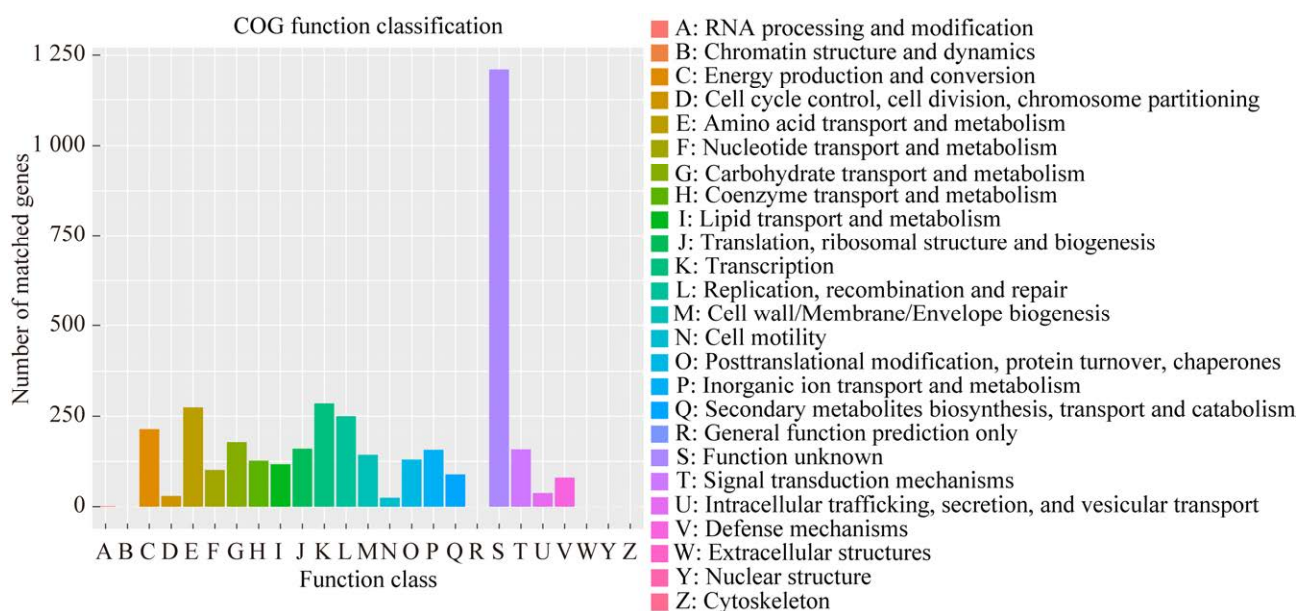


图4 *Nocardioides* sp. InS609-2 的 COG 功能分类图

Figure 4 COG functional classification map of *Nocardioides* sp. InS609-2.

COG 聚类分析发现其蛋白质功能主要集中在能量产生及传递(energy production and conversion) (214 个基因)、氨基酸转运与代谢(amino acid transport and metabolism) (274 个基因)、碳水化合物转运与代谢(carbohydrate transport and metabolism) (178 个基因)、翻译以及核糖体结构和生物合成(translation, ribosomal structure and biogenesis) (160 个基因)、转录(transcription) (285 个基因)、复制以及重组与修复(replication, recombination and repair) (250 个基因)、无机离子转运与代谢(inorganic ion transport and metabolism) (157 个基因)和信号转导机制(signal transduction mechanisms) (158 个基因)这些方面, 上述各功能注释的基因占 COG 功能注释总基因数的百分比依次为 5.69%、7.29%、4.73%、4.25%、7.58%、6.65%、4.17%和 4.20%。其中, S 代表的未知功能(function unknown)的基因数最多, 表明 *Nocardioides* sp. InS609-2 基因组序列具有的新颖性。

2.3 基因通用功能注释

2.3.1 GO 功能分类

在 *Nocardioides* sp. InS609-2 基因组中, 3 052 个基因能够在 GO 数据库提取得到注释信息(图 5), 按各基因功能可将其分为生物学过程(biological process)、细胞组分(cellular component)和分子功能(molecular function)这三大亚类, 其中, 生物学过程亚类中涉及基因最多的是生物学过程(biological process)、细胞的氮化合物代谢过程 (cellular nitrogen compound metabolic process)、生物合成过程(biosynthetic process)和小分子代谢过程 (small molecule metabolic process); 细胞组分亚类中涉及基因最多的是细胞(cell)、细胞组成(cellular component)、胞内(intracellular)和细胞质(cytoplasm); 分子功能亚类中涉及基因最多的是分子功能 (molecular function)和离子结合(ion bonding)。GO 功能注释可以对所获得的预测基因进行深层次的标准化功能分类, 便于研究者们理解基因背后所代表的生物学意义。

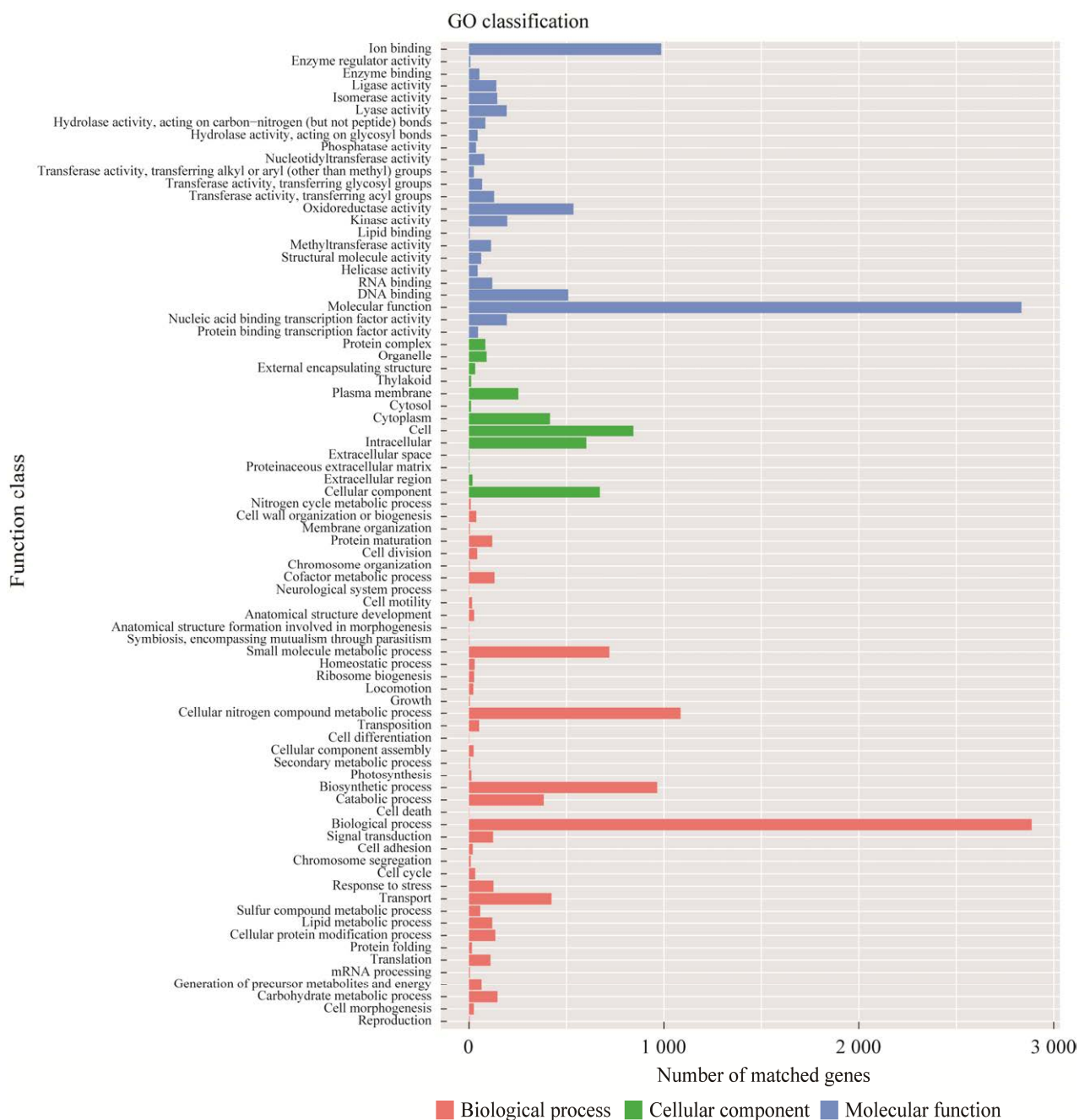


图5 *Nocardioidees* sp. InS609-2 GO 功能分类图

Figure 5 GO functional classification map of *Nocardioidees* sp. InS609-2.

2.3.2 KEGG 代谢途径分析

KEGG 是一个系统分析基因功能的数据库, 将基因组信息与更高阶的功能信息联系起来, 其中包含细胞过程的图形表示, 如代谢、膜转运、信号转导和细胞周期。

KEGG 代谢分析表明, 在 *Nocardioidees* sp. InS609-2 基因组中, 有 1 767 个基因在该数据库中得到有意义的注释(图 6)。大多数基因注释集中在碳水化合物代谢(carbohydrate metabolism)、氨基酸代谢(amino acid metabolism)、蛋白质家

族: 信号和细胞过程(protein families: signaling and cellular processes)、蛋白质家族: 代谢(protein families: metabolism)、蛋白质家族: 遗传信息处理 (protein families: genetic information processing)中, 分别有相关基因 360、321、439、231 和 486 个。此外, 在 metabolism 中与次级代谢产物生物合成(biosynthesis of other secondary metabolites)相关的基因有 44 个。

2.3.3 Swiss-Prot 数据库和 NR 数据库注释

Swiss-Prot 是一个蛋白质序列数据库, 它致力于提供高水平的注释[如蛋白质功能的描述、结构域结构、翻译后修饰 (post-translational modification, PTM)和变体等]、最低程度的冗余及与其他数据库的高度集成。经过比对,

*Nocardioide*s sp. InS609-2 基因组有 2 725 个基因的蛋白序列功能在 Swiss-Prot 数据库得到注释。

NR 数据库是 NCBI 中一个非冗余的蛋白质数据库。它包含从 GenBank 核酸序列翻译而来的非冗余序列, 并且还收录了其他蛋白数据库的非冗余序列。*Nocardioide*s sp. InS609-2 基因组中有 4 134 个基因在 NR 数据库上得到注释, 其中与 *Nocardioide*s sp. Soil805 具有最多的同源基因数目, 总共 404 个, 占全部注释结果的 9.77%。从中可以看出, 即使是基因组同源性与 *Nocardioide*s sp. InS609-2 相似度最高的 *Nocardioide*s sp. Soil805, 相似的比例也比较低, 说明 *Nocardioide*s sp. InS609-2 菌株的新颖性, 有待后续进一步的种属鉴定。

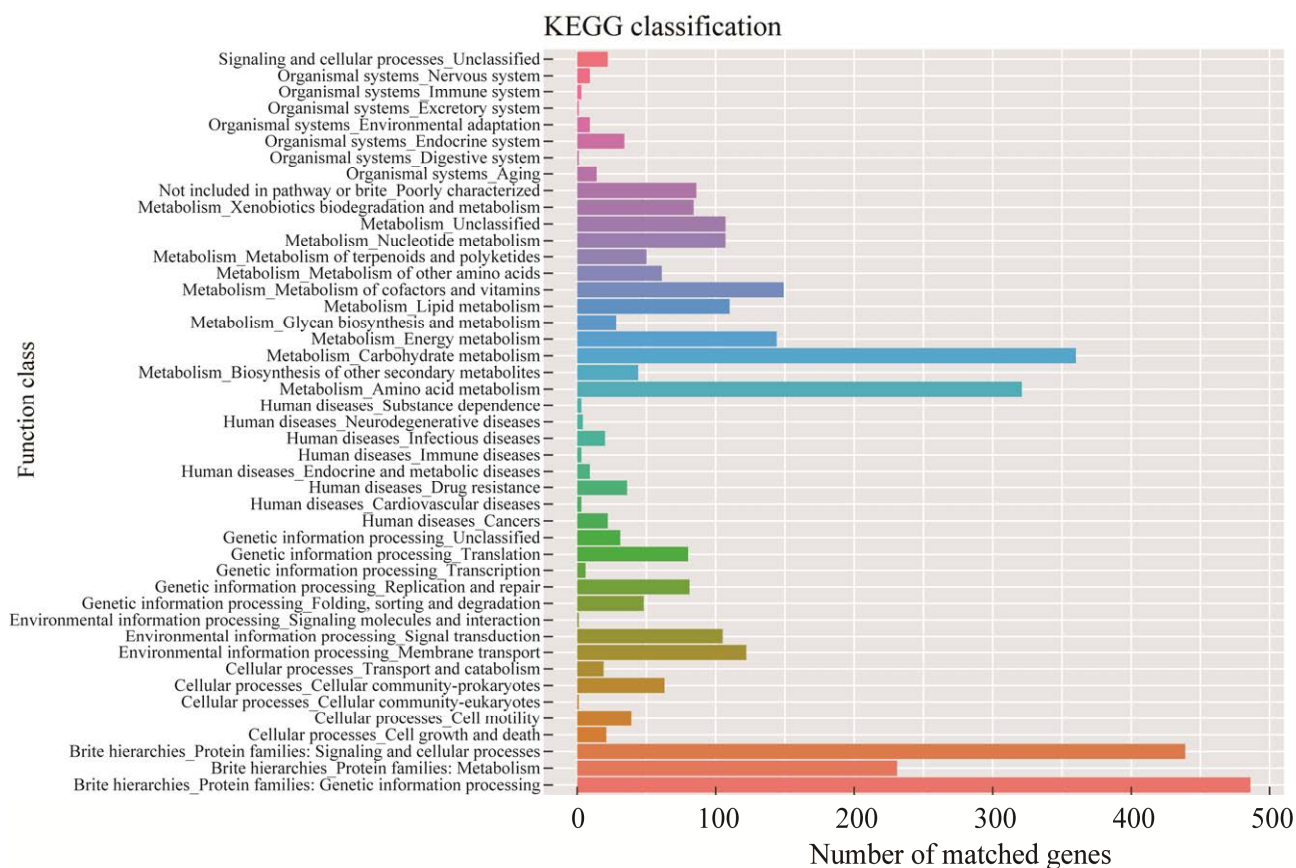


图 6 *Nocardioide*s sp. InS609-2 KEGG 通路分布图

Figure 6 KEGG pathway distribution of *Nocardioide*s sp. InS609-2 genome sequence.

2.4 次级代谢产物合成基因簇预测分析结果

antiSMASH 软件预测 *Nocardioide* sp. InS609-2 基因组得到 19 个基因簇(表 2)。根据 eggNOG 注释, 88 个基因参与次级代谢产物的

产生。antiSMASH 7 beta 软件预测得到了 19 个次级代谢物生物合成基因簇。基因簇大小为 9–78 kb。其中, 编码糖类的基因簇有 14 个, 占预测到的基因簇总和的 73.7%, 剩下的基因簇

表 2 *Nocardioide* sp. InS609-2 次级代谢产物合成基因簇

Table 2 Gene clusters of secondary metabolite of *Nocardioide* sp. InS609-2

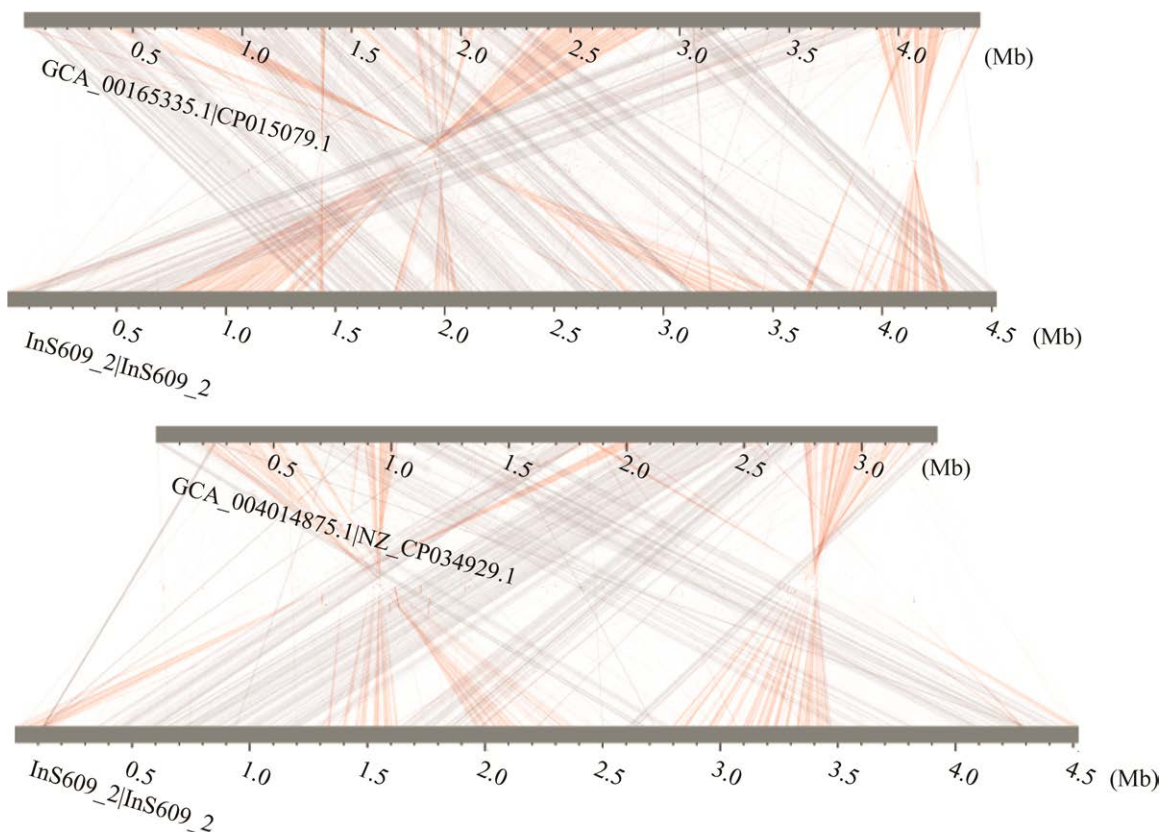
基因簇 Cluster	类型 Type	长度 Length (nt)	最相似的已知基因簇 Most similar known cluster	相似度 Similarity (%)	已知功能 Known function
1	核糖体合成和翻译后修饰多肽 Ribosomally synthesized and post-translationally modified peptides, RiPP-like	9 436	Unknown		
2	糖类 Saccharide	31 978	SF2575		
3	糖类 Saccharide	24 047	Madurastatin D1/Madurastatin D2/(-)-madurastatin C1		
4	糖类 Saccharide	24 806	Unknown		
5	糖类 Saccharide	34 531	Unknown		
6	糖类 Saccharide	29 923	Cervinomycin B3/Cervinomycin C1/Cervinomycin C2/Cervinomycin C3/Cervinomycin C4		
7	糖类 Saccharide	22 727	Unknown		
8	糖类 Saccharide	37 488	Unknown		
9	糖类 Saccharide	49 990	O and K-antigen		
10	糖类 Saccharide	38 028	Branched-chain fatty acids	100	Anti-inflammatory, anticancer
11	III 型聚酮化合物合酶 Type III polyketide synthases, T3PKS	41 071	Alkylresorcinol	100	Antibacterial
12	糖类 Saccharide	28 127	TVA-YJ-2		
13	非核糖体肽合成酶 Nonribosomal peptide synthetase, NRPS-like	42 508	Indigoidine	20	Natural blue pigment
14	糖类 Saccharide	27 495	Accramycin A	8	Antibacterial
15	糖类 Saccharide	78 411	Leucomycin		
16	非核糖体肽合成酶, 核糖体合成和翻译后修饰多肽 Nonribosomal peptide synthetase, NRPS-like, ribosomally synthesized and post-translationally modified peptides, RiPP-like	50 602	Pacidamycin 1/Pacidamycin 2/Pacidamycin 3/Pacidamycin 4/Pacidamycin 5/Pacidamycin 6/Pacidamycin 7/Pacidamycin D	63	Antibiotic
17	糖类 Saccharide	35 039	Deoxyhangtaimycin		
18	脂肪酸, 非核糖体肽合成酶 Fatty acid, nonribosomal peptide synthetase, NRPS	51 443	Asukamycin		
19	糖类 Saccharide	22 075	Unknown		

类型为核糖体合成和翻译后修饰多肽(ribosomally synthesized and post-translationally modified peptides, RiPP-like)、III 型聚酮化合物合酶(T3KPS)和非核糖体肽合酶(nonribosomal peptide synthetase, NRPS-like)和脂肪酸(fatty acid)类型。*Nocardioide*s sp. InS609-2 编码的基因簇中, 基因簇 10 与 *Streptomyces filamentosus* 的支链脂肪酸合成基因簇的相似性为 100%, 基因簇 11 与 *Streptomyces griseus* 的烷基间苯二酚合成基因簇的相似性为 100%, 基因簇 16 与 *Streptomyces coeruleorubidus* 的 pacidamycin 合成基因簇的相似性为 63%。除了这 3 个次级代谢产物基因簇与已知基因簇的相似度高之外, 大多数区域与已知基因簇的相似度总体较低(低于 35%), 但具有较大的预测核心簇(20–100 kb), 表明这些区域具有合成新颖化合物的潜力。具体是否能够合成这些预测的次级代谢产物还与培养

条件等因素有关。由于 *Nocardioide*s sp. InS609-2 菌株新颖且其次级代谢产物相关文献报道很少, 具有研究的潜力, 有待进一步深入挖掘。

2.5 比较基因组学分析

由 NCBI 上数据分析发现, *Nocardioide*s sp. InS609-2 基因组与 *Nocardioide*s *dokdonensis*、*N. yefusunii*、*N. euryhalodurans*、*N. seonyuensis*、*N. daphniae* 和 *N. okcheonensis* 这 6 株菌的基因组数据相似性比较高(与 *Nocardioide*s sp. InS609-2 进行 BLAST 比对, identity 分别为 94.35%、94.08%、94.93%、94.35%、94.99%和 94.39%)。*Nocardioide*s sp. InS609-2 基因组与上述 6 株类诺卡氏菌的全基因组进行共线性分析, 结果见图 7。从基因组间的连线可以看出它们之间的匹配关系, 其连线并不是非常紧密, 代表共线性程度一般。推测该菌可能是一个潜在的新种, 有待进一步的试验来验证。



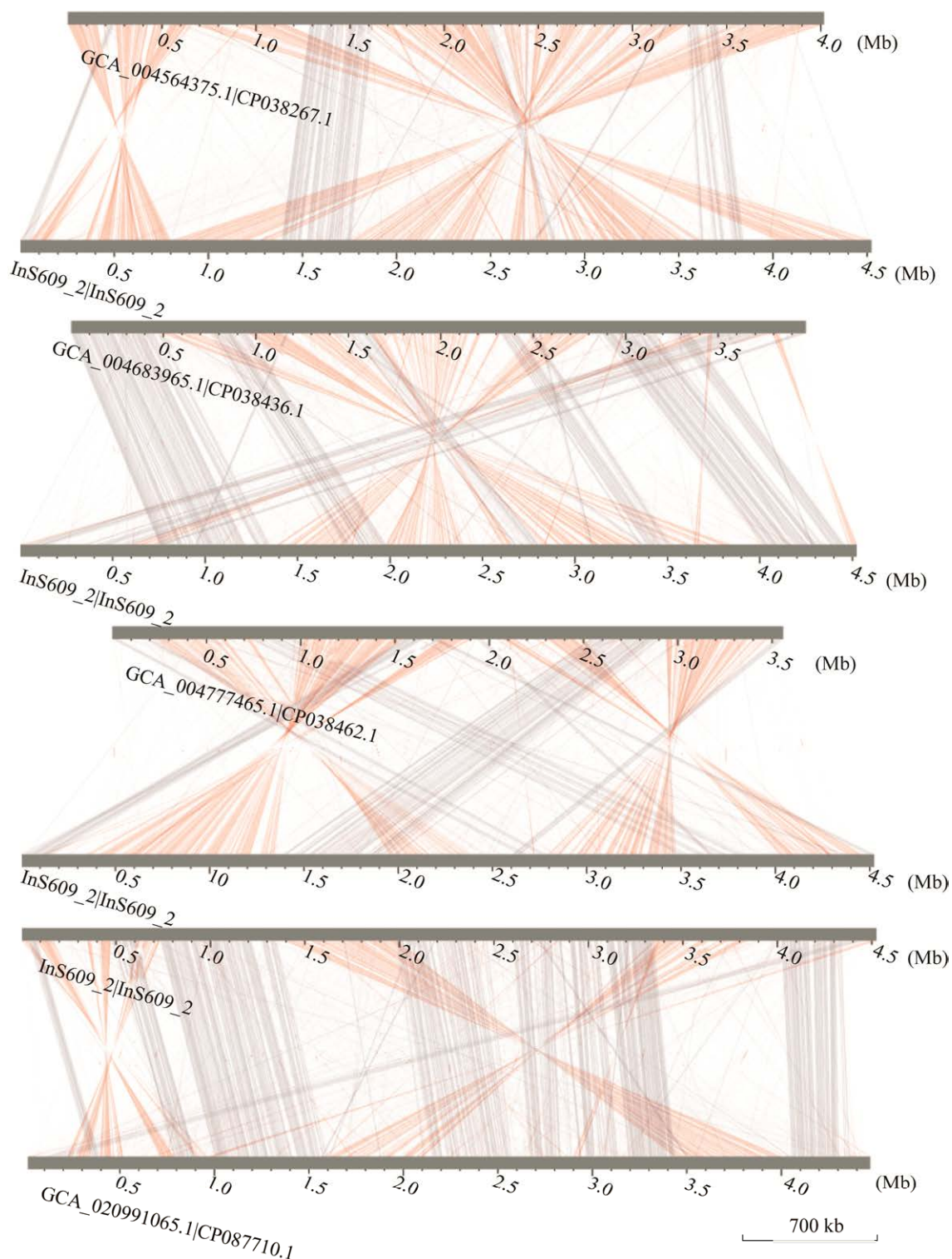


图 7 *Nocardioideae* sp. InS609-2 与 *Nocardioideae dokdonensis*、*Nocardioideae yefusunii*、*Nocardioideae euryhalodurans*、*Nocardioideae seonyuensis*、*Nocardioideae daphniae* 和 *Nocardioideae okcheonensis* 的共线性分析
 Figure 7 Synteny analysis of *Nocardioideae* sp. InS609-2, *Nocardioideae dokdonensis*, *Nocardioideae yefusunii*, *Nocardioideae euryhalodurans*, *Nocardioideae seonyuensis*, *Nocardioideae daphniae* and *Nocardioideae okcheonensis*. CP015079.1 refers to *N. dokdonensis*, CP034929.1 refers to *N. yefusunii*, CP038267.1 refers to *N. euryhalodurans*, CP038436.1 refers to *N. seonyuensis*, CP038462.1 refers to *N. daphniae*, CP087710.1 refers to *N. okcheonensis*.

利用 BLASTClust 对 *Nocardioide*s sp. InS609-2 和参比菌株 *Nocardioide*s *dokdonensis* CP015079.1、*N. yefusunii* CP034929.1、*N. euryhalodurans* CP038267.1、*N. seonyuensis* CP038436.1、*N. daphniae* CP038462.1 和 *N. okcheonensis* CP087710.1 的全基因组蛋白进行聚类分析,如图 8 所示,44 个蛋白聚类簇是 7 个菌株基因组共有的,4 353 个蛋白聚类簇仅 *Nocardioide*s sp. InS609-2 含有,其他 6 个基因组中不存在。

2.6 基于 16S rRNA 基因序列构建的系统发育树

在 NCBI 数据库中找到 20 株类诺卡氏菌 16S rRNA 基因序列用于和 *Nocardioide*s sp. InS609-2 一起构建系统发育树,结果见图 9。

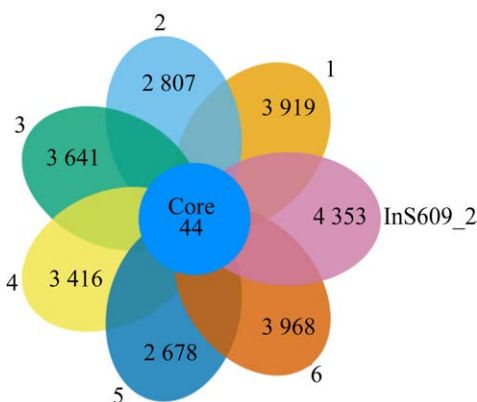


图 8 BLASTClust 蛋白聚类分类图

Figure 8 BLASTClust classification of proteins into clusters. 1: *Nocardioide*s *dokdonensis* CP015079.1; 2: *Nocardioide*s *yefusunii* CP034929.1; 3: *Nocardioide*s *euryhalodurans* CP038267.1; 4: *Nocardioide*s *seonyuensis* CP038436.1; 5: *Nocardioide*s *daphniae* CP038462.1; 6: *Nocardioide*s *okcheonensis* CP087710.1.

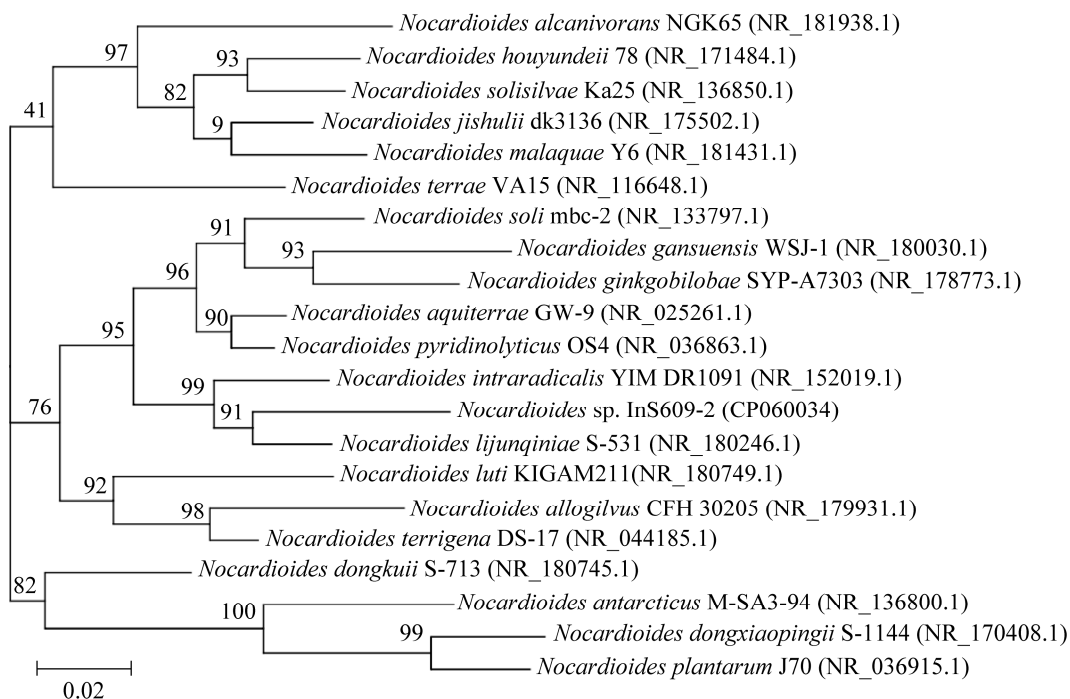


图 9 *Nocardioide*s sp. InS609-2 与相关菌株基于 16S rRNA 基因序列构建的系统发育树 括号中序号是相关菌株的登录号;分支点上的数字是 bootstrap 支持率;标尺刻度 0.02 是序列差异的分支长度

Figure 9 The phylogenetic tree based on 16S rRNA gene sequences of *Nocardioide*s sp. InS609-2 and related strains. The numbers in parentheses are accession numbers of related strains; The numbers in each branch points are percentages supported by bootstrap; Bar=0.02 is nucleotide divergence.

2.7 核心基因组分析

在 NCBI 数据库中找到 20 株类诺卡氏菌(表 3 中的前 20 株菌)基因组用于和 *Nocardioide*s sp. InS609-2 一起进行核心基因组分析, 分析结果见图 10。

2.8 泛基因组分析

在 NCBI 数据库中, 有 30 株类诺卡氏菌菌株(表 3)具有基因组相关数据。对这 30 株菌以及 *Nocardioide*s sp. InS609-2 总共 31 株菌的基因组进行泛基因组分析。从结果可以看出(图 11), 随着基因组个数的增加, 菌的泛基因总数为上升趋势,

说明菌的遗传物质具有开放性, 同时也说明菌具有相对较高的遗传多样性。

2.9 基于基因组序列构建的系统发育树

在 NCBI 数据库中找到 20 株类诺卡氏菌(表 3 中的前 20 株菌)基因组用于和 *Nocardioide*s sp. InS609-2 一起构建基因组系统发育树, 结果见图 12。

2.10 基因组组装及评价结果

由图 13 可知, 散点图呈现泊松分布的形状, 所以说明该测序数据不存在 GC 的偏向性。由图 14 可知, 在 *k*-mer 深度达到 78 的时候, 覆盖率最高。

表 3 核心基因组分析和泛基因组分析中使用的菌株

Table 3 Strains information used in core genome analysis and pan-genome analysis

编号 No.	菌株名称 Strain	登录号 Accession No.
1	<i>Nocardioide</i> s sp. BP30	CP123620.1
2	<i>Nocardioide</i> s sp. B-3	CP087968.1
3	<i>Nocardioide</i> s sp. CF8	CM001852
4	<i>Nocardioide</i> s sp. HDW12B	CP049867.1
5	<i>Nocardioide</i> s sp. JQ2195	CP050902.1
6	<i>Nocardioide</i> s sp. JS614	CP000508.1
7	<i>Nocardioide</i> s sp. Kera G14	CP085839.1
8	<i>Nocardioide</i> s sp. LMS CY	CP076104.1
9	<i>Nocardioide</i> s sp. L-11A	CP124680.1
10	<i>Nocardioide</i> s sp. QY071	CP124681.1
11	<i>Nocardioide</i> s sp. Root122	GCA_001425025.1
12	<i>Nocardioide</i> s sp. Root190	GCA_001428565.1
13	<i>Nocardioide</i> s sp. S5	CP022296.1
14	<i>Nocardioide</i> s sp. S-1144	CP040695.2
15	<i>Nocardioide</i> s sp. Soil777	GCA_001428425.1
16	<i>Nocardioide</i> s sp. T2.26MG-1	GCA_950101795.1
17	<i>Nocardioide</i> s sp. TF02-7	CP092535.1
18	<i>Nocardioide</i> s sp. ex-173	GCA_021117365.1
19	<i>Nocardioide</i> s sp. dk884	CP045649.1
20	<i>Nocardioide</i> s sp. Zg-1228	GCA_017086465.1
21	<i>Nocardioide</i> s sp. AX2bis	GCA_902506435.1
22	<i>Nocardioide</i> s sp. IC4_145	GCA_011250565.1
23	<i>Nocardioide</i> s sp. J9	GCA_007829375.1
24	<i>Nocardioide</i> s sp. KC13	GCA_011045035.1
25	<i>Nocardioide</i> s sp. OK12	BNET00000000.1
26	<i>Nocardioide</i> s sp. Root151	GCA_001427925.1
27	<i>Nocardioide</i> s sp. Root240	GCA_001428825.1
28	<i>Nocardioide</i> s sp. Root79	GCA_001428765.1
29	<i>Nocardioide</i> s sp. STR2	GCA_026712185.1
30	<i>Nocardioide</i> s sp. Soil796	GCA_001429525.1

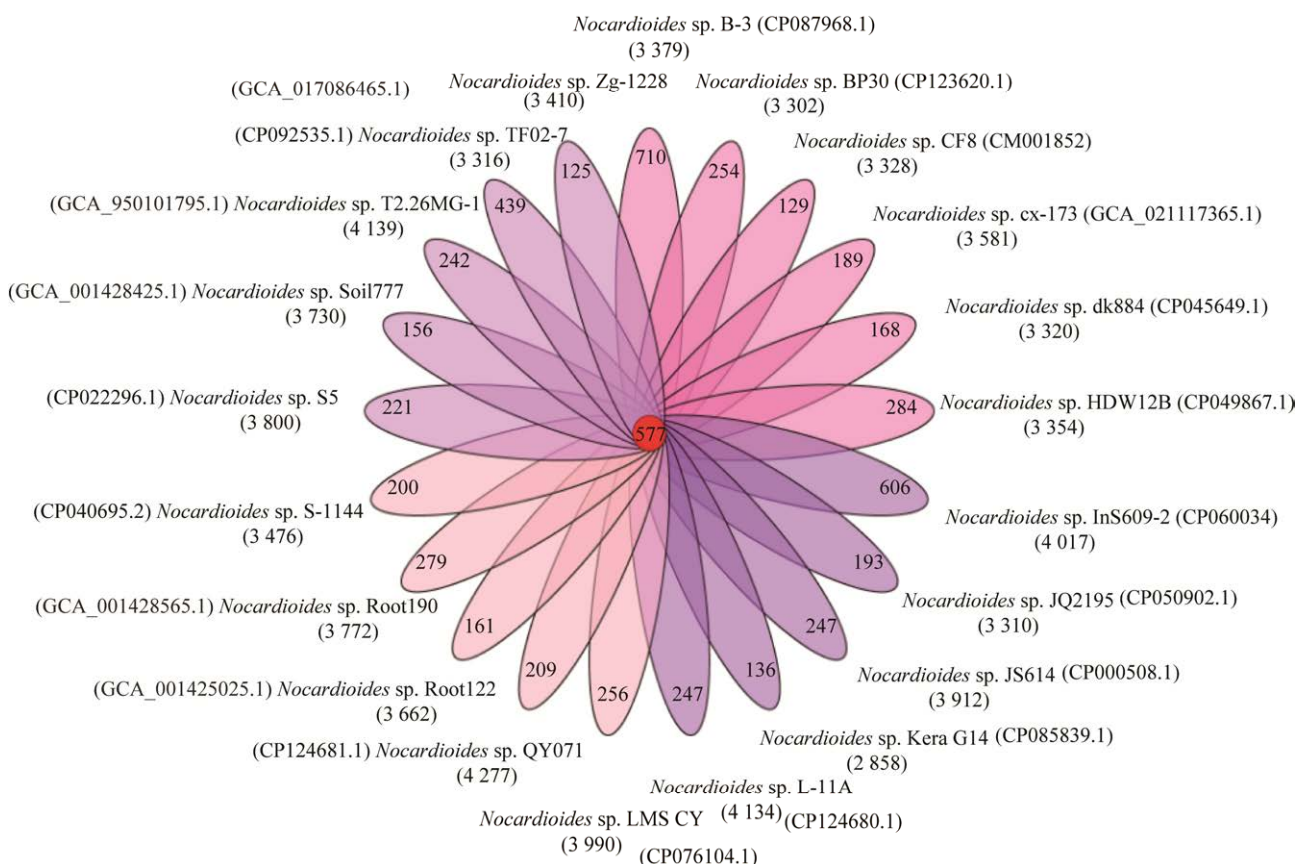


图 10 *Nocardioides* sp. InS609-2 的核心基因组分析

Figure 10 Core genome analysis of *Nocardioides* sp. InS609-2.

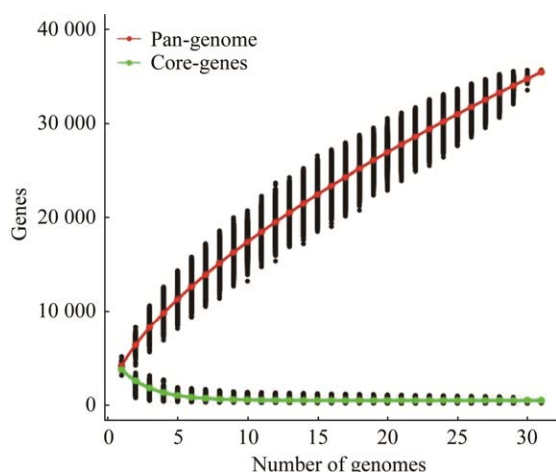


图 11 *Nocardioides* sp. InS609-2 的泛基因组分析

Figure 11 Pan genome analysis of *Nocardioides* sp. InS609-2.

3 讨论与结论

类诺卡氏菌 *Nocardioides* sp. InS609-2 全基因组测序分析结果表明, COG 聚类分析中, S 代表的未知功能(function unknown)的基因数最多; NR 数据库中, 与 *Nocardioides* sp. InS609-2 相似度最高的 *Nocardioides* sp. Soil805, 相似的比例也比较低 (average nucleotide identity, ANI=76.59); 比较基因组学分析中, 与 6 株参比菌株 *Nocardioides dokdonensis* CP015079.1、*N. yefusunii* CP034929.1、*N. euryhalodurans* CP038267.1、*N. seonyuensis* CP038436.1、*N. daphniae* CP038462.1 和 *N. okcheonensis* CP087710.1 的共线性程度低。这些结果提示 *Nocardioides* sp. InS609-2 可能是一个新菌, 有待进一步的种属鉴定。

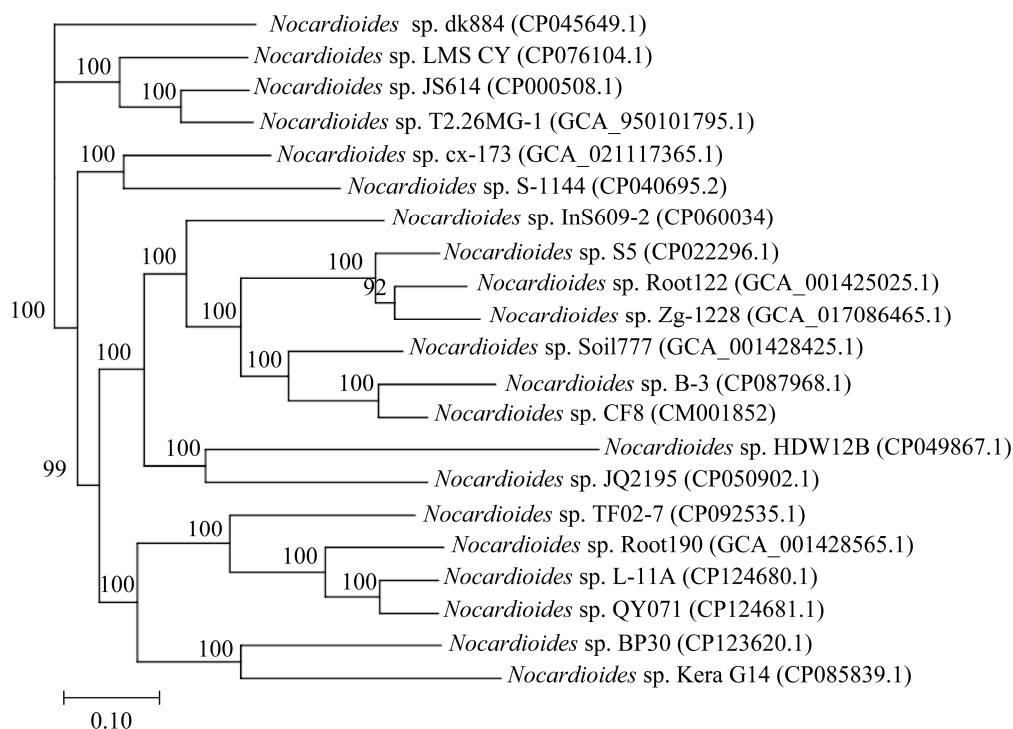


图 12 *Nocardioideae* sp. InS609-2 与相关菌株基于基因组序列构建的系统发育进化树 括号中序号是相关菌株的登录号；分支点上的数字是 bootstrap 支持率；标尺刻度 0.01 是序列差异的分支长度

Figure 12 Phylogenetic tree of genome based on nucleic acid sequences of *Nocardioideae* sp. InS609-2 and related strains. The numbers in parentheses are accession numbers of related strains; The numbers in each branch points are percentages supported by bootstrap; Bar=0.01 is nucleotide divergence.

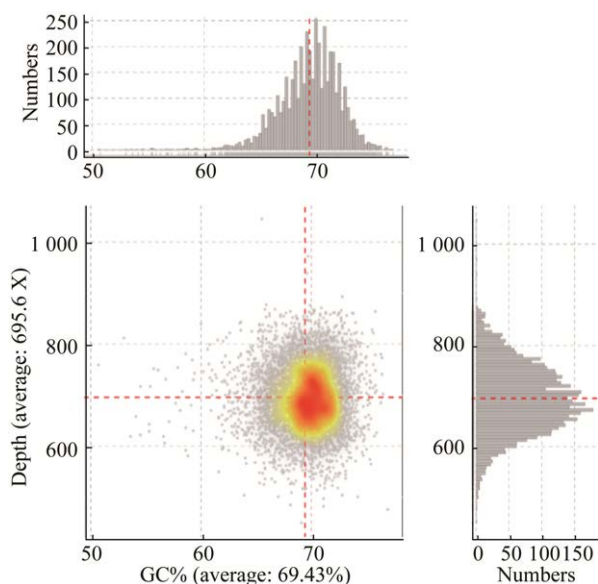


图 13 GC-depth 分布图

Figure 13 GC-depth distribution diagram.

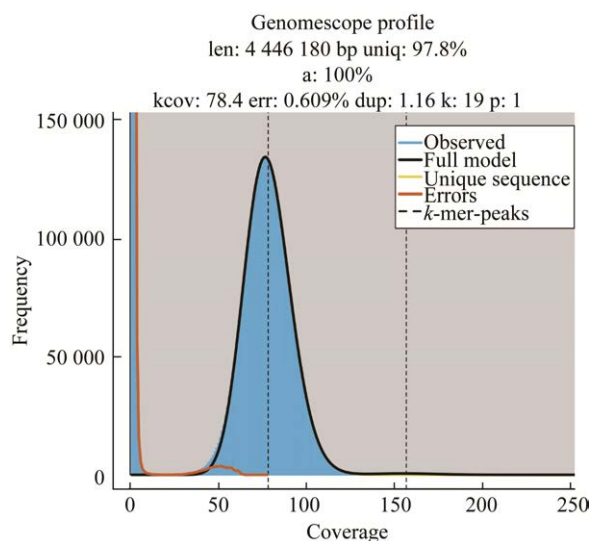


图 14 *k*-mer 分布曲线图

Figure 14 *k*-mer distribution diagram.

截至目前,关于 *Nocardioide*s 全基因组测序的研究报道比较少。有研究报道了一株从与柳树 (*Salix alba*) 有关的土壤中分离出来的类诺卡氏菌 *Nocardioide*s sp. WS12^[18], 对其进行了全基因组测序, 揭示了与橡胶降解相关的基因, 一个完整的异戊二烯单加氧酶基因簇, 以及相关的异戊二烯降解途径基因的存在, 表明 *Nocardioide*s sp. WS12 可能具有降解聚顺-1,4-异戊二烯的能力。然而异戊二烯是一种气候活性气体, 其在全球的广泛生产主要源于陆地植物的排放。该结果有助于研究异戊二烯的合成和释放对大气污染的重要作用。另有研究报道 2005 年在独岛海滩的泥沙中分离到 *Nocardioide*s *dodonensis* FR1436^{T[21]}, 对其进行序列分析检测到数十个参与类固醇代谢, 特别是类固醇降解的基因。2015 年完成对 *Nocardioide*s *simplex* VKM Ac-2033D^[20] 的全基因组测序, 显示该菌株有 5 421 个蛋白质编码序列, 4 633 个编码序列(coding sequence, CDS)和 816 个假基因, 对这些序列的初步分析揭示了几个参与胆固醇代谢的基因簇(主要通过侧链降解、类固醇核心降解和转运这 3 种方式)。类固醇是生物体内重要的生物分子, 还被用于治疗从炎症到癌症的各种疾病, 其改性研究在改善人类健康方面具有较高的潜力。*Nocardioide*s sp. strain JS614^[22] 的全基因组序列是 *Nocardioide*s 的第一个基因组序列, 同时也是氯乙烯(vinyl chloride, VC)/乙烯氧化细菌的第一个基因组序列。JS614 生长在以乙烯和氯乙烯为唯一碳源和能源的环境中, 其全基因组测序分析结果表明了烯烃氧化的遗传基础, 支持了正在进行的乙烯和氯乙烯生长的生理和生物化学研究, 并提供了生物标记物。

本研究对 *Nocardioide*s sp. InS609-2 的全基因组序列进行分析, 最后得到的总长度为 4 524 052 bp, G+C 含量平均为 69.42%, 预测得

到 4 656 个基因、56 个 tRNA 和 6 个 rRNA。通过 antiSMASH 预测得到 19 个次级代谢产物合成基因簇, 在通过该基因组预测获得的 19 个基因簇中, 有 5 个显示出与先前已知的基因簇有比较高的相似度(图 1), 其中 2 个最有可能生物合成已知的化合物。一种是糖(表 2 中的基因簇 10), 负责生物合成化合物支链脂肪酸^[27]。在人类脂肪细胞中, 支链脂肪酸能够改变脂质生成和炎症相关基因的表达, 这为肥胖相关代谢疾病的研究提供了一个方向。另一种是 T3KPS (表 2 中的基因簇 11), 负责生物合成化合物烷基间苯二酚^[6]。它已在体内和体外被证明可以改变与免疫系统和代谢控制相关的各种生理和病理过程。除了这两个次级代谢物基因簇与已知基因簇的相似度为 100%外, 大多数区域与已知天然产物的相似度总体较低(低于 35%), 但具有较大的预测核心簇(20–100 kb), 这表明这些途径可能编码新的天然产物。以上结果证明 *Nocardioide*s sp. InS609-2 有合成具有新结构和活性的次级代谢产物的能力。

*Nocardioide*s sp. InS609-2 系统发育树的构建可以将 *Nocardioide*s 内不同个体的遗传距离与进化关系可视化, 了解它们之间的进化关系。通过对包括 *Nocardioide*s sp. InS609-2 在内的类诺卡氏菌进行核心基因组分析, 得到了 *Nocardioide*s sp. InS609-2 的 577 个核心基因, 这些核心基因很可能涉及 *Nocardioide*s 菌体的基础代谢功能, 其目的是维持最基本的生命特征。通过对包括 *Nocardioide*s sp. InS609-2 在内的类诺卡氏菌进行泛基因组分析, 可以看到其基因组呈开放型, 导致这一结果的原因可能是 *Nocardioide*s 的分布广泛, 可以与外界的各种遗传物质发生交换。通过比较基因组学揭示了类诺卡氏菌不同菌株的遗传信息, 发现了不同菌株之间的差异, 为不同菌株的区分鉴定和应用奠定了遗传学的基础。*Nocardioide*s 之间基因组

的差异性可能与复杂的生活环境有关,是适应复杂的生长环境的一种体现。本研究也从这个方面证明了类诺卡氏菌的基因组多样性以及研究潜力。

REFERENCES

- [1] JAGANNATHAN SV, MANEMANN EM, ROWE SE, CALLENDER MC, SOTO W. Marine actinomycetes, new sources of biotechnological products[J]. Mar Drugs, 2021, 19(7): 365.
- [2] JOSE PA, MAHARSHI A, JHA B. Actinobacteria in natural products research: progress and prospects[J]. Microbiological Research, 2021, 246: 126708.
- [3] ALIAS-VILLEGAS C, JURADO V, LAIZ L, MILLER AZ, SAIJ-JIMENEZ C. *Nocardioide albertanoniae* sp. nov., isolated from Roman catacombs[J]. International Journal of Systematic and Evolutionary Microbiology, 2013, 63(Pt_4): 1280-1284.
- [4] TUO L, DONG YP, HABDEN X, LIU JM, GUO L, LIU XF, CHEN L, JIANG ZK, LIU SW, ZHANG YB, ZHANG YQ, SUN CH. *Nocardioide deserti* sp. nov., an actinobacterium isolated from desert soil[J]. International Journal of Systematic and Evolutionary Microbiology, 2015, 65(Pt_5): 1604-1610.
- [5] KIM I, CHHETRI G, KIM J, KANG M, SO Y, LEE B, JANG W, SEO T. *Nocardioide donggukensis* sp. nov. and *Hyunsoonleella aquatilis* sp. nov., isolated from Jeongbang Waterfall on Jeju Island[J]. International Journal of Systematic and Evolutionary Microbiology, 2021, 71(12).
- [6] YOON DJ, CHO ES, HWANG CY, NAM YD, PARK SL, LIM SI, SEO MJ. *Nocardioide luti* sp. nov., belonging to the family *Nocardioideaceae* isolated from kaolinite, exhibiting the biosynthesis potential of alkylresorcinol[J]. Antonie Van Leeuwenhoek, 2021, 114(7): 983-995.
- [7] YAN ZF, LIN P, LI CT, KOOK M, YI TH. *Nocardioide pelophilus* sp. nov., isolated from freshwater mud[J]. International Journal of Systematic and Evolutionary Microbiology, 2018, 68(6): 1942-1948.
- [8] DONG K, LU S, YANG J, PU J, LAI XH, JIN D, LI JQ, ZHANG G, WANG XX, LIANG JR, TIAN Z, ZHANG SH, HUANG YY, GE YJ, ZHOU J, REN ZH, WU XM, HUANG Y, WANG SP, XU JG. *Nocardioide jishulii* sp. nov., isolated from faeces of Tibetan gazelle (*Procapra picticaudata*)[J]. International Journal of Systematic and Evolutionary Microbiology, 2020, 70(6): 3665-3672.
- [9] KIM DY, JEONG IC, LEE SY, JEONG YS, HAN JE, TAK EJ, LEE JY, KIM PS, HYUN DW, BAE JW. *Nocardioide palaemonis* sp. nov. and *Tessaracoccus palaemonis* sp. nov., isolated from the gastrointestinal tract of lake prawn[J]. International Journal of Systematic and Evolutionary Microbiology, 2022, 72(12).
- [10] GLAESER SP, MCINROY JA, BUSSE HJ, KÄMPFER P. *Nocardioide zeae* sp. nov., isolated from the stem of *Zea mays*[J]. International Journal of Systematic and Evolutionary Microbiology, 2014, 64(Pt_7): 2491-2496.
- [11] HWANG K, CHOE H, KIM KM. Complete genome of *Nocardioide aquaticus* KCTC 9944T isolated from meromictic and hypersaline ekho lake, Antarctica[J]. Marine Genomics, 2022, 62: 100889.
- [12] 江婷, 蒲洪, 段燕文, 颜晓晖, 黄勇. 深海、沙漠、火山、极地来源链霉菌新天然产物(2009—2020)[J]. 有机化学, 2021, 41(5): 1804-1820.
- JIANG T, PU H, DUAN YW, YAN XH, HUANG Y. New natural products of *Streptomyces* sourced from deep-sea, desert, volcanic, and polar regions from 2009 to 2020[J]. Chinese Journal of Organic Chemistry, 2021, 41(5): 1804-1820 (in Chinese).
- [13] BRUNTNER C, BINDER T, PATHOM-AREE W, GOODFELLOW M, BULL AT, POTTERAT O, PUDER C, HÖRER S, SCHMID A, BOLEK W, WAGNER K, MIHM G, FIEDLER HP. Frigocyclinone, a novel angucyclinone antibiotic produced by a *Streptomyces griseus* strain from Antarctica[J]. The Journal of Antibiotics, 2005, 58(5): 346-349.
- [14] IVANOVA V, KOLAROVA M, ALEKSIEVA K, GRÄFE U, DAHSE HM, LAATSCH H. Microbiaeratin, a new natural indole alkaloid from a *Microbispora aerate* strain, isolated from Livingston Island, Antarctica[J]. Preparative Biochemistry and Biotechnology, 2007, 37(2): 161-168.
- [15] SHEN JJ, FAN YQ, ZHU GL, CHEN H, ZHU WM, FU P. Polycyclic macrolactams generated via intramolecular diels-alder reactions from an Antarctic *Streptomyces* species[J]. Organic Letters, 2019, 21(12): 4816-4820.
- [16] SINGH TA, PASSARI AK, JAJOO A, BHASIN S, GUPTA VK, HASHEM A, ALQARAWI AA, ABD_ALLAH EF. Tapping into actinobacterial genomes for natural product discovery[J]. Frontiers in Microbiology, 2021, 12: 655620.
- [17] 胥加龙, 王慧. 基因组学在放线菌次级代谢产物发现中的应用[J]. 药物生物技术, 2018, 25(4): 350-353.
- XU JL, WANG H. Application of genomics in secondary metabolite discovery in actinomycete[J]. Pharmaceutical Biotechnology, 2018, 25(4): 350-353 (in Chinese).

- [18] GIBSON L, LARKE-MEJÍA NL, MURRELL JC. Complete genome of isoprene degrading *Nocardioides* sp. WS12[J]. Microorganisms, 2020, 8(6): 889.
- [19] DUBOURG G, SANKAR SA, RATHORED J, LAGIER JC, ROBERT C, COUDERC C, PAPAIZIAN L, RAOULT D, FOURNIER PE. Noncontiguous finished genome sequence and description of *Nocardioides massiliensis* sp. nov. GD13(T)[J]. New Microbes and New Infections, 2016, 10: 47-57.
- [20] SHTRATNIKOVA VY, SCHELKUNOV MI, PEKOV YA, FOKINA VV, LOGACHEVA MD, SOKOLOV SL, BRAGIN EY, ASHAPKIN VV, DONOVA MV. Complete genome sequence of steroid-transforming *Nocardioides simplex* VKM Ac-2033D[J]. Genome Announcements, 2015, 3(1): e01406-14.
- [21] KWAK MJ, KWON SK, KIM JF. Complete genome sequence of the sand-sediment actinobacterium *Nocardioides dokdonensis* FR1436^T[J]. Standards in Genomic Sciences, 2017, 12(1): 1-7.
- [22] COLEMAN NV, WILSON NL, BARRY K, BRETTIN TS, BRUCE DC, COPELAND A, DALIN E, DETTER JC, del RIO TG, GOODWIN LA, HAMMON NM, HAN SS, HAUSER LJ, ISRANI S, KIM E, KYRPIDES N, LAND ML, LAPIDUS A, LARIMER FW, LUCAS S, et al. Genome sequence of the ethene- and vinyl chloride-oxidizing actinomycete *Nocardioides* sp. strain JS614[J]. Journal of Bacteriology, 2011, 193(13): 3399-3400.
- [23] PUKALL R, LAPIDUS A, GLAVINA del RIO T, COPELAND A, TICE H, CHENG JF, LUCAS S, CHEN F, NOLAN M, LABUTTI K, PATI A, IVANOVA N, MAVROMATIS K, MIKHAILOVA N, PITLUCK S, BRUCE D, GOODWIN L, LAND M, HAUSER L, CHANG YJ, et al. Complete genome sequence of *Kribbella flavida* type strain (IFO 14399^T)[J]. Standards in Genomic Sciences, 2010, 2(2): 186-193.
- [24] 马乐, 张博, 刘元峥, 苗静, 高志国, 薄晓菲. R2A 培养基的应用及有效期的验证[J]. 生物化工, 2021, 7(1): 73-76, 80.
- MA L, ZHANG B, LIU YZ, MIAO J, GAO ZG, BO XF. Application and validity verification of R2A medium[J]. Biological Chemical Engineering, 2021, 7(1): 73-76, 80 (in Chinese).
- [25] BESEMER J, LOMSADZE A, BORODOVSKY M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions[J]. Nucleic Acids Research, 2001, 29(12): 2607-2618.
- [26] SCHATTNER P, BROOKS AN, LOWE TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs[J]. Nucleic Acids Research, 2005, 33(suppl_2): W686-W689.
- [27] CZUMAJ A, ŚLEDZIŃSKI T, MIKA A. Branched-chain fatty acids alter the expression of genes responsible for lipid synthesis and inflammation in human adipose cells[J]. Nutrients, 2022, 14(11): 2310.