

研究报告

植物内生链霉菌 *Streptomyces* sp. SAT1 的基因组测序和比较基因组分析王莎¹ 窦桂铭² 马玉超^{*1}

1 北京林业大学 北京 100089

2 中国林业科学研究院 北京 100089

摘要:【背景】一直以来,链霉菌都是活性物质的主要生产者,近年来随着抗生素滥用引起的环境和微生物抗药性问题越发严重,挖掘高效生物防治因子和新型抗生素成为了解决以上问题的重要手段。【目的】通过获得植物内生链霉菌 SAT1 全基因组序列和次级代谢基因簇信息,利用比较基因组学和泛基因组学分析 SAT1 菌株的特殊性以及与其他链霉菌的共性,为阐明 SAT1 抑菌和内生机制提供理论基础,为揭示链霉菌的生态功能提供可靠数据。【方法】通过三代测序平台 PacBio Sequel 完成 SAT1 基因组测序,利用生物信息学技术进行注释和功能基因分类;分别利用 RAxML 和 PGAP 软件进行系统发育树的构建和泛基因组分析;次级代谢基因簇的预测和分析通过 antiSMASH 网站完成。【结果】获得 SAT1 菌株的全基因组完成图,该菌线性染色体长度约 7.47 Mb,包含有 4 个质粒,GC 含量近 73%,共预测到 7 550 个蛋白编码基因,含有 37 个次级代谢基因簇,分属 29 个类型,其中默诺霉素基因簇与加纳链霉菌具有较高相似性。42 株代表链霉菌中,单个菌株次级代谢基因簇数量为 20–55 个,主要类型为 PKS 类、Terpene 类和 Nrps 类,而且含有大量杂合基因簇,各个菌株中特有基因数目较为庞大。【结论】链霉菌 SAT1 菌株在基因组特点以及次级代谢基因簇的数量和类型上与其余 41 株链霉菌具有一定的共性,其中潮霉素 B 基因簇和默诺霉素基因簇合成的相关物质可能与 SAT1 抑菌活性密切相关。42 株链霉菌中次级代谢基因簇数量的多少与基因组大小成正相关,同时大量杂合基因簇以及庞大的特有基因数目的存在说明链霉菌在长期进化过程中存在了很高程度的水平基因转移现象,可能具有重要的生态功能。

关键词: 链霉菌, 全基因组测序, 次级代谢基因簇, 比较基因组

Foundation item: Fundamental Research Funds for the Central Universities (2017ZY14)

***Corresponding author:** E-mail: mayuchao@bjfu.edu.cn

Received: 06-12-2020; **Accepted:** 01-03-2021; **Published online:** 11-06-2021

基金项目: 中央高校基本科研业务费(2017ZY14)

***通信作者:** E-mail: mayuchao@bjfu.edu.cn

收稿日期: 2020-12-06; 接受日期: 2021-03-01; 网络首发日期: 2021-06-11

Genome sequencing and comparative genome analysis of *Streptomyces* sp. SAT1

WANG Sha¹ DOU Guiming² MA Yuchao^{*1}

¹ Beijing Forestry University, Beijing 100089, China

² Chinese Academy of Forestry, Beijing 100089, China

Abstract: **[Background]** *Streptomyces* has always been the main producer of bioactive compounds. However, as the abusing of antibiotics, environmental pollution and drug resistance are becoming increasingly serious problem. The discovery of efficient bio-control factor and novel antibiotics becomes the main methods to solve these problems. **[Objective]** Obtain the whole genome sequence of the *Streptomyces* sp. SAT1 and the information about its secondary metabolite gene clusters; analyze the particularity and generality with other streptomycetes by the technology of comparative genomics and pan genomics. Form this, we could provide theoretical basis for illuminating the mechanism of bacteriostasis and growth-promoting in SAT1, and reliable data to reveal the ecological function of *Streptomyces*. **[Methods]** The sequence of SAT1 was completed by the third generation sequencing platform PacBio Sequel, then annotated and classified by bio-information technology; the software RAXML and PGAP was used to construct phylogenetic tree and analyze pan-genome, respectively. The prediction and analyze of the secondary metabolite gene clusters was achieved by antiSMASH. **[Results]** From the complete genome map of SAT1, the length of linear chromosome is 7.47 Mb, with 73% GC content, and four plasmids exist in the strain. Additionally, there are 7 550 genes which encoded proteins and 37 secondary metabolite gene clusters which classified by 29 types in SAT1. And the moenomycin gene cluster was highly homologous to *Streptomyces ghanaensis* ATCC14672 moenomycin gene cluster. In the 42 streptomycetes, it exists about 20–55 secondary metabolites gene clusters in each strain which classified into PKS, Terpene, Nrps and Heterozygous gene clusters. The dispensable genome was huge in these research objectives. **[Conclusion]** *Streptomyces* sp. SAT1 has many common points in the trait of genome and secondary gene clusters with other streptomycetes. We speculate the moenomycin and hygromycin_B gene clusters play an important role in the antibacterial activity of SAT1. In the 42 research objectives, the number of gene clusters and the size of genome has a positive correlation. In addition, the existence of abundant heterozygous gene clusters and large number of dispensable genome illustrate *Streptomyces* has high levels of horizon gene transfer over long periods of evolution, which possesses important environmental functions.

Keywords: *Streptomyces*, whole genome sequencing, secondary metabolites gene clusters, comparative genomics

植物内生链霉菌不但次级代谢产物丰富,而且与宿主长期协同进化形成了良好的互利共生关系,是研究植物-微生物互作机制的良好材料,同时也是作为促生及防病生物制剂应用于农业领域的潜在资源^[1]。司洪阳筛选得到一株抗真菌效果良好的链霉菌,经纯化分析后发现该菌能够合成具有较高抑菌活性的新型化合物 6-(5-Hydroxy-6-Methylheptyl)-5,6-Dihydro-2H-Pyran-2-One^[2];葛优优等研究发现内生链霉菌 SSD49 具有很好的植物

促生和防病效果^[3]。尽管链霉菌属基因组内存在丰富的次级代谢基因簇(几个到几十个),但在实验条件下有些基因簇沉默^[4],阻碍了新型活性物质的发现和链霉菌应用潜力的开发。而且,链霉菌基因组较大(4.5–12 Mb),GC 含量高(>70%),代谢调控机制复杂^[5],非模式菌株的遗传操作较为困难。

随着测序技术和生物信息学技术的逐渐成熟,利用基因组挖掘法可以准确地预测链霉菌次级代谢基因簇的数量和类型、编码基因的位置、

大小、序列信息和功能,以及代谢调控方式和代谢产物的骨架结构等^[6],这有利于沉默基因簇的诱发表达、目标化合物的选择和纯化,从而避免重复分离已知化合物,更快捷地发现新型代谢物。例如,法呢基二苯并二杂酮 ECO-04601^[7]就是利用基因组挖掘技术从小单孢子菌属中分离得到,目前正在进行抗实体瘤一期临床试验,此外,通过分析菌株 *Streptosporangium* sp. CGMCC 4.7309 的基因组,发现了 II 型聚酮类化合物合成基因,并通过基因簇敲除以及互补实验证实了该菌株所产生的活性物质的类型^[8]。综上,将基因组挖掘与传统天然产物发现策略相结合,大大提高了发现新型化合物的准确率和效率,也能够更加深入地研究内生链霉菌的进化方式和在生态系统的重要功能。

实验室前期从药用植物茅苍 (*Adenophora trachelioides*) 的根部分离出一株内生链霉菌 SAT1^[9],通过平板对峙实验发现该菌株对细菌以及真菌型病原菌具有较好的抑菌效果。为更好地研究 SAT1 的抑菌机制和链霉菌的生态功能,通过三代测序技术完成了 SAT1 菌株的全基因组测序、功能基因注释和次级代谢基因簇分析;同时将该基因组与 NCBI Genome 数据库中的 41 株链霉菌基因组进行比较分析,揭示 SAT1 基因组和次级代谢基因簇特异性,比较分析了链霉菌基因组和次级代谢的共性特征,为深入研究 SAT1 的抑菌和共生机制提供可靠依据,以期为揭示链霉菌的生态功能提供理论基础。

1 材料与方法

1.1 SAT1 的基因组测序和拼接

链霉菌 SAT1 全基因组测序平台为 PacBio Sequel^[10],由天津诺禾致源生物信息科技有限公司完成。

1.2 链霉菌 SAT1 的基因组组分分析

使用 GeneMarkS^[11]软件预测基因组的编码基因;通过 RepeatMasker^[12]和 TRF 软件分别预测散在重复序列和串联重复序列;利用 rRNA 库比对或 rRNAmmer 软件^[13]预测 rRNA;利用 tRNAscan-SE

软件^[14]预测 tRNA 区域和 tRNA 的二级结构;分别使用 Rfam 软件^[15]和 IslandPath-DIOMB^[16]预测 sRNA 序列和基因岛。

1.3 SAT1 的基因组功能注释

(1) GO 数据库注释^[17]:使用 IPRscan 软件比对得到细胞组分(Cellular Component)、分子功能(Molecular Function)和生物学途径(Biological Process)这 3 种功能基因数目的整体分布。

(2) COG (Cluster of Orthologous Groups of Proteins)数据库注释^[18-19]:通过 BLAST 将链霉菌 SAT1 的氨基酸序列与 COG 数据库进行比对,推测出蛋白序列的功能。

(3) KEGG (Kyoto Encyclopedia of Genes and Genomes)数据库注释^[20-22]:通过 BLAST 将 SAT1 氨基酸序列与 KEGG 数据库进行比对,找到与行使某一类功能相关(KEGG Ontology, KO)的所有注释上的基因。

1.4 SAT1 基因组的比较分析

1.4.1 完成图基本信息比较

在 NCBI 数据库中搜索早期已经进行过全基因组测序的链霉菌,按照时间顺序选择具有一定代表性的链霉菌,共 41 株,下载其完成图,比较包含 SAT1 在内的 42 株链霉菌的基因组基本信息。

1.4.2 系统发育分析

以 *Kitasatospora setae* KM-6054 为外群,随机提取所有菌株的 100 个不同类型的核心单拷贝蛋白,使用 ClustalW 软件进行比对,连接成 Super-Protein Sequences。使用 RAxML (Randomized Axelerated Maximum Likelihood)构建系统发育树软件,自展值(Bootstrap),即检验所计算的进化树分支可信度设置为 100,建树模型为 PROTGAMMALGF,构建最大相似性系统进化树。

1.5 SAT1 菌株中次级代谢基因簇的预测与分析比较

次级代谢基因簇预测:利用 antiSMASH 网站 (<http://antismash.secondarymetabolites.org/>)进行预测;目标基因簇基因片段比对:利用 NCBI 网站中 BLAST 工具对目标基因片段分别进行比对;数据处理:利用 Origin 8.0 进行统计分析。

1.6 泛基因组分析

利用 PGAP (Pan-Genomes Analysis Pipeline) 软件对包括 SAT1 菌株在内的 42 株链霉菌进行泛基因组分析。首先通过相应的 perl 程序准备 PGAP 能识别的核苷酸序列、氨基酸序列以及对应的功能文件；选择 Map-to-Pan 法，即基于参考基因组的重头组装方法，选择默认参数(覆盖度 50%、比对分值 50)进行直系同源基因鉴定，进而完成功能基因聚类分析，最终输出数据并进行分析统计。

2 结果与分析

2.1 SAT1 菌株的基因组测序及分析

通过第三代测序技术获得了拮抗菌株 SAT1

的表观修饰分布圈图(图 1)，各圈从外至内依次代表：基因组位置、正义修饰位点分布、反义链修饰位点分布、GC 含量和 GC 偏好性。该基因组包含 1 条线性染色体和 4 个质粒，染色体长度 7 472 530 bp，GC 含量 73.15%；质粒 Plas1 1 095 bp，GC 含量 75.80%；质粒 Plas2 343 282 bp，GC 含量 72.38%；质粒 Plas3 840 523 bp，GC 含量 73.34%；质粒 Plas4 92 538 bp，GC 含量 70.03%。共含有 6 个 rRNA 操纵元(18 个 rRNA 基因)和 69 个 tRNA 基因。基因组中共预测出 7 550 个蛋白编码基因，其个数占整个基因组的 86.11% (表 1)。

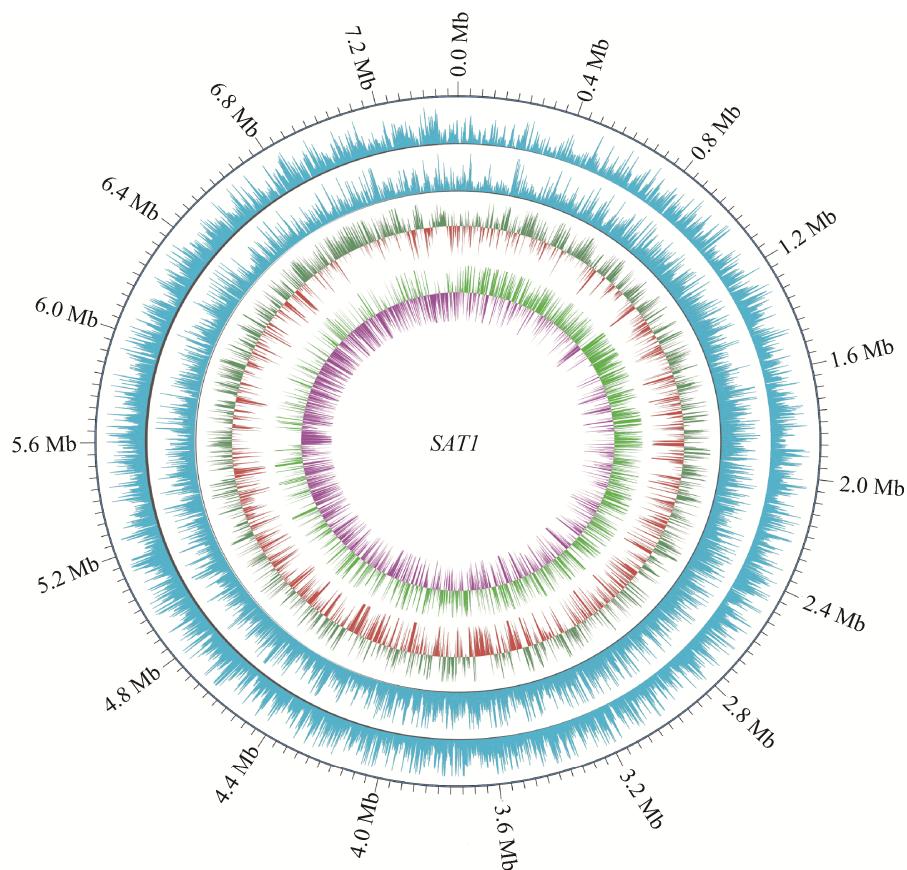


图 1 菌株 *Streptomyces* sp. SAT1 的表观修饰分布圈图

Figure 1 Epigenetic modifications of genome sequencing of *Streptomyces* sp. SAT1

表 1 *Streptomyces* sp. SAT1 的基因组特征

Table 1 Genomic features of *Streptomyces* sp. SAT1

Features	Chromosome	Plas1	Plas2	Plas3	Plas4
Length (bp)	7 472 530	1 095	343 282	840 523	92 538
GC content (%)	73.15	75.80	72.38	73.34	70.03
CDS	6 441	0	338	664	107
rRNA (Operons)	18 (6)	0	0	0	0
tRNA	69	0	0	0	0
Accession No.	CP015849	CP016081	CP016082	CP016083	CP016084

2.2 SAT1 菌株的基因组功能注释

在 SAT1 菌株中, 共有 5 066 个基因在 GO 数据库获得注释(图 2), 注释结果分为 3 类, 即分子功能 (Molecular Function), 细胞组分 (Cellular Component)和生物学过程(Biological Process)。在分子功能中, 参与 Binding 和 Catalytic Activity 的基因比例比较大, 分别为 34.56%和 37.07%。在细胞组分类别中, 参与 Cell Part 和 Cell Junction 的独立基因比较多, 占比均约 20%。然而, 在生物学过程这一分子功能中, 参与 Metabolic Process 和 Cellular Process 的基因较多, 分别为 39.48%和

35.03%。

通过与 COG 数据库进行比对, 共有 5 395 个基因获得注释结果(图 3), 分为 24 类, 其中 Transcription 和 General Function Prediction Only 类别含有基因最多, 分别为 763 个和 724 个, 占全部注释基因的 14.14%和 13.42%。其余功能基因占比高于 5%的分类有 Carbohydrate Transport and Metabolism (539, 9.99%)、Amino Acid Transport and Metabolism (534, 9.90%)、Signal Transduction Mechanisms (444, 8.23%)、Coenzyme Transport and Metabolism (356, 6.60%)、Energy Production

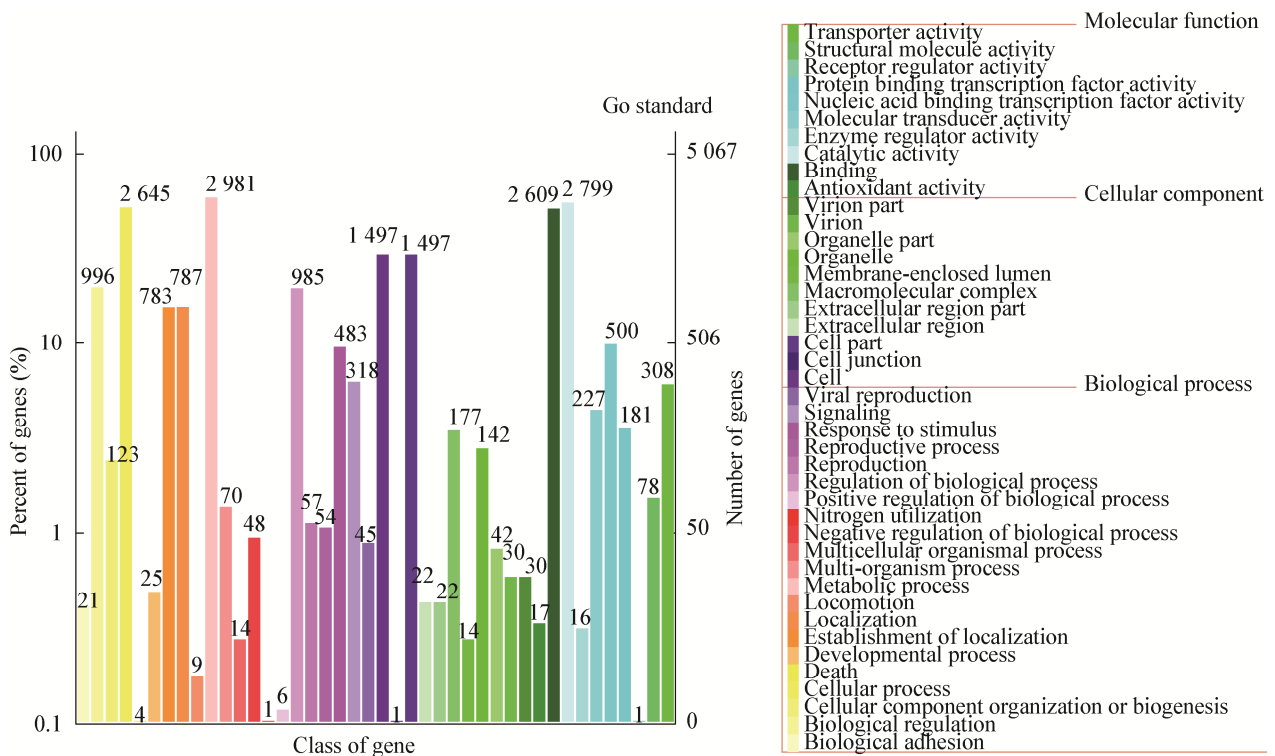


图 2 数据库 GO 的功能注释

Figure 2 Gene annotation and functional classification in GO database

and Conversion (335, 6.21%)、Lipid Transport and Metabolism (328, 6.08%)、Inorganic Ion Transport and Metabolism (321, 5.95%)、Cell Wall/Membrane/Envelope Biogenesis (292, 5.41%)、Translation, Ribosomal Structure and Biogenesis (286, 5.30%)和 Secondary Metabolites Biosynthesis, Transport and Catabolism (285, 5.28%)。占比最少的类别为 RNA Processing and Modification (2, 0.04%)和 Chromatin Structure and Dynamics (2, 0.04%)。

共有 3 239 个基因在 KEGG 数据库中获得注释结果,其中有 811 个基因未得到明确分类(图 4),这些基因共参与 115 个代谢通路,主要参与 Membrane Transport (68 KOs)、Nucleotide Metabolism (66 KOs)和 Translation (52 KOs)。

2.3 链霉菌 SAT1 基因组的比较分析

2.3.1 链霉菌完成图基本信息比较

对包含 SAT1 在内的共 42 株菌基因组完成图基本信息进行比较(表 2),发现基因组大小范围为

5.96–11.94 Mb,其中,基因组最小和最大的菌株分别为 *S. xiamenensis* strain 318 和 *S. bingchenggensis* BCW-1,蛋白质编码范围 5 340–9 793 个,GC 含量均高于 70%,其中 GC 含量最高的菌株为 *S. albus* J1074 和 *S. albus* strain SM254 (73.3%)。多数链霉菌为线性(Linear)染色体,并含有 1–4 个质粒(23 株),仅 6 株是环状(Circular)染色体,包括 *S. violaceusniger* Tu 4113、*Streptomyces* sp. SirexAA-E、*S. collinus* Tu 365、*S. fulvissimus* DSM 40593^T、*Streptomyces* sp. CNQ-509 和 *Streptomyces* sp. CCM_MD2014。

2.3.2 系统发育分析

基于随机选取的 100 个单拷贝蛋白,以 *Kitasatospora setae* KM-6054 (链霉菌近源属模式菌)为外群,利用 RAxML 软件构建 43 株菌(42 株链霉菌和 1 株外群菌株)的最大相似性系统进化树(图 5)。结果显示, *Kitasatospora setae* KM-6054 与其他的链霉菌菌株形成 2 个独立的分支;41 株链霉菌分成了 4 个大的分支;菌株 *S. cattleya* NRRL

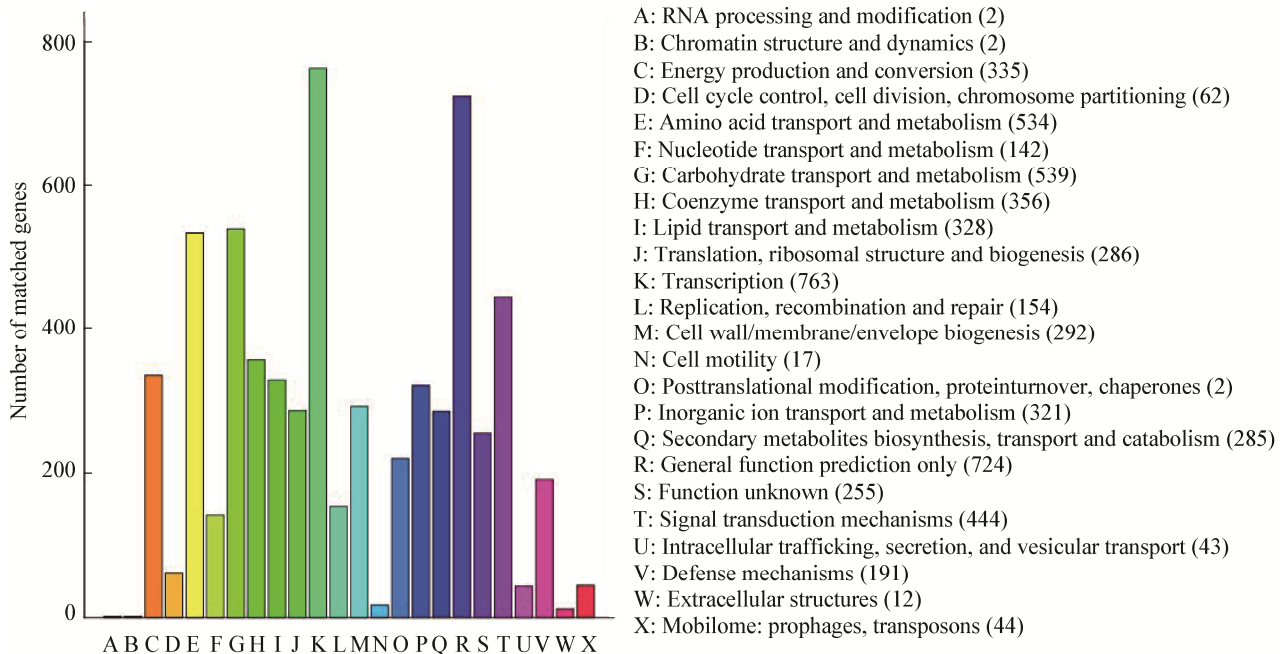


图 3 数据库 COG 的功能注释

Figure 3 Gene annotation and functional classification in COG database

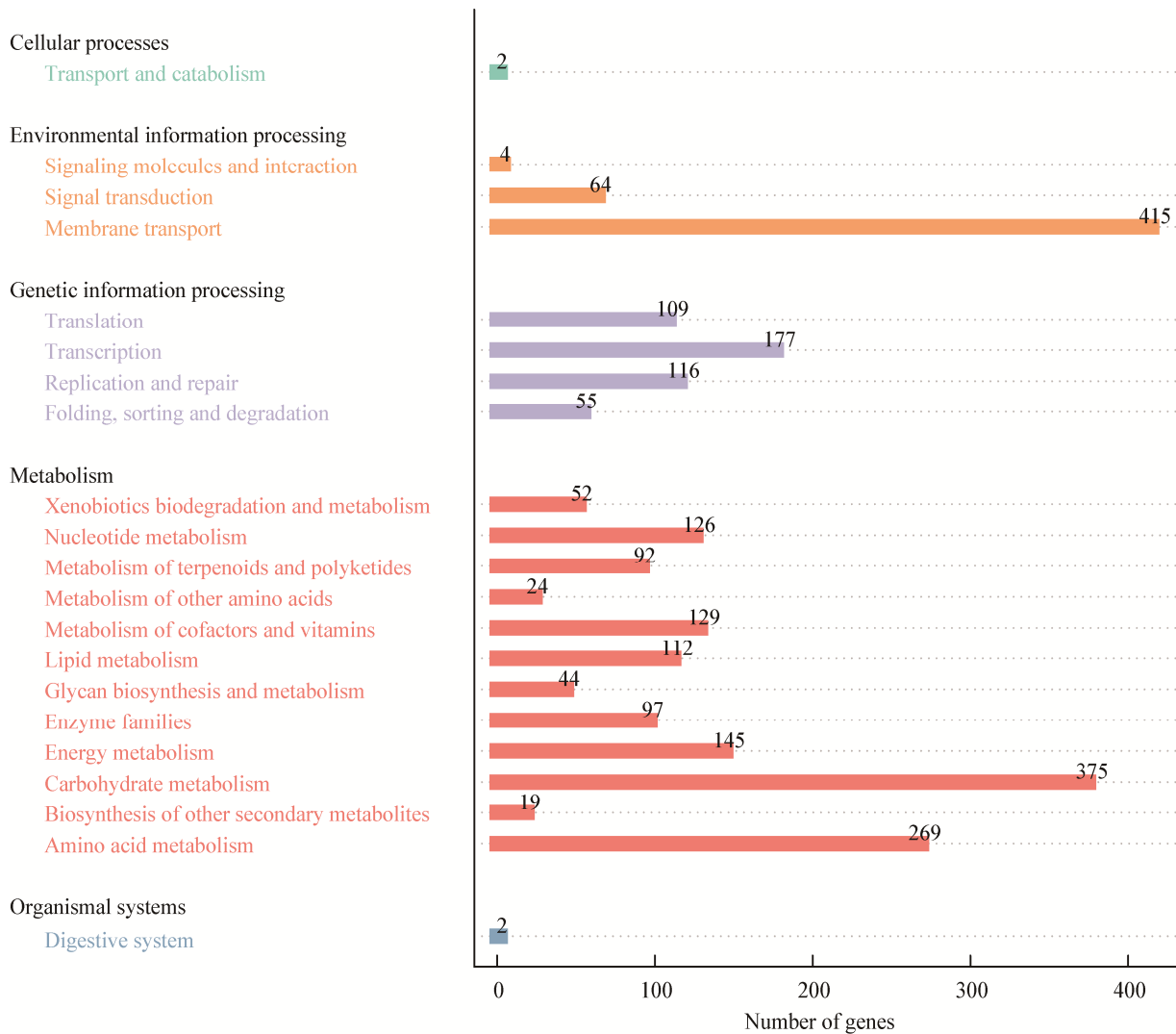


图4 数据库 KEGG 的功能注释

Figure 4 Gene annotation and functional classification in KEGG database

8057^T=DSM 46488^T 处于链霉菌属的边缘。菌株 *Streptomyces* sp. SAT1 与 *S. collinus* Tu 365 亲缘关系最近, 并与 *S. reticuli*、*S. hygroscopicus* subsp. *limoneus* strain KCTC 1717、*S. hygroscopicus* subsp. *jinggangensis* TL01 和 *S. hygroscopicus* subsp. *jinggangensis* 5008 等共同形成一个较大分支。

2.4 SAT1 菌株中次级代谢基因簇的预测与分析比较

2.4.1 SAT1 菌株次级代谢基因簇的预测

通过 antiSMASH 网站预测, *Streptomyces* sp. SAT1 基因组共含有 11 种类型的 29 个已知次级代

谢基因簇(表 3)。PKS 类、Terpene 类和 Nrps 类是主要类型, 共预测到 19 种可能次级代谢产物。其中, 相似性不低于 50% 的次级代谢基因簇的预测产物有 Spore Pigment、2-Methylisoborneol、Hopene、Albaflavenone、Coelibactin、Informatipeptin、Nataxazole、Ectoine、Melanin、Hygromycin_B、Moenomycin、Isorenieratene、Bafilomycin、Coelichelin、Herboxidiene、Sceliphrolactam、Lactonamycin、Scabichelin、Streptobactin、Geosmin 和 Apoptolidin。

表 2 42 株链霉菌基因组基本信息

Table 2 Complete genome sequencing of 42 *Streptomyces* strains

序号 No.	菌株 Strain	NCBI 序列编号 NCBI accession No.	基因组大小 Size (Mb)	质粒数目 Plasmid number	质粒名称 Plasmid name	质粒大小 Plasmid size (Mb)
1	<i>S. violaceusniger</i> Tu 4113	NC_015957.1	10.66	2	pSTRVI01 pSTRVI02	0.290 0.190
2	<i>S. bingchenggensis</i> BCW-1	NC_016582.1	11.94			
3	<i>Streptomyces</i> sp. SirexAA-E	NC_015953.1	7.41			
4	<i>S. griseus</i> subsp. <i>griseus</i> NBRC 13350	NC_010572.1	8.55			
5	<i>S. avermitilis</i> NBRC 14893	NC_003155.5	9.03	1	Plsm SAP1	0.090
6	<i>S. scabiei</i> 87.22	NC_013929.1	10.15			
7	<i>S. coelicolor</i> A3(2)	NC_003888.3	8.67	2	SCP1 SCP2	0.360 0.030
8	<i>S. albus</i> J1074	NC_020990.1	6.84			
9	<i>S. cattleya</i> DSM 46488 ^T	NC_016111.1	6.28	1	pSCATT	1.180
10	<i>S. collinus</i> Tu 365	NC_021985.1	8.27	2	pSOC1 pSOC2	0.090 0.020
11	<i>S. davawensis</i> strain JCM4913	NC_020504.1	9.47	1	pSDA1	0.090
12	<i>S. fulvissimus</i> DSM 40593 ^T	NC_021177.1	7.91			
13	<i>S. hygrosopicus</i> subsp. <i>jinggangensis</i> 5008	NC_021177.1	10.15	2	pSHJG1 pSHJG2	0.160 0.070
14	<i>S. hygrosopicus</i> subsp. <i>jinggangensis</i> TL01	NC_020895.1	9.84	2	pSHJGH1 pSHJGH2	0.160 0.070
15	<i>Streptomyces</i> sp. PAMC26508	NC_021055.1	7.53	1	Psp01	0.100
16	<i>Streptomyces albus</i> strain SM254	NZ_CP014485.1	7.17			
17	<i>S. reticuli</i> LN997842.1	LN997842.1	8.35	3	Plsm1 Plsm2 Plsm3	0.810 0.090 0.080
18	<i>S. globisporus</i> C-1027	NZ_CP013738.1	7.61	2	SGLP1 PSGL1	0.170 0.010
19	<i>S. venezuelae</i> strain ATCC 15439 ^T	NZ_CP013129.1	9.05			
20	<i>S. hygrosopicus</i> subsp. <i>limoneus</i> strain KCTC 1717	NZ_CP013219.1	8.65	1	Chr II	1.890
21	<i>Streptomyces</i> sp. CNQ-509	NZ_CP011492.1	8.04			
22	<i>S. ambofaciens</i> ATCC 23877 ^T	NZ_CP012382.1	8.30	1	pSAM1	0.090
23	<i>S. pristinaespiralis</i> strain HCCB 10218	NZ_CP011340.1	8.53			
24	<i>Streptomyces</i> sp. CFMR 7	NZ_CP011522.1	8.21	1	Unnamed	0.100
25	<i>S. glaucescens</i> strain GLA.O	NZ_CP009438.1	7.45	1	pSglau1	0.170
26	<i>S. lividans</i> TK24	NZ_CP009124.1	8.35			
27	<i>Streptomyces</i> sp. CCM_MD2014	NZ_CP009754.1	8.27			
28	<i>S. pratensis</i> ATCC 33331 ^T	NC_016114.1	7.34	2	pSFLA01 pSFLA02	0.190 0.130
29	<i>S. cyaneogriseus</i> subsp. <i>noncyanogenus</i> strain NMWT 1	NZ_CP010849.1	7.76			
30	<i>S. vietnamensis</i> strain GIM4.0001	NZ_CP010407.1	8.87	1	pSVL 1	0.290
31	<i>S. leeuwenhoekii</i> sleC34	NZ_LN831790.1	7.90	2	Psle1 Psle2	0.090 0.130
32	<i>S. xiamenensis</i> strain 318	NZ_CP009922.3	5.96			
33	<i>S. albulus</i> strain NK660	NZ_CP007574.1	9.36	1	pNKL	0.010
34	<i>S. albus</i> strain DSM 41398 ^T	NZ_CP016825.1	8.38			
35	<i>S. lydicus</i> A02	NZ_CP007699.2	9.30			
36	<i>Streptomyces</i> sp. CdTB01	NZ_CP013743.1	9.90	1	Plsm	0.290
37	<i>Streptomyces</i> sp. Mg1	NZ_CP011664.1	7.87	3	Psmg1-1 Psmg1-2 Psmg1-3	0.530 0.130 0.180
38	<i>Streptomyces</i> sp. 769	NZ_CP003987.1	10.10	1	pSGZL	0.240
39	<i>Streptomyces</i> sp. 4F	NZ_CP013142	8.05			
40	<i>Streptomyces</i> sp. S10 (2016)	HF674977	9.08			
41	<i>S. albulus</i> ZPM	NZ_CP006871	9.78			
42	<i>Streptomyces</i> sp. SAT1	NZ_CP015849.1	7.47	4	Plas1 Plas2 Plas3 Plas4	0.001 0.340 0.840 0.090

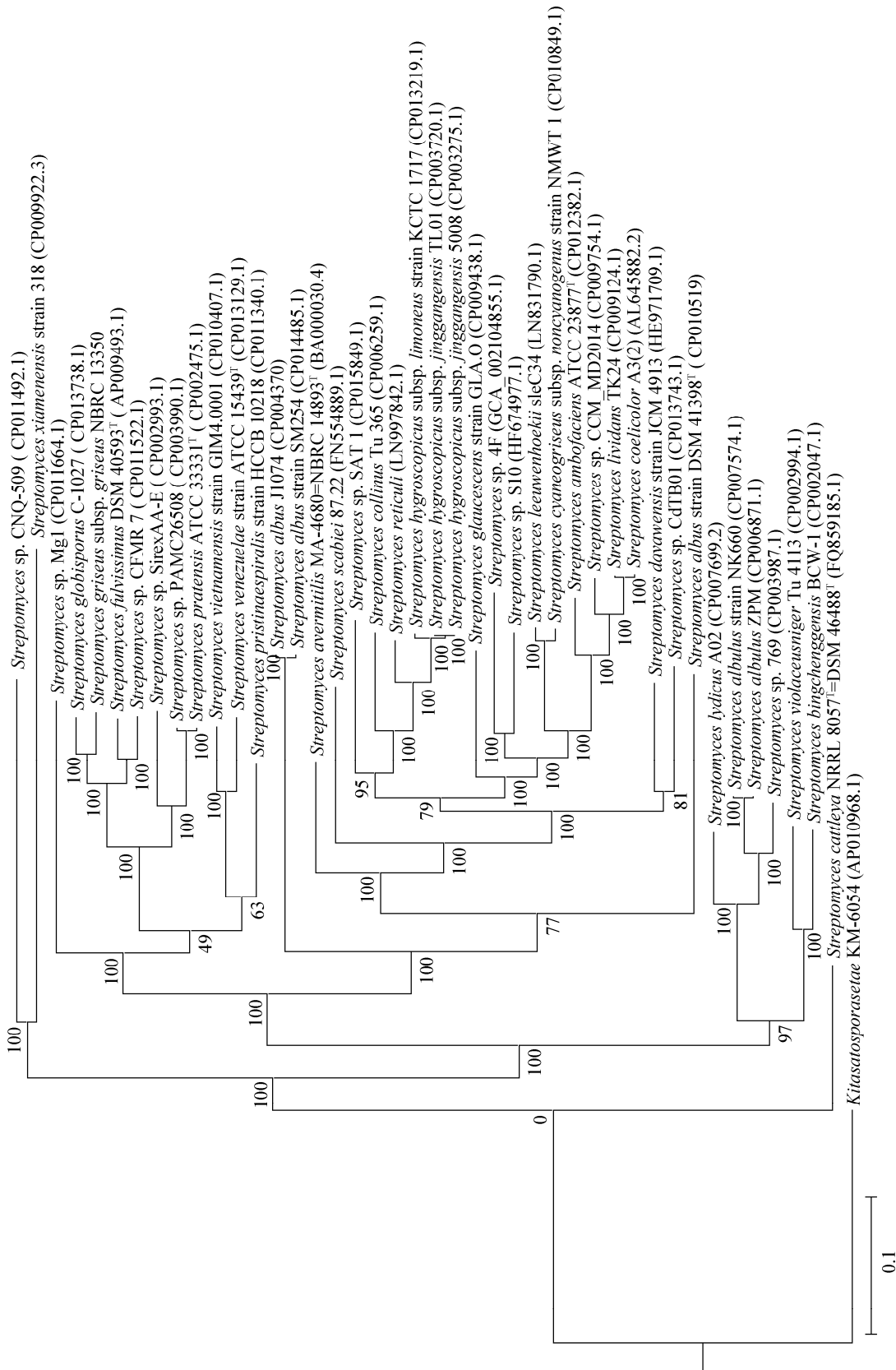


图5 基于单拷贝蛋白构建的最大相似性系统进化树

Figure 5 Phylogenetic tree based on single-copy core proteins using ML method

注: 标尺代表由序列差异表现的分支长度; 节点处的数字代表自展值; 括号内的编号是各基因序列的GenBank登录号

Note: The bar shows the branch length represented by sequence differences; Numbers at the node represent the Bootstrap values; Numbers in parenthesis are the GenBank accession number of each gene sequence

表 3 SAT1 基因组中次级代谢基因簇的预测
Table 3 Prediction of the secondary metabolites in *Streptomyces* sp. SAT1

类型 Type	预测产生的物质 Predicted production	相似性 Similarity (%)
PKS	Apoptolidin	53
	Sceliphrolactam	16
	Herboxidiene	57
	Spore pigment	83
	Nataxazole	100
	Bafilomycin	94
	Sceliphrolactam	80
	Lactonamycin	57
	Botryenolol	20
	Terpene	2-methylisborneol
Geosmin		100
Hopene		92
Albaflavenone		100
Isorenieratene		85
Bacteriocin		Informatipeptin
Nrps	Coelibactin	100
	Coelichelin	100
	Scabichelin	100
	Stenothricin	28
	Streptobactin	58
	CC-1065	9
Nrps-Pks	Oxazolomycin	12
	Cinnabaramide	27
Lantipeptide	Lantipeptide	2
Butyrolactone	Butyrolactone	14
Ectoine	Ectoine	100
Melanin	Melanin	100
Amglyccycl	Hygromycin_B	78
Phosphoglycolipid	Moenomycin	90

2.4.2 SAT1 菌株中默诺霉素基因簇同源性比对

根据 antiSMASH 预测结果, SAT1 菌株中存在的默诺霉素基因簇与加纳链霉菌(*Streptomyces ghanaensis* ATCC14672)中默诺霉素基因簇的相似性到了 90%以上。将 2 菌株默诺霉素因簇中各个基因片段比对后发现, 除 *moeR5*、*moeS5* 这 2 个基因外, 其他 21 个 *moe* 基因片段的相似程度均 $\geq 69\%$, 2 菌株中默诺霉素基因簇具有高度的同源性(表 4), 推测 SAT1 菌株具备产生默诺霉素物

质的能力。

2.4.3 SAT1 菌株与其他链霉菌次级代谢基因簇的比较分析

经 antiSMASH 预测和分析, 42 株链霉菌次级代谢基因簇类型丰富, 单个菌株基因簇数量多。在这些链霉菌中, 预测到有近 1 500 个次级代谢基因簇, 单个菌株的基因簇数量范围为 20–55 个(图 6), 不同菌株次级代谢基因簇数目的差异性呈现基因组越大次级代谢基因簇越多的趋势。按 antiSMASH 的分类标准, 这些次级代谢基因簇分属于 28 个大类, 而且每个菌株都含有 10 种类型以上的基因簇。PKS 类、Terpene 类和 Nrps 类为主要类型, 共有 584 个, 约占总数的 41.3%。值得注意的是, 基因簇中含有大量的杂合基因簇(Heterozygous Cluster), 而且杂合类型多样, 如 Nrps-T1pks-Others、Bacteriocin-Terpene-Nrps 及 Nrps-Pks 等。

2.5 泛基因组分析

如图 7 所示, 每个椭圆形代表一个菌株, 中间重合的圆圈中的数字表示核心基因家族和核心基因(括号中)的数目, 非重合部分显示了菌株的特有基因家族和基因(括号中)的数目。42 株链霉菌共有核心基因 1 610 个, 约占总基因数目的 17%–32%; 聚类分析得到 51 370 个直系同源基因簇, 其中 28 074 个直系同源基因簇只存在于某一特定菌株, 而这些直系同源基因簇约占泛基因组的 55%。

3 讨论

链霉菌属是生物活性物质的主要来源。*Streptomyces* sp. SAT1 是从药用植物莽苣(*Adenophora trachelioides*)中筛选得到, 该菌株抑菌谱较广且抑菌活性较强, 对真菌型病原菌以及细菌型病原菌均有较好的抑菌效果, 此外, 实验室前期研究发现该菌株的发酵液粗提物和挥发性物质对辣椒疫霉菌丝的生长有明显的抑制作用, 接种链霉菌 SAT1 的孢子液可提高辣椒苗对辣椒疫霉的抗性并明显促进辣椒苗的生长^[23], 由此可见,

表 4 SAT1 与加纳链霉菌默诺霉素基因簇比对

Table 4 The comparison about the moenomycin gene cluster of SAT1 and *Streptomyces ghanaensis* ATCC14672

<i>S. ghanaensis</i> ATCC14672			<i>Streptomyces</i> sp. SAT1		Predicted function
Gene	Protein accession No. (NCBI)	ORF (aa)	ORF (aa)	aa similarity (%)	
<i>moeA4</i>	ABJ90146.1	516	501	73	Long-chain fatty acid--CoA ligase
<i>moeB4</i>	ABJ90147.1	521	514	72	Amide synthetase
<i>moeC4</i>	ABJ90148.1	474	414	82	Aminolevulinate synthase
<i>moeS5</i>	ABJ90171.1	282	307	43	Hexose-4-ketoreductase
<i>moeR5</i>	ABJ90170.1	374	333	41	Hexose-4,6-dehydratase
<i>moeGT3</i>	ABJ90169.1	414	292	79	Glycosyltransferase
<i>moeGT2</i>	ABJ90168.1	286	272	88	Glycosyltransferase family 2 protein
<i>moeGT5</i>	ABJ90167.1	312	400	71	Glycosyltransferase
<i>moeP5</i>	ABJ90166.1	223	223	80	ABC transporter ATP-binding protein
<i>moeX5</i>	ABJ90165.1	226	226	82	ABC transporter membrane protein
<i>moeO5</i>	ABJ90164.1	281	281	79	Prenyl 3-phosphoglycerate synthase
<i>moeN5</i>	ABJ90163.1	260	260	73	Prenyltransferase
<i>moeM5</i>	ABJ90162.1	530	530	85	Carbamoyltransferase
<i>moeGT4</i>	ABJ90161.1	427	424	73	Glycosyltransferase
<i>moeK5</i>	ABJ90160.1	407	503	94	Radical SAM protein (methyltransferase)
<i>moeH5</i>	ABJ90159.1	513	505	81	Aminotransferase subunit 2
<i>moeGT1</i>	ABJ90158.1	402	408	82	Glycosyltransferase
<i>moeF5</i>	ABJ90157.1	645	648	77	Aminotransferase subunit 1
<i>moeE5</i>	ABJ90156.1	340	331	80	NDP-hexose 4-epimerase
<i>moeJ5</i>	ABJ90155.1	564	574	79	ABC transporter ATP-binding protein
<i>moeD5</i>	ABJ90154.1	591	699	79	ABC transporter ATP-binding protein
<i>moeA5</i>	ABJ90153.1	394	408	84	5-aminolevulinate synthase
<i>moeB5</i>	ABJ90152.1	301	527	69	Long-chain fatty acid--CoA ligase

该菌株作为植物内生链霉菌在生物防治应用方面具有一定的潜力。

根据测序结果发现 SAT1 菌株中含有 4 个大小不等的质粒, 其中 Plas2 以及 Plas3 质粒中分别有 1 个和 7 个基因簇。在链霉菌中, 质粒不但能够编码产生活性物质, 还可能具有某些调控基因以控制链霉菌的新陈代谢的作用, 此外, 链霉菌自身的质粒也可以用来构建新的基因高效表达载体, 作为链霉菌基因克隆的工具^[24], 以上为后续针对 SAT1 菌株的进一步研究及改造提供了理论基础。

通过 SAT1 全基因组测序以及 antiSMASH 预测

结果可以看出, 该菌株基因组较大, GC 含量高, 而且次级代谢基因簇种类和数量都较多, 这些次级代谢基因簇所编码的产物是 SAT1 菌株的主要抑菌活性来源。其中, 潮霉素 B (Hygromycin_B) 基因簇和默诺霉素 (Moenomycin) 基因簇为拮抗菌 SAT1 所特有, 田文佳^[9]通过平板对峙实验和发酵液抑菌实验发现该菌株及其发酵液能够对包含真菌、细菌在内的 8 种病原菌: 栗疫菌 (*Cryphonectria parasitica*)、杨树溃疡病真菌型 7897/7926 (*Botryosphaeria dothidea* 7897/7926)、苹果轮纹病病原菌 (*Botryosphaeria berengeriana*)、辣椒疫

霉(*Phytophthora capsici*)、细菌型杨树溃疡病原菌(*Lonsdalea quercina* subsp. *populi* N-5-1)、耐甲氧西林金黄色葡萄球菌(MRSA)和白假丝酵母(*Candida albicans* ATCC 10231)产生拮抗作用。

根据默诺霉素基因簇中各个基因片段的比对结果可以看出,除 *moeR5*、*moeS5* 基因外, SAT1 菌株的默诺霉素基因簇的其余基因均与加纳链霉菌(*Streptomyces ghanaensis* ATCC 14672)中相对应的基因具有较程度的相似性,具有一定合成默诺霉素类物质的能力。默诺霉素是一种磷酸糖脂类抗生素,该物质能够直接作用于细菌细胞壁合成过程中的转糖基酶,效价远远高于临床上使用的万古霉素,但由于该物质结构中具有长链脂肪醇结构,半衰期较长而无法应用于临床^[25],如果能够得到不含有该结构的

默诺霉素同类化合物,将为医用抗生素增添新的一员,但 SAT1 菌株所产默诺霉素类型还需要进一步研究。

近年来,随着抗生素的滥用,病原菌抗药性也逐渐增强,虽然不断有研究发现链霉菌所产生的未知抑菌活性物质,但作为新型抗生素的研究报道少之又少。Núñez-Montero 等通过对链霉菌 *Streptomyces fildesensis* So13.3 的基因组进行测序分析,发现该菌株中至少有 11 个未知基因簇,对其菌体粗提物中的活性成分进行质谱分析与 GNP 天然产物数据库对比后发现,质荷比在 148–624 之间有一未知低极性组分,不排除是新型活性物质^[26]。为挖掘 SAT1 菌株与其他链霉菌的特性及共性,本研究对 42 株链霉菌的次级代谢基因簇进行了预测以及比较分析,

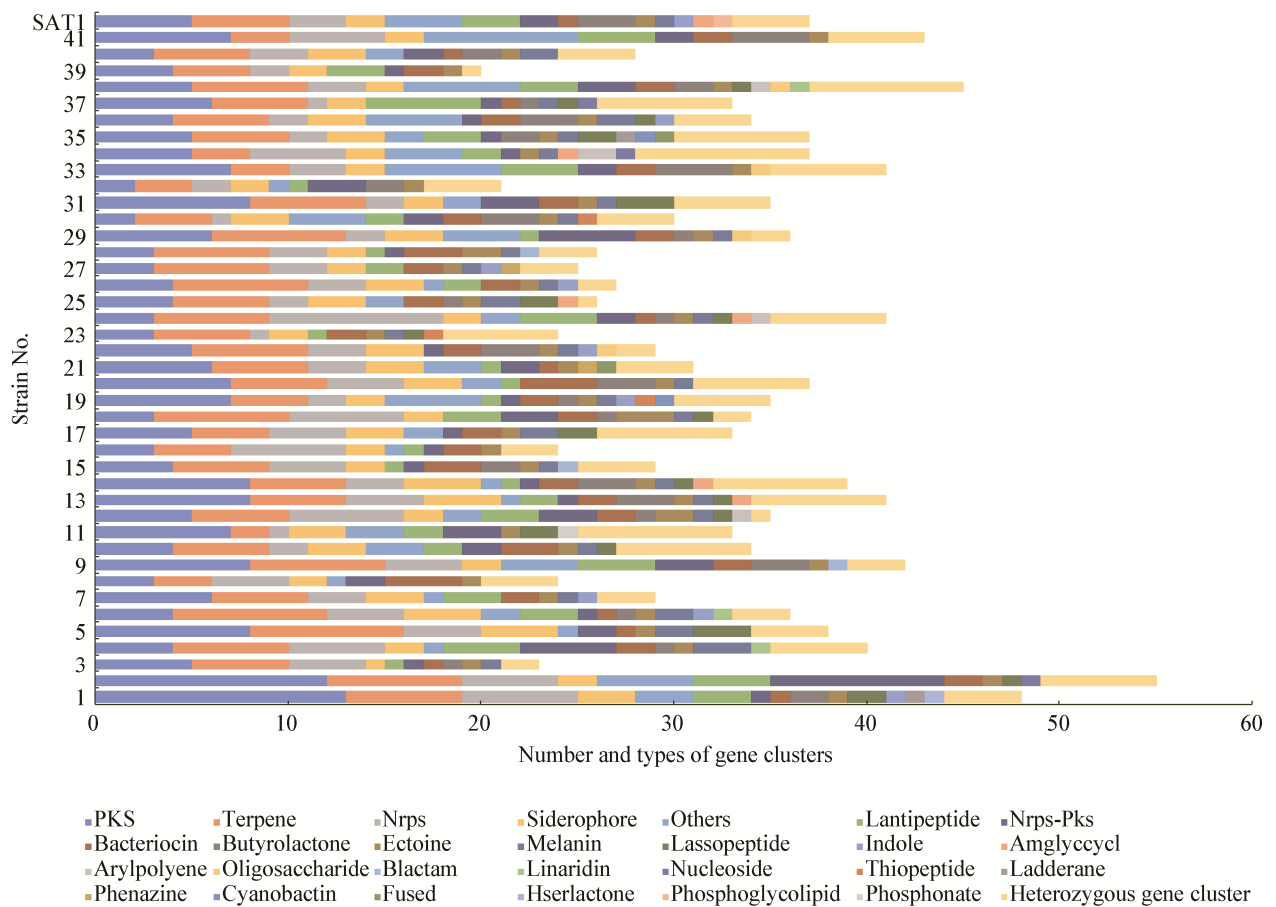


图 6 42 株链霉菌基因组中次级代谢基因簇的分布

Figure 6 The distribution of the secondary metabolic gene clusters in 42 *Streptomyces* genomes

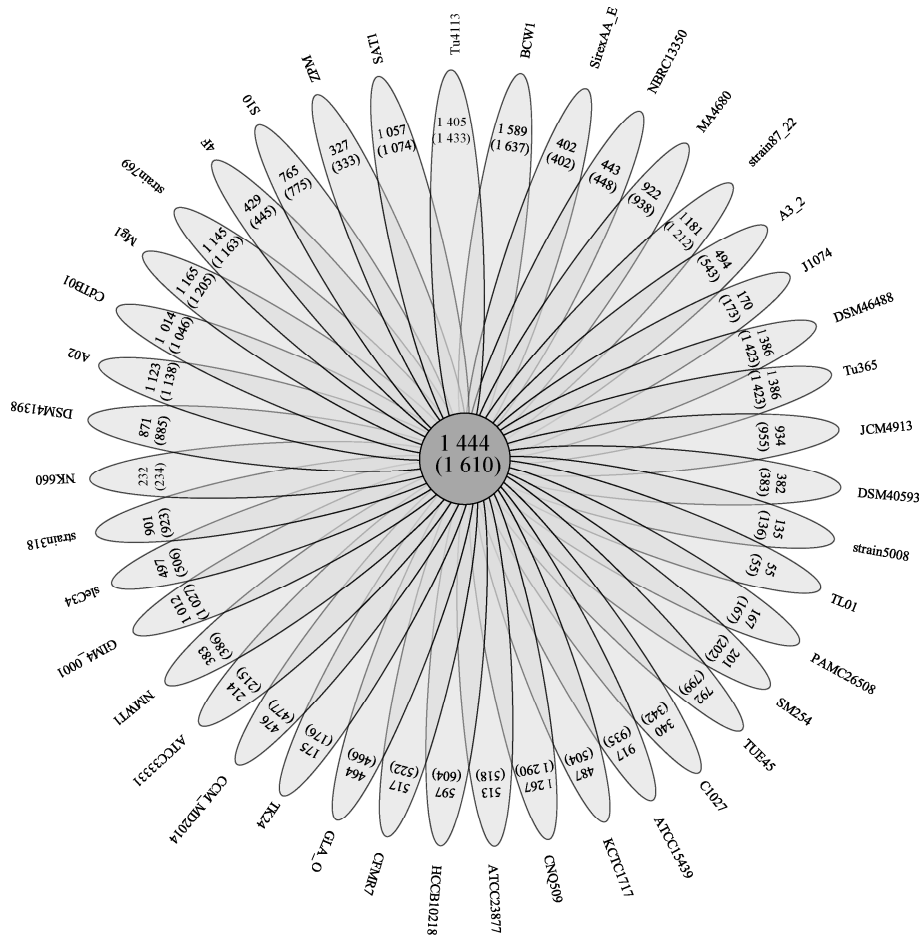


图 7 42 株链霉菌泛基因组分析结果
Figure 7 Analysis of Pan-genome diversity of 42 *Streptomyces*

发现在以上链霉菌中，均存有不同数量类型的未知基因簇，而且在这些次级代谢基因簇中，PKS 类和 Nrps 类最多，这 2 类基因簇指导合成的大环内酯类和非核糖体多肽类多为活性物质。链霉菌之间次级代谢基因簇的差异性可能与其复杂的生活环境有关，是适应复杂的生长环境的一种体现。基于此，本研究也从侧面证实了链霉菌天然活性物质的多产性，以及从链霉菌中挖掘新型抗生素的可能性。

通过泛基因组分析能够挖掘单一菌株的特有基因，特有基因能够使菌株具有区别于其他菌株的表型，比如植物内生、抑菌、促生等，使菌株对环境具有更好的适应能力。研究特有

基因可以解释菌株某些方面的能力(如抑菌机制)等。通过对以上菌株进行泛基因组分析，结果表明 55%的直系同源基因簇只存在于某一特定菌株，说明该属的特有基因数目庞大，在基因组层面上具有很高的杂合性。此外，42 株链霉菌共有 1 610 个核心基因，大约占到总基因数目的 17%–32%，这些核心基因很可能涉及菌体的基础代谢功能，维持其最基本生命特征，而 SAT1 菌株的 COG 功能注释结果也印证了我们的猜想(图 3)，显示多为参与离子转运(无机离子转运与代谢等)、核酸相关(转录，DNA 复制、结合与修复)等使细胞具有基础的代谢功能、维持细胞最基本生命特征等功能的途径。同时，次

级代谢基因簇预测结果表明, 以上 42 株链霉菌均含有四氢嘧啶(Ectoine)类次级代谢基因簇, 大部分微生物在长期的进化过程中都形成了在胞质内积累相容性溶质来对抗外界不断变化的渗透压机制, 其中四氢嘧啶类物质就是一种重要的相容性溶质, 其使链霉菌更好地适应复杂多变的环境。我们还发现该属中存有大量的杂合基因簇, 这表明链霉菌在长期进化过程中存在较高等度的水平基因转移现象, 具有重要的生态功能。

4 结论

内生链霉菌 SAT1 的线性染色体长度约 7.47 Mb, 包含有 4 个质粒, GC 含量近 73%, 含有 37 个次级代谢基因簇, 分属 28 个类型, 与本研究中的其余链霉菌具有很大的共同点, 而且基因簇数目及类型都较多, 其中默诺霉素基因簇和潮霉素 B 基因簇可能与该菌的抑菌活性有关, 而从默诺霉素基因簇的比对结果来看, 该菌株具备一定合成默诺霉素类物质的能力, 后续可以对 SAT1 菌株所产默诺霉素的类型进行进一步研究。此外, 本研究发现 42 株链霉菌中都存有大量的杂合基因簇, 而且特有基因数目庞大, 说明该属在长期进化的过程中具有较高等度的水平基因转移现象。本研究为阐明 SAT1 的内生机制和进一步开发 SAT1 菌株作为微生物生防菌剂提供有力依据, 同时为揭示链霉菌的生态功能提供了理论基础。

REFERENCES

- [1] Liu YX, Zhao Y, Zhang BX, Yang YM, Fan ST, Li CY, Wang Y, Xu PL, Qin HY, Lu WP. Research progress on the source and biological function of plant endophytic bacteria[J]. Special Wild Economic Animal and Plant Research, 2020, 42(4): 60-67 (in Chinese)
刘迎雪, 赵滢, 张宝香, 杨义明, 范书田, 李昌禹, 王月, 许培磊, 秦红艳, 路文鹏. 植物内生细菌来源及生物学功能研究进展[J]. 特产研究, 2020, 42(4): 60-67
- [2] Si HY. Isolation, purification and structure identification of active metabolites from *Streptomyces fungicidicus*[D]. Shenyang: Master's Thesis of Shenyang Agricultural University, 2019 (in Chinese)
司洪阳. 杀真菌链霉菌(*Streptomyces fungicidicus*)代谢产物的分离纯化及结构鉴定[D]. 沈阳: 沈阳农业大学硕士学位论文, 2019
- [3] Ge YY, Liu XY, Dou GM, Ma YC. Antimicrobial activities and efficacy of endophytic *Streptomyces* sp. SSD49 in plant disease control and plant-growth-promoting[J]. Biotechnology Bulletin, 2017, 33(6): 121-127 (in Chinese)
葛优优, 刘晓瑜, 窦桂铭, 马玉超. 内生链霉菌 SSD49 的抑菌活性和防病促生效果[J]. 生物技术通报, 2017, 33(6): 121-127
- [4] Paterson J, Jahanshah G, Li Y, Wang Q, Mehnaz S, Gross H. The contribution of genome mining strategies to the understanding of active principles of PGPR strains[J]. FEMS Microbiology Ecology, 2017, 93(3): fiw249
- [5] Hopwood DA. Soil to genomics: the *Streptomyces* chromosome[J]. Annual Review of Genetics, 2006, 40(1): 1-23
- [6] Yin Y, Ge M, Chen DJ. New methods, new technologies and discovery of novel antibiotics[J]. Microbiology China, 2013, 40(10): 1874-1884 (in Chinese)
殷瑜, 戈梅, 陈代杰. 新方法新技术与新型抗生素发现[J]. 微生物学通报, 2013, 40(10): 1874-1884
- [7] Farnet CM, Dimitriadou V, Bachmann BO. Farnesyl dibenzodiazepinones and methods of treating cancer using same: US, US7186713[P]. 2007-03-06
- [8] Tian J, Chen HY, Guo ZY, Liu N, Li JE, Huang Y, Xiang WS, Chen YH. Discovery of pentangular polyphenols hexaricins A-C from marine *Streptosporangium* sp. CGMCC 4.7309 by genome mining[J]. Applied Microbiology and Biotechnology, 2016, 100(9): 4189-4199
- [9] Tian WJ. The antimicrobial activity of the endophytic *Streptomyces* sp. SAT1 and the regulatory function of *wblA*[D]. Beijing: Master's Thesis of Beijing Forestry University, 2019 (in Chinese)
田文佳. 内生链霉菌 SAT1 的抑菌活性特征及 *wblA* 基因的调控功能[D]. 北京: 北京林业大学硕士学位论文, 2019
- [10] Berlin K, Koren S, Chin CS, Drake JP, Landolin JM, Phillippy AM. Correction: corrigendum: assembling large genomes with single-molecule sequencing and locality-sensitive hashing[J]. Nature Biotechnology, 2015, 33(10): 623-630
- [11] 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

- [12] Saha S, Bridges S, Magbanua ZV, Peterson DG. Empirical comparison of ab initio repeat finding programs[J]. *Nucleic Acids Research*, 2008, 36(7): 2284-2294
- [13] 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
- [14] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence[J]. *Nucleic Acids Research*, 1997, 25(5): 955-964
- [15] Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, et al. Rfam: updates to the RNA families database[J]. *Nucleic Acids Research*, 2009, 37(Database issue): D136-D140
- [16] Bertelli C, Brinkman FSL. Improved genomic island predictions with IslandPath-DIMOB[J]. *Bioinformatics*, 2018, 34(13): 2161-2167
- [17] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology[J]. *Nature Genetics*, 2000, 25(1): 25-29
- [18] Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, et al. The COG database: an updated version includes eukaryotes[J]. *BMC Bioinformatics*, 2003, 4(1): 41
- [19] Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families[J]. *Science*, 1997, 278(5338): 631-637
- [20] Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. From genomics to chemical genomics: new developments in KEGG[J]. *Nucleic Acids Research*, 2006, 34(Suppl_1): D354-D357
- [21] Kanehisa M. A database for post-genome analysis[J]. *Trends in Genetics*, 1997, 13(9): 375-376
- [22] Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome[J]. *Nucleic Acids Research*, 2004, 32(Database issue): D277-D280
- [23] Liu XY, Dou GM, Ma YC. Screening and identification of antagonistic SAT1 and its potential for biocontrol of pepper *Phytophthora* blight[J]. *Guangdong Agricultural Sciences*, 2014, 41(19): 75-79,85 (in Chinese)
刘晓瑜, 窦桂铭, 马玉超. 生防链霉菌 SAT1 的分离、鉴定及其对辣椒疫霉的生物防治潜力[J]. *广东农业科学*, 2014, 41(19): 75-79,85
- [24] DiCenzo GC, Finan TM. The divided bacterial genome: structure, function, and evolution[J]. *Microbiology and Molecular Biology Reviews*, 2017, 81(3): e00019-17
- [25] Ostash B, Doud EH, Lin C, Ostash I, Perlstein DL, Fuse S, Wolpert M, Kahne D, Walker S. Complete characterization of the seventeen step moenomycin biosynthetic pathway[J]. *Biochemistry*, 2009, 48(37): 8830-8841
- [26] Núñez-Montero K, Lamilla C, Abanto M, Maruyama F, Jorquera MA, Santos A, Martínez-Urtaza J, Barrientos L. Antarctic *Streptomyces fildesensis* So13.3 strain as a promising source for antimicrobials discovery[J]. *Scientific Reports*, 2019, 9(1): 7488