

陆紫云,查双龙,刘江崑,等.地芽孢杆菌(*Geobacillus*) YHL的全基因组测序及序列分析[J].江西师范大学学报(自然科学版),2022,46(2):147-155.

LU Ziyun, ZHA Shuanglong, LIU Jiangyin, et al. The whole-genome sequencing and sequence analysis of *Geobacillus* sp. YHL [J]. Journal of Jiangxi Normal University (Natural Science), 2022, 46(2): 147-155.

文章编号:1000-5862(2022)02-0147-09

地芽孢杆菌(*Geobacillus*) YHL 的全基因组测序及序列分析

陆紫云,查双龙,刘江崑,马羊帅,肖力婷,杨慧林*

(江西师范大学生命科学院,江西省亚热带植物资源保护与利用重点实验室,江西南昌 330022)

摘要:该文采用 Illumina 高通量测序技术对地芽孢杆菌(*Geobacillus* sp. YHL)进行全基因组测序,使用 Velet 软件进行组装,利用 Glimmer 软件对菌株进行基因预测,得到的蛋白质通过与 COG、KEGG 等数据库进行比对来获得相应的注释信息.利用多种绘图工具对注释信息进行汇总及分析,获得了 COG、KEGG 等多种基础注释信息,对这些信息进行挖掘分析,研究结果发现:该菌株具有多种编码酶基因,包括糖苷水解酶、葡糖苷酶、木聚糖酶、淀粉酶、新普鲁兰酶、支链淀粉酶和脂肪酶,是一种嗜热的多酶编码菌,有一定的应用潜力.重点关注了在基因组中编码热应激蛋白基因,这些基因信息最终可以提供关于细菌的热适应机制的初步解释.

关键词:嗜热微生物;地芽孢杆菌;全基因组测序;基因组分析

中图分类号:TS 252.54 **文献标志码:**A **DOI:**10.16357/j.cnki.issn1000-5862.2022.02.06

0 引言

嗜热微生物是指可在温度较高的环境中生存并繁殖的一类微生物群体,它们在火山、温泉等自然栖息地和堆肥、工业生产等人为环境中都有一定的分布.许多嗜热微生物具有适于生物技术和商业应用的特性^[1](如具有多种热稳定酶^[2]),可用于工业应用,对某些矿物具有特殊的浸溶能力^[3]及生物修复能力^[4].

Geobacillus 菌属是在 2001 年被 T. N. Nazina 等^[5]提出从 *Bacillus* 菌属中分离出来作为一个新的菌属,该菌属内成员专性嗜热,是典型的嗜热微生物菌群.迄今为止,该菌属在油田、火山口、温泉等自然环境和人为高温环境中都有发现,且该菌属物种在许多中温或低温地区也被分离得到^[6].该菌属在生物技术和工业领域中具有重要应用.近年来,*Geobacillus* 菌属受到研究者的广泛研究. R. E. Cripps 等^[7]

利用代谢工程方法对 2 株 *Geobacillus* 嗜热菌进行改造,最后得到高效生产乙醇的菌株; A. Verma 等^[8]纯化和表征菌种 *Geobacillus thermocatenulatus* MS5 中的漆酶,能够在纺织工业中广泛有效地去除对环境造成污染的染料,在合成染料的生物修复中发挥着重要作用;唐赞等^[9]发现可使苯酚降解的菌株; S. Bilge 等^[10]从嗜热菌中纯化和表征出耐高温淀粉酶等多种热稳定酶.

自从 2004 年 T. Hideto 等^[11]提出第 1 个完整的嗜热杆菌相关物种 *Geobacillus kaustophilus* HTA426 的基因组序列至今,已经有大约 200 个该菌属物种的基因组序列被进行高通量测序.目前,基于生物信息技术对基因组序列进行分析已成为人们广泛认可的分析方式.依托计算机学科和数学学科对基因组序列进行分析已成为普遍的分析方式^[12],生物信息学研究从发展至今已为许多科学研究提供了一个探索方向^[13].在利用生物信息数据研究生物的遗传密码以及物种进化的内容中^[14],对基因组的分析是

收稿日期:2021-06-03

基金项目:国家自然科学基金(31760449),江西省自然科学基金(20181BAB214003)和江西省教育厅研究生创新基金(YC2020-S184)资助项目.

通信作者:杨慧林(1986—),男,江西宜春人,副教授,博士,主要从事微生物多样性及生物信息学研究. E-mail: yanghl@jxnu.edu.cn

提供关于细菌如何应对高温的信息的重要一步,通过对基因组分析确定在其基因组中是否存在热相关蛋白,为微生物适应高温提供相关见解^[15].有研究者将 *Geobacillus* 菌属物种与其他嗜中温细菌基因组进行对比分析发现其部分特有基因塑造了其对温度的某种耐受性^[11],这也说明对于全基因组的测定并在进行比较基因分析中挖掘出有用的信息是非常必要的,这些独特基因对微生物适应极端环境至关重要.

本文研究的 *Geobacillus* 菌株是在工业管道中发现的 1 株嗜热菌,它能够在 55 °C 下生长繁殖.为了获得该嗜热菌的分子信息,初步了解其嗜热机制,本文利用 NGS 技术测定该菌株的全基因组序列,同时利用生物信息学软件预测菌株的功能基因,对其基因功能进行进一步分析,为挖掘其潜在的生物学意义提供基础.

1 材料与方法

1.1 菌株 YHL1 的培养与基因组 DNA 提取

将菌株 YHL 接至 LB 培养基中于 55 °C 条件下培养 2 d,然后在超净工作台中取出样品,在低温(4 °C)条件下离心后弃上清、收集菌体,基因组提取参照文献^[16]的方法,在获得高质量基因组后送至测序公司进行高通量测序.

1.2 基因组测序、组装及注释

在提取基因组 DNA 后,对其进行质量鉴定,利用 Qubit 3.0 对提取的 DNA 浓度进行测定,当质量达标后测定菌株的全基因组,测序平台为 Illumina HiSeq 2000.在测序完成后,利用拼接软件 Velvet 1.2.10^[17]优化 Kmer 值,将测定的序列进行组装.然后利用软件 Glimmer 3.02^[18]对菌株进行基因预测,得到的蛋白质通过与 COG 数据库、NR 数据库、Swiss-prot 数据库、Interpro 数据库、KEGG 和 GO 数据库进行比对来获得相应的注释信息.最后利用多种绘图工具对注释信息进行汇总.全基因组序列数据已提交到 NCBI,登录号为 JAEIGB000000000.

1.3 系统发育进化树构建

将 16s rRNA 序列的 contigs 进行拼接,基于 16s rRNA 的同源序列在 EZbio 上比对其同源序列,选取其亲缘关系最近的 18 株菌,使用 MAGE X 软件对共 19 株菌的 16s rDNA 序列构建 Neighbor-joining(NJ)系统进化树.

1.4 次级代谢产物合成基因簇分析

采用次级代谢产物合成基因簇在线预测软件

antiSMASH(<https://antismash.secondarymetabolites.org>)^[19]对菌株次级代谢产物进行预测,获得代谢产物的预测结果.

2 结果与分析

2.1 基因组组装

在通过高通量测序对菌株基因组进行测序后,使用 Velvet 1.2.10 软件对测序的片段进行拼接,最终得到 47 个 contigs,N50 的长度为 145 810 bp,N90 长度为 42 680 bp.基因序列全长为 3 426 484 bp,GC 含量为 52.35%,包含了 9 个 rRNA 和 85 个 tRNA.利用 Glimmer 预测软件对编码基因预测,预测得到编码基因 3 609 个,占总基因组的 86.19%,总长度为 2 953 284 bp,平均总长度为 818.31 bp.基因组圈图如图 1 所示.

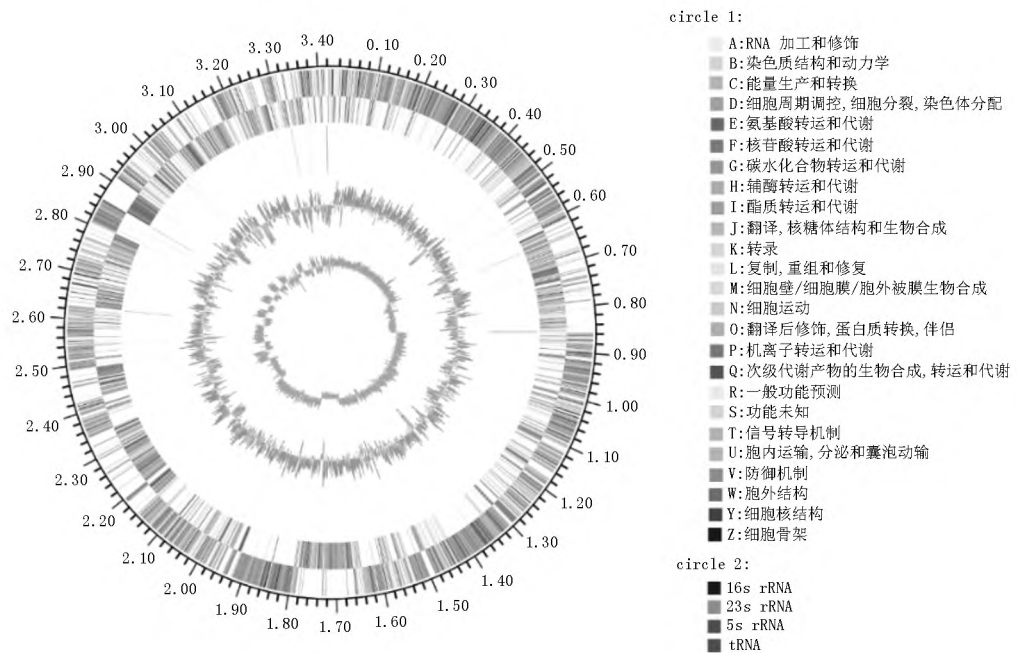
2.2 系统发育树

基于 16s rRNA 基因序列信息,使用 MAGE X 软件构建邻接(NJ)系统进化树(见图 2),以证明该菌株在谱系中的地位.根据进化树的分支距离可以发现:实验菌株和高温烷烃地芽孢杆菌 *Geobacillus thermoleovorans* KCTC 3570(T)具有较高的同源性,分布在同一个系统分支上.

目前在 NCBI 上可查询到的属于 *Geobacillus* 属的全基因组测序的菌株有 100 多株,对目前已完成全基因组序列的属于 *Geobacillus* 属的 31 株菌进行统计分析,结果如表 1 所示.从表 1 可以看出它们的基本信息大致相似.菌株全基因序列大小相似,GC 含量均在 52% 左右.本文研究的菌株 GC 含量为 52.35%.同时,基于 16s rRNA 基因构建系统发育树可以初步确定菌株 YHL 归类为 *Geobacillus* 属.本文研究菌株命名为 *Geobacillus* sp. YHL.

2.3 基因注释

利用 Glimmer 软件对 YHL 菌株的蛋白基因数进行预测,共获得 3 609 个;接着对上述的蛋白序列进行 COG 注释,将注释结果与 COG 数据库进行比对,设定其 $E_{\text{value}} \leq 1 \times 10^{-5}$,最终得到 3 051 个蛋白注释(见图 3).从图 3 可以看出:在 YHL 菌株的 COG 聚类中所有的假设基因和部分编码功能未知蛋白的基因对于基因组预测基因总数的占比最大.此外,预测得到的基因集中在氨基酸转运和代谢,碳水化合物转运和代谢,能量生产和转换,转录、复制、重组以及修复 5 个方面,它们分别对应图 3 中的 E、G、C、K、L.这也反映出菌株需要不断修复 DNA 和蛋白质来应对极端环境以确保自身的生存.



注:在 YHL 菌株基因组圈图中,最外圈为基因大小;第 2、第 3 圈分别为正、负链上的 CDS;第 4 圈为 rRNA 和 tRNA;第 5 圈为 GC 含量,指向外圈表示其 GC 含量高于平均水平;第 6 圈为 GC-skew 值。

图 1 *Geobacillus* sp. YHL 基因组圈图

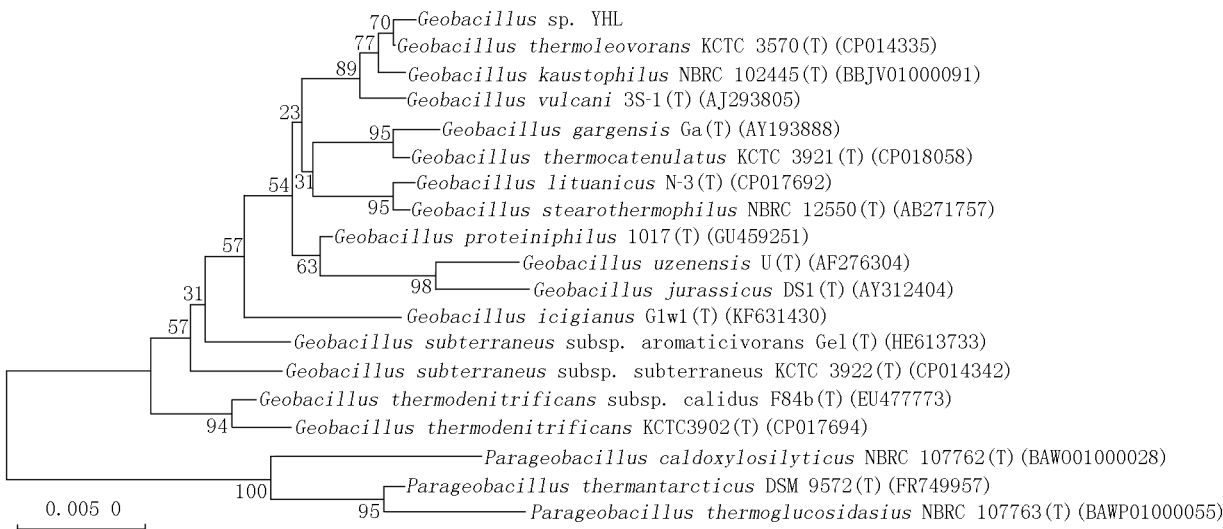


图 2 *Geobacillus* sp. YHL 邻接系统进化树

表 1 部分已完成全基因组测序的地芽孢菌属基本特征比较分析

菌株名称	大小/Mb	GC 含量/%	GenBank 登录号	框架	基因数量	蛋白数量
YHL	3.426 48	52.35	JAEIGB000000000	136	3 609	
cJS12	3.721 49	52.00	NZ_CP014749.1/CP014749.1	1	3 846	3 095
C56-T3	3.650 81	52.50	NC_014206.1/CP002050.1	1	3 750	3 439
Y412MC52	3.673 94	52.31	NC_014915.1/CP002442.1	2	3 760	3 421
Y412MC61	3.667 90	52.31	NC_013411.1/CP001794.1	2	3 750	3 458
GHH01	3.583 13	52.30	NC_020210.1/CP004008.1	1	3 682	3 313
C56-T2	3.545 94	52.40	NZ_VJWB000000000	3	3 534	3 169
FJAT-46040	3.361 15	52.30	NZ_NISS000000000	5	3 608	3 048
46C-IIa	3.474 91	52.10	NZ_NADR000000000	51	3 544	3 276
LEMMJ02	3.436 27	52.60	NZ_VKJO000000000	433	3 581	3 322
CAMR5420	3.499 82	51.90	NZ_JHUS000000000	96	3 646	3 306

表 1(续)

菌株名称	大小/Mb	GC 含量/%	GenBank 登录号	框架	基因数量	蛋白数量
Sah69	2.991 50	52.60	NZ_LLKS000000000	73	3 313	2 787
PK12	3.631 61	51.60	NZ_SDKL000000000	189	3 703	3 363
LEMMY01	3.586 07	51.90	NZ_MVKA000000000	76	3 849	3 387
A8	3.348 31	52.40	NZ_AUXP000000000	173	3 502	3 126
MAS1	3.497 41	52.20	NZ_AYSF000000000	121	3 754	3 263
47C-IIb	3.347 49	49.60	NZ_NADS000000000	167	3 411	3 185
A8	2.451 96	53.00	NZ_JZIT000000000	173	3 502	3 126
B4113_201601	3.651 13	51.30	NZ_LQYX000000000	122	3 743	3 326
1017	3.574 95	51.80	NZ_MQMG000000000	186	3 782	3 379
Manikaran-105	3.191 35	52.50	NZ_PIJG000000000	140	3 363	3 016
WSUCF-018B	3.229 18	52.50	NZ_PIJF000000000	145	3 386	3 053
FW23	3.486 83	52.20	NZ_JGCJ000000000	245	3 648	3 224
T6	3.661 36	51.90	NZ_LDNZ000000000	294	3 829	3 403
WSUCF1	3.402 38	52.20	NZ_ATCO000000000	346	3 483	2 866
ZGt-1	3.483 11	52.10	NZ_LDPD000000000	241	3 407	2 977
LC300	3.532 58	52.14	CP008903.1	2	3 330	2 847
12AMOR1	3.442 72	52.05	CP011832.1	2	3 511	3 357
CAMR12739	3.414 74	52.20	JHUR000000000	74	3 511	1 703
BCO2	3.417 03	52.20	LJAJ000000000	154	4 700	4 649
15	3.363 00	52.40	LVHZ000000000	254	3 414	3 100

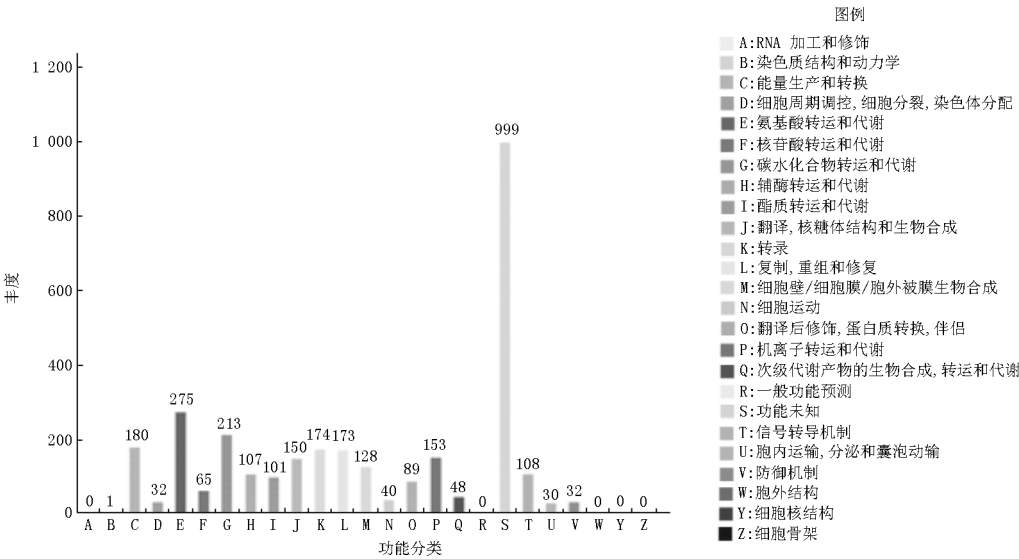
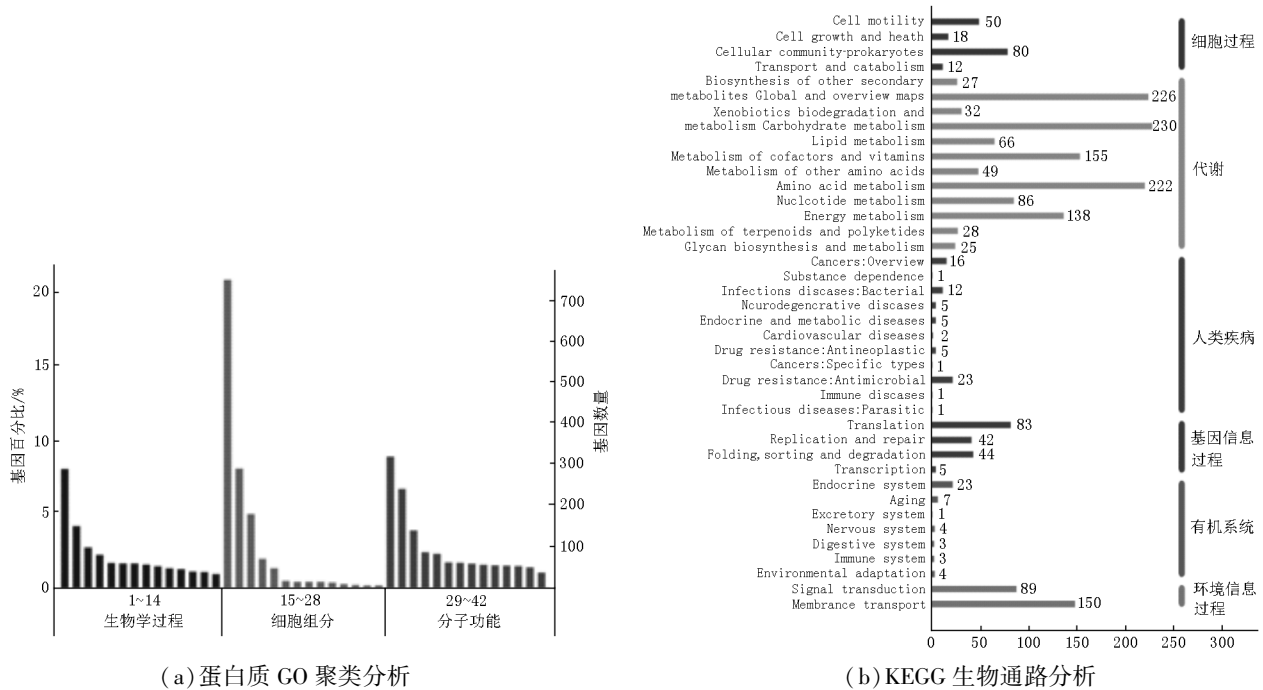


图 3 *Geobacillus* sp. YHL 蛋白质 COG 聚类分析

基于上述预测结果,利用 GO 注释对预测结果进行注释,同时通过在线工具 WEGO(<http://wego.genomics.org.cn/cgi-bin/wego/index.pl>)对预测的蛋白结果进行 GO 功能分类图的绘制(见图 4(a)),GO 功能分类(细胞组分、生物过程、分子功能)占据优势的基因数量主要是氧化还原过程(oxidation-reduction process)、膜的整体组成部分(integral component of membrane)、细胞质(cytoplasm)、细胞质膜(plasma membrane)、ATP 结合(ATP binding)、DNA 结合(DNA binding).数量众多的基因注释到细胞膜

和 DNA 方面.同 GO 注释一样,将预测蛋白与 KEGG 数据库进行比对(见图 4(b)),KEGG 生物通路主要分布在代谢(metabolism)、基因信息加工(genetic information processing)和环境信息过程(environmental information processing)中.在代谢途径通路中基因显著富集的有糖代谢(carbohydrate metabolism)、global and overview maps、氨基酸代谢(amino acid metabolism)、辅助因子和维生素的代谢(metabolism of co-factors and vitamins)、能量代谢(energy metabolism).而在环境信息过程中基因富集的有膜转运(mem-

brane transport). 在基因信息加工途径中基因富集 的有翻译(translation).



注:为简化图示,图(a)中立柱从左至右分别以1~42表示,其含义如下:1为氧化还原过程;2为转录调控,DNA模板;3为转运;4为蛋白质水解;5为磷脂酶信号转导系统;6为代谢过程;7为磷酸化;8为翻译;9为跨膜转运;10为碳水化合物代谢过程;11为核酸磷酸二酯键水解;12为甲基化;13为细胞分裂;14为蛋白质磷酸化的信号转导;15为细胞质膜组成成分;16为细胞质;17为细胞质膜;18为胞内;19为核糖体;20为细胞膜组成成分;21为细胞;22为膜;23为病毒衣壳;24为ABC结合盒转运体复合物;25为胞外区域;26为细菌型鞭毛基体;27为染色体;28为核糖体大亚基;29为ATP结合;30为DNA结合;31为金属离子结合;32为水解酶活性;33为转录因子活性,序列特异性DNA结合;34为镁离子结合;35为转运活性;36为ATP酶活性;37为锌离子结合;38为氧化还原酶;39为转移酶活性;40为核糖体结构成分;41为RNA结合;42为黄素腺嘌呤二核苷酸结合。

图4 *Geobacillus* sp. YHL 蛋白质 GO 聚类分析和 KEGG 生物通路分析

2.4 次级代谢产物分析

次级代谢产物是指微生物培养到生长后期通过代谢产生的物质,它往往具有一定的应用价值,对人类的生

产、生活具有一定意义. 因此通过 antiSMASH 软件对 YHL 菌株进行预测,获得该菌株在基因组中基因簇分布情况,共预测到 5 个次级代谢产物合成相关的基因簇(见表 2),包括聚酮(PKS)、萜类(terpene)、铁载体类(siderophore)、细菌素类(bacteriocin)。

表2 YHL 次级代谢产物合成基因簇预测

序号	重叠群	位置	基因数量	基因簇类型
1	contigs2	1-21353	36	T3PKS
2	contigs3	144631-154861	9	bacteriocin
3	contigs9	817-14750	12	siderophore
4	contigs18	14425-25278	13	bacteriocin
5	contigs35	1-16311	18	terpene

2.5 热相关候选基因

在菌株 YHL 基因组中发现了与温度应激相关的基因,包括热休克蛋白、冷休克蛋白和相关分子伴侣的基因,结果如表 3 所示. 从表 3 可以发现:其中热休克蛋白 Hsp20 有较多基因数量. 该蛋白属于一种小的热休克蛋白,也有研究发现 Hsp20 蛋白可能参与了低温和高温保护蛋白质的伴随过程, Hsp20

是一个重要的氧化应激和温度应激反应基因^[20], 这或许帮助菌株 YHL 细胞在高温中维持蛋白的稳定. Hsp100 家族对于蛋白质聚集物有清除功能. 最新研究发现在 Hsp100 家族中的成员之一的依赖 ATP 酶活性的分子伴侣 ClpB 可通过拉动作用将蛋白链从蛋白聚集物中单个拉出. 而当分子伴侣 ClpB 被去除以后,之前被拉出的蛋白链又可以重新进行折叠,同

时恢复原有的功能^[21]. 这对于由在高温影响下菌株 YHL 的蛋白质的错误折叠所产生的异常蛋白质的清除起到重要作用. 在高温刺激后的蛋白质聚集物的再活化对于修复菌体自身机能有着重要的意义, 这可以节省再翻译的时间以及避免消耗众多的能量. 蛋白质聚集物的再活化需要 1~2 个协同作用的伴侣系统. 在原核生物中, 由伴侣蛋白 DnaK、共伴侣 DnaJ、GrpE 和 Hsp100 伴侣蛋白 ClpB 组成. 伴侣间的协同工作是细菌应对环境的重要功能. 这或许是菌株 YHL 应对高温的重要策略之一. 在冷相关蛋白(如冷休克蛋白 Csp 家族)中的冷休克蛋白 CspB 富含 polyU 的 5'UTR 以高亲和力结合富含 T 的单链 DNA(ssDNA), 这可防止在温度下降时 mRNA 2 级结构的形成, 从而稳定 mRNA 结构, 有助于翻译的启动^[22-23]. 这些基因对于菌株 YHL 应对温度波动以维持细胞活力有重要作用.

2.6 菌株 YHL 的生物技术潜力

在工业上, 通过微生物优化发酵或重组 DNA 技术来克隆嗜热菌酶基因异源表达是获得工业酶的重要手段^[24]. 在菌株 YHL 中预测到多种编码酶基因, 这表明 YHL 是个多酶编码菌. YHL 能够编码糖苷水解酶、葡糖苷酶、木聚糖酶、淀粉酶、新型普鲁兰酶和脂肪酶. 此外, 编码多铜多酚氧化酶(漆酶)和抗癌酶谷氨酰胺酶(见表 4), 这些酶可用于食品、化学、制药和环境生物技术行业. α -淀粉酶和支链淀粉酶的组合在淀粉加工产业中扮演重要角色, 淀粉酶的热稳定性对伴随高温的糖化过程非常重要. 脂

肪酶广泛应用于皮革、乳制品和生物柴油生产工业中^[25]. 来自嗜热细菌 *Geobacillus thermocatenulatus* KCTC 3921 的脂解酶基因在大肠杆菌中异源表达出具有对长链脂肪酸有强脂解活性的热稳定脂肪酶^[26]. *L*-谷氨酰胺酶可以对抗肿瘤细胞, 它可将 *L*-谷氨酰胺水解为 *L*-谷氨酸和氨, 通过阻断从头蛋白质合成和通过促进癌细胞死亡的氧化应激增加超氧化物水平来选择性地抑制肿瘤生长, 是治疗癌症的有效药物^[27-28].

表 3 菌株 YHL 温度应激相关蛋白基因

温度应激相关蛋白	基因数
33 kDa 分子伴侣	1
伴侣蛋白热休克蛋白 HSP-70 辅助因子	1
分子伴侣(热休克蛋白 33)	1
冷休克蛋白	1
冷休克蛋白 CspB	2
热休克蛋白 33	1
热休克蛋白 Hsp33	1
热诱导转录抑制因子 hrcA	1
热休克蛋白 Hsp20	13
热休克蛋白 HtpX	2
分子伴侣 DnaJ	5
分子伴侣 DnaK	3
PBS 裂解酶类热重复蛋白	1
小分子热休克蛋白	2
ATP 依赖性伴侣 ClpB	5
ATP 依赖性 Clp 蛋白酶 ATP 结合亚基	6
ATP 依赖性 Clp 蛋白酶 ATP 结合亚基 ClpX	5
ATP 依赖性 clp 蛋白酶蛋白水解亚基	4

表 4 菌株 YHL 编码酶基因

基因功能描述	基因 ID	比对到的参考基因	基因长度
糖苷水解酶	gene1782	gi 505188142 ref WP_015375244. 1	83
	gene2243	gi 502910615 ref WP_013145591. 1	74
糖苷水解酶 43 家族蛋白	gene0890	gi 985489327 ref WP_060787919. 1	535
β -葡糖苷酶	gene0867	gi 519663231 ref WP_020279319. 1	470
	gene1358	gi 1004923764 gb KYD32640. 1	57
	gene3530	gi 502909841 ref WP_013144817. 1	455
耐热性的类 β -葡糖苷酶蛋白	gene0867	gi 190366247 gb ACE77068. 1	470
蛋白酶	gene0382	gi 499551244 ref WP_011232027. 1	422
α -葡糖苷酶	gene0867	gi 505186905 ref WP_015374007. 1	562
木聚糖酶	gene0912	gi 73332107 gb AAZ74783. 1	331
麦芽糖淀粉酶	gene1098	gi 306416106 gb ADM86931. 1	588
α -淀粉酶	gene1102	gi 295854559 gb ADG45817. 1	513
	gene1398	gi 2642326 gb AAB86961. 1	537
I 型普鲁兰酶	gene0085	gi 696487078 ref WP_033026023. 1	718
支链淀粉酶	gene1400	gi 696487448 ref WP_033026384. 1	1 583
GDSL 家族脂肪酶	gene0771	gi 530700091 gb EQB94529. 1	253

表 4(续)

基因功能描述	基因 ID	比对到的参考基因	基因长度
脂肪酶	gene0627	gi 499550354 ref WP_011231137.1	248
	gene0771	gi 499550478 ref WP_011231261.1	253
	gene0827	gi 923020781 ref WP_053414863.1	250
	gene0896	gi 502910224 ref WP_013145200.1	223
	gene3623	gi 503962055 ref WP_014196049.1	365
酰基甘油酯酶	gene0827	gi 505187814 ref WP_015374916.1	250
漆酶	gene1992	gi 499549847 ref WP_011230630.1	274
谷氨酰胺酶 A	gene3273	gi 765536756 ref WP_044743734.1	300
转谷氨酰胺酶	gene3459	gi 618755020 dbj GAJ57688.1	238
γ -谷氨酰胺转移酶	gene3459	gi 499551602 ref WP_011232385.1	238

3 讨论

本文对 1 株嗜热菌进行全基因组测序,获得 47 个 contigs,基于 16s rRNA,将其归类为 *Geobacillus* 属,对于确定到种,还需结合生理生化实验来确定. COG 聚类分析结果表明其主要代谢途径为氨基酸转运和代谢(E)、碳水化合物转运和代谢(G)2 个方面. 这表明其在高温环境中的基础生理活动代谢旺盛. 在 KEGG 分析中菌株 YHL 主要集中在代谢方面,在代谢条目下的子条目和数量都是最多的,这说明 YHL 具有旺盛的初级代谢过程,其中主要包括与机体提供能量和能源的糖代谢、为生命提供物质基础的氨基酸代谢、维持机体健康和促进生长的维生素的代谢以及辅助因子. 菌株 YHL 与 D. N. Singh 等^[2]研究的菌株 RL 有高相似的 COG 聚类情况;菌株 RL 是一种嗜热多酶编码细菌,在降解木质纤维素和废水脱色等方面有广泛作用;在 YHL 中也被证实含有多种碳水化合物酶基因,菌株 YHL 具有基因工程的可操作性和工业应用潜能. 菌株 YHL 在 COG 聚类分析中发现有高达 999 个蛋白未能进行归类,由此可推测因其中含有某些独特的基因而塑造了 YHL 耐高温的特点,菌株 YHL 的功能仍有待挖掘.

铁载体是微生物摄取在环境中的铁元素而形成的对铁离子高亲和的化合物. 目前关于铁载体生物合成的方式主要包括 2 种:(i)由非核糖体肽合成酶(non-ribosomal peptide synthetases, NRPSs)的模块化多酶家族指导合成;(ii)不依赖于 NRPS(NRPS-independent, NIS)指导合成. 以 NIS 合成途径合成的铁载体涉及一个由 Iuc 酶为主导的新合成酶家族,主要由 2 种铁载体合成酶 IucA 和 IucC 来催化形成关键酰胺键,而 IucA 和 IucC 这 2 个合成酶也成为 NIS 生物合成途径的重要标志^[29-32]. 目前,有关 NIS

的铁载体生物合成途径的研究仍处于起步阶段,需要进一步研究来了解 NIS 合成酶的底物特异性和产物控制的分子机制. 在 YHL 基因组中预测到的铁载体基因簇中发现了铁载体生物合成蛋白 IucA,这说明菌株 YHL 对在环境中铁元素的摄取是通过 NIS 合成途径来完成的. 众多研究表明:由 NIS 途径产生的铁载体是病原菌的毒力因子,可增强菌株致病性. 徐水宝等^[33]研究发现由 NIS 途径合成的 Aerobactin 可增强高毒力肺炎克雷伯菌的毒力. 联系到 YHL 菌株是在工业生产中发现的 1 株菌,可尝试以开发途径所涉及的通路蛋白抑制剂的方法来抑制该菌株,以达到在工业上有效清除杂菌的目的.

微生物基因组分析为人们提供由生活在热环境中的嗜热细菌遗传组成的信息. 菌株 YHL 含有许多编码热应激相关蛋白质基因,这些蛋白质可能对细菌适应高温具有重要意义. 基因组数据分析对确定微生物与温度适应相关的基因功能很重要. 研究表明:细菌通过上调热休克蛋白、增强蛋白质合成和降低碳分解代谢来抵抗高温^[34]. 在 YHL 中发现了许多热休克蛋白基因和蛋白质分子伴侣,并且发现了和其他菌株的热应激不同基因,在菌株 YHL 中存在 CspB,而在其他的菌株(如 *Geobacillus* sp. TFF3)中存在 CspD,它们同为 Csp 蛋白家族的一员,但 CspB 是冷激诱导的^[35],而 CspD 则由其他条件诱导产生(如出现在大肠杆菌营养缺乏时期)^[36]. 这说明不同的菌株对抗外界环境的作用模式不完全相似.

菌株 YHL 编码白色生物技术需要水解酶糖苷水解酶、葡糖苷酶、木聚糖酶、淀粉酶、新普鲁兰酶、支链淀粉酶和脂肪酶的混合物. 此外,编码绿色生物催化剂多铜多酚氧化酶(漆酶)和抗癌酶谷氨酰胺酶的基因的存在分别反映了菌株 YHL 对于灰色和红色生物技术的潜力. 克隆重组第 1 步要获得编码酶基因,而基因来源宿主的特性是选择的重要指标.

通过对嗜热菌来源的遗传操作来获得具有热稳定性的酶,这是常用的手段.随着嗜热菌等极端微生物的特性越来越被重视,越来越需要研究和遗传改造这些物种,并在各种生物技术和工业应用中利用它们的特性.

对菌株 YHL 进行全基因组测序而获得其相关信息是有必要的,完整的基因组研究可能带来新的见解(如此类细菌代谢的差异以及各种蛋白质和酶在分子水平上的功能和热稳定性),并将提供许多与嗜热细菌细胞各方面相关的信息,同时也可丰富 *Geobacillus* 属的物种遗传信息,为后续遗传操作提供基础.

4 参考文献

- [1] MEHTA R, SINGHAL P, SINGH H, et al. Insight into thermophiles and their wide-spectrum applications [J]. 3 Biotech, 2016, 6: 81.
- [2] SINGH D N, SOOD U, SINGH A K, et al. Genome sequencing revealed the biotechnological potential of an obligate thermophile *Geobacillus thermoleovorans* strain RL isolated from hot water spring [J]. Indian Journal of Microbiology, 2019, 59(3): 351-355.
- [3] VILCÁEZ J, SUTO K, INOUE C. Bioleaching of chalcopryrite with thermophiles: temperature-pH-ORP dependence [J]. International Journal of Mineral Processing, 2008, 88(1/2): 37-44.
- [4] 邢睿智, 赵子强, 赵文琪, 等. 嗜热脂肪地芽孢杆菌对聚苯乙烯的降解性能 [J]. 环境科学, 2021, 42(6): 3056-3062.
- [5] NAZINA T N, TOUROVA T P, POLTARAUS A B, et al. Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans* [J]. International Journal of Systematic and Evolutionary Microbiology, 2001, 51(2): 433-446.
- [6] RAHAMN T J, MARCHANT R, BANAT I M. Distribution and molecular investigation of highly thermophilic bacteria associated with cool soil environments [J]. Biochemical Society Transactions, 2004, 32(2): 209-213.
- [7] CRIPPS R E, ELEY K, LEAK D J, et al. Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production [J]. Metabolic Engineering, 2009, 11(6): 398-408.
- [8] VERMA A, SHIRKOT P. Purification and characterization of thermostable laccase from thermophilic *Geobacillus thermocatenulatus* MS5 and its application in removal of textiles dyes [J]. Journal of Biosciences, 2014, 2(8): 479-485.
- [9] 唐赞, 刘沐之, 梁凤来, 等. 一株嗜热菌的分离鉴定及其苯酚降解特性 [J]. 微生物学通报, 2006, 33(5): 39-44.
- [10] SARI B, FAIZ O, GENÇ B, et al. New xylanolytic enzyme from *Geobacillus galactosidasius* BS61 from a geothermal resource in Turkey [J]. International Journal of Biological Macromolecules, 2018, 119: 1017-1026.
- [11] TAKAMI H, TAKAKI Y, CHEE G J, et al. Thermoadaptation trait revealed by the genome sequence of thermophilic *Geobacillus kaustophilus* [J]. Nucleic Acids Research, 2004, 32(21): 6292-6303.
- [12] 王正华, 王勇献. 后基因组时代生物信息学的新进展 [J]. 国防科技大学学报, 2003, 25(1): 1-6, 44.
- [13] 何晓峰, 杨慧林, 张志斌, 等. 小单孢菌 (*Micromonospora rosaria*) DSM803 全基因组测序及分析 [J]. 基因组学与应用生物学, 2016, 35(8): 2055-2063.
- [14] 段民孝. 基因组学研究概述 [J]. 北京农业科学, 2001(2): 6-10.
- [15] CHING Xinjie, TEOH C P, LEE D J H, et al. Genome of a thermophilic bacterium *Geobacillus* sp. TFV3 from Deception Island, Antarctica [J]. Advances in Polar Science, 2020, 31(2): 146-152.
- [16] 徐玲霞, 王筱兰, 杨慧林, 等. 雄烯二酮高产菌新金分枝杆菌 MN4 的全基因组测序及序列分析 [J]. 微生物学报, 2016, 56(8): 1358-1367.
- [17] ZERBINO D R, BIRNEY E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs [J]. Genome Research, 2008, 18(5): 821-829.
- [18] DELCHER A L, BRATKE K A, POWERS E C, et al. Identifying bacterial genes and endosymbiont DNA with Glimmer [J]. Bioinformatics, 2007, 23(6): 673-679.
- [19] BLIN K, WOLF T, CHEVRETTE M G, et al. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification [J]. Nucleic Acids Research, 2017, 45(1): 36-41.
- [20] RHEE J S, KIM R O, CHOI H G, et al. Molecular and biochemical modulation of heat shock protein 20 (Hsp20) gene by temperature stress and hydrogen peroxide (H_2O_2) in the monogonont rotifer, *Brachionus* sp. [J]. Comparative Biochemistry Physiology Part C: Toxicology & Pharmacology, 2011, 154(1): 19-27.
- [21] AVELLANEDA M J, FRANKE K B, SUNDERLIKOVA V,

- et al. Publisher correction: processive extrusion of polypeptide loops by a Hsp100 disaggregase [J]. *Nature*, 2020, 578(7796): 317-320.
- [22] LOPEZ M M, YUTANI K, MAKHATADZE G I. Interactions of the cold shock protein CspB from *Bacillus subtilis* with single-stranded DNA: importance of the T base content and position within the template [J]. *Journal of Biological Chemistry*, 2001, 276(18): 15511-15518.
- [23] PHADTARE S. Recent developments in bacterial cold-shock response [J]. *Current Issues in Molecular Biology*, 2004, 6(2): 125-136.
- [24] HAKI G D, RAKSHIT S K. Developments in industrially important thermostable enzymes: a review [J]. *Bioresource Technology*, 2003, 89(1): 17-34.
- [25] ANDUALEMA B, GESSESSE A. Microbial lipases and their industrial applications: review [J]. *Biotechnology*, 2012, 11(3): 100-118.
- [26] EUNHYE J, KIM J, LEE A, et al. Identification and characterization of a novel thermostable GDSL-type lipase from *Geobacillus thermocatenulatus* [J]. *Journal of Microbiology and Biotechnology*, 2021, 31(3): 483-491.
- [27] MAHAJAN R V, KUMAR V, RAJENDRAN V, et al. Purification and characterization of a novel and robust L-asparaginase having low-glutaminase activity from bacillus licheniformis; in vitro evaluation of anti-cancerous properties [J]. *PLOS ONE*, 2014, 9(6): e99037.
- [28] MOSTAFA Y S, ALAMRI S A, ALFAFI M Y, et al. L-glutaminase synthesis by marine *halomonas meridiana* isolated from the red sea and its efficiency against colorectal cancer cell lines [J]. *Molecules*, 2021, 26(7): 1963.
- [29] BARRY S M, CHALLIS G L. Recent advances in siderophore biosynthesis [J]. *Current Opinion in Chemical Biology*, 2009, 13(2): 205-215.
- [30] CHALLIS G L. A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases [J]. *ChemBioChem*, 2005, 6(4): 601-611.
- [31] GROSS R, ENGELBRECHT F, BRAUN V. Genetic and biochemical characterization of the aerobactin synthesis operon on pColV [J]. *Molecular & General Genetics*, 1984, 196(1): 74-80.
- [32] 黄婷婷, 林双君, 邓子新. 放线菌中铁载体生物合成机制研究进展 [J]. *微生物学通报*, 2011, 38(5): 765-773.
- [33] 徐水宝, 金嘉琳. 高毒力肺炎克雷伯菌的分子致病机制 [J]. *微生物与感染*, 2017, 12(5): 320-326.
- [34] WANG Mengmeng, MIAO Jiaxi, WANG Xuanqing, et al. Genomic and transcriptome analyses of a thermophilic bacterium *Geobacillus stearothermophilus* B5 isolated from compost reveal its enzymatic basis for lignocellulose degradation [J]. *Microorganisms*, 2020, 8(9): 1357.
- [35] 柳巨雄, 栾新红, 杨焕民, 等. 冷休克蛋白表达的分子机制 [J]. *中国兽医学报*, 2003, 23(6): 607-610.
- [36] 陈丽萍. cspB 基因对烟草的遗传转化 [D]. 淮北: 淮北师范大学, 2011: 43-46.

The Whole-Genome Sequencing and Sequence Analysis of *Geobacillus* sp. YHL

LU Ziyun, ZHA Shuanglong, LIU Jiangyin, MA Yangshuai, XIAO Liting, YANG Huilin*

(Key Laboratory of the Conservation and Sustainable Utilization for Subtropical Plant Resources of Jiangxi Province, College of Life Science, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

Abstract: The Illumina high-throughput sequencing technology is used to sequence the whole genome of microorganisms, Velet software is used for assembly, and then software Glimmer is used for gene prediction of bacterial strains, and the predicted proteins are compared with COG, KEGG and other databases to obtain the corresponding annotation information in this paper. Using a variety of drawing tools for annotation information on summary and analysis of the results, the COG, KEGG and other basic annotation information are obtained, and these information are analyzed by using the data mining. The results find that the strain has a variety of enzyme gene encoding, including glycoside hydrolase, glucoside enzyme new pullulan, xylanase, amylase, enzyme, branched chain, amylase and lipase. It is a thermophilic multienzyme encoding bacterium with important potential for industrial use. The focus is also on genes encoding heat stress proteins in the genome, which may ultimately provide a preliminary explanation of the mechanism of bacterial heat adaptation.

Key words: thermophilic microorganisms; *Geobacillus*; whole genome sequencing; genome analysis

(责任编辑: 刘显亮)