



## 论文

# 应用 BIGIS-4 测序系统完成 *Glacielola mesophila* sp. nov. 的全基因组测序和组装

袁丽娜<sup>①†</sup>, 任鲁风<sup>①†</sup>, 李运涛<sup>②</sup>, 韩伟静<sup>②</sup>, 俞勇<sup>③</sup>, 楚亚男<sup>①</sup>, 刘贵明<sup>①</sup>, 于丹<sup>①</sup>, 滕明静<sup>①</sup>, 王亮<sup>①</sup>, 王绪敏<sup>①</sup>, 周晓光<sup>②</sup>, 俞育德<sup>②\*</sup>, 于军<sup>①\*</sup>

① 中国科学院北京基因组研究所, 中国科学院基因组科学及信息重点实验室, 北京 100029;

② 中国科学院半导体研究所, 集成光电子学国家重点实验室, 北京 100083;

③ 中国极地研究中心, 国家海洋局极地科学重点实验室, 上海 200136

† 同等贡献

\* 联系人, E-mail: yudeyu@semi.ac.cn; junyu@big.ac.cn

收稿日期: 2011-05-09; 接受日期: 2011-06-10

中国科学院科研装备研制项目(批准号: YZ200823)、国家自然科学基金(批准号: 61007033, 30971610 和 40906097)和中国科学院北京基因组研究所自主立项经费资助

**摘要** BIGIS-4 是中国新一代基于焦磷酸测序技术的测序仪. 本实验利用 BIGIS-4 完成了对 *Glacielola mesophila* sp. nov. (*Gmn*) 的全基因组测序. *Gmn* 是一株从海洋无脊椎动物体内分离到的革兰氏阴性菌. BIGIS-4 测序得到 152043 个高质量的测序片段, 平均读长 406 bp. 测序片段由 BIGIS-4 系统后处理模块组装. 除单核苷酸同聚体引起的测序错误外, 测序结果中没有检测到其他低质量错误. *Gmn* 基因组全长 5144318 bp, 共注释得到 4303 个基因, 其中有大量的代谢基因, 与菌种在海洋表面非脊椎动物的生长环境相关. *Gmn* 的冷适应和信号转导相关基因因其对海洋低温环境的适应提供了依据.

**关键词**

*Glacielola mesophila* sp. nov.  
BIGIS-4 测序系统  
海洋环境的适应

*Glacielola* 属细菌作为优势微生物广泛存在于海冰、极区海水、海洋沉积物等海洋低温环境<sup>[1-8]</sup>. 该属的多数成员能产生氧化酶、过氧化氢酶、酯酶、淀粉酶、琼胶酶、 $\alpha/\beta$ 半乳糖苷酶、木聚糖等多种大分子及细菌多糖降解酶类, 同时还参与海水中原油的生物降解<sup>[9]</sup>. 这些生物学特征表明, *Glacielola* 属细菌在海洋低温生态系统的物质循环中起重要作用. 以 *Glacielola* 作为模式生物开展微生物嗜冷机制研究, 将为理解微生物在海洋低温生态系统中的生态功能

提供直接线索. *Glacielola mesophila* sp. nov. (*Gmn*) 由俄罗斯日本海 Troitsa 海湾 *Halocynthia aurantium* 标本的内液分离得到<sup>[1]</sup>. 实验室中, *Gmn* 在 3%(质量体积比) NaCl, pH 7.5, 25°C 条件下生长状态最好, 在低于 10°C 和高于 37°C 时不能存活. 当盐浓度降至 1%(质量体积比)NaCl 时, 菌株扩增缓慢, 高于 6% 时生长受到明显抑制(数据未显示). *Gmn* 不能水解尿素、明胶或酪蛋白, 能利用碳水化合物微弱产酸, 但又不能利用 L-阿拉伯糖、柠檬酸、苯乙酸、己二酸、L-苯丙

氨酸、L-组氨酸、L-鸟氨酸、L-苏氨酸和腐胺。Guo 等人<sup>[10]</sup>在 *G. mesophila* KMM 241 中克隆表达了一个冷适应和耐盐酶——木聚糖酶。木聚糖酶能有效地把低聚木糖和成木二糖转化成木二糖和木三糖而不产生木糖, 并且在高钠离子浓度和 4℃ 低温下仍保持高效的活力和催化活性。对 *Glaciicola* 全基因组序列的分析将为海洋生物对其独特生境的适应研究提供更好的基础。

BIGIS-4 是一个高度集成的可扩展型测序仪, 基于焦磷酸测序原理, dNTP 顺序进入反应仓, 在 DNA 聚合酶、ATP 硫酸化酶、荧光素酶和腺苷三磷酸双磷酸酶共同作用下实现序列读取<sup>[11]</sup>。BIGIS-4 由光学模块、液路模块、计算机模块、反应模块和一个数据后处理模块构成。其中, 反应模块分割成 4 个独立的反应仓, 共用一套液路系统。根据不同的测序样本量和通量需求, 1~4 个反应模块可独立或随机同时启动。测序芯片载有独立的微反应池, 一次测序运行可以得到 560 Mb 数据量。与市场化成熟的二代测序仪相比, 兼具低成本和长读长的优势。

本项目应用 BIGIS-4 测序系统对中温菌 *Gmn* 全基因组测序分析, 为微生物在海洋低温生态系统中的生态功能和微生物冷适应进化过程等研究积累理论和数据基础, 同时可以检测自主研发的 BIGIS-4 系统在实际应用中的工作能力。

## 1 材料与方法

### 1.1 菌株培养

*Gmn* 菌株利用人工海水培养基(MgCl<sub>2</sub> 5 g/L, MgSO<sub>4</sub> 2 g/L, CaCl<sub>2</sub> 0.5 g/L, KCl 1 g/L, 蛋白胨 5 g/L, 酵母提取物 0.1 g/L, 3%(质量体积比)NaCl 补足, 用 1 N KOH 调节至 pH 7.5)25℃ 下培养 2 天。

### 1.2 基因组测序

DNA 提取纯化后, 根据 454 文库准备和 emPCR 扩增方案进行上机前文库构建。固定了已扩增 DNA 文库的微球通过离心载入 BIGIS-4 测序芯片, 同时载入的还有载酶磁珠(包括 ATP 硫酸化酶和荧光素酶)和 DNA 聚合酶。准备测序芯片的同时, 启动时长 1 h 的预冲洗流程。预冲洗结束后, 用 BIGIS-4 测序试剂盒(包括 dATP, dTTP, dCTP, dGTP, 底物, ATP, 腺苷三磷酸双磷酸酶和测序缓冲液)替换预冲洗试剂盒, 反应仓中预冲洗用的旧测序芯片用新准备好的测序芯片

替换。测序芯片背面和 CCD 光纤面板相对, 形成化学荧光信号获取的暗环境。测序得到的原始数据传送到 BIGIS-4 后处理模块, 进行碱基识别和数据处理。

### 1.3 序列组装和分析

测序序列由 BIGIS-4 后处理软件进行组装。组装片段之间 >2 kb 的空缺由鸟枪法填补, <2 kb 的由 PCR 测序直接填补。已组装好的大片段和空缺处序列通过 Consed 程序包进行序列组装<sup>[12]</sup>, 编码序列通过 Glimmer3 程序进行预测<sup>[13]</sup>, 运用 BLASTP 将所有预测得到的蛋白序列与 NCBI 非冗余蛋白数据库进行比对, E 值阈值设为  $1 \times 10^{-10}$ 。通过 InterProScan 程序包进行蛋白分析<sup>[14]</sup>, 根据 COG 数据库对蛋白进行分类。基因组中的 tRNA 应用 tRNAscan-SE 工具搜索<sup>[15]</sup>, rRNA 经由 RNAmmer1.2 服务器(<http://www.cbs.dtu.dk/services/RNAmmer/>)识别。基因组中的插入序列采用 ISFinder 预测<sup>[16]</sup>。比较基因组分析借助 MUMmer 完成<sup>[17]</sup>。

## 2 实验结果

### 2.1 基因组拼接

BIGIS-4 系统单模块测序运行共产生 152043 条高质量读出序列, 数据产量约 64 Mb。图 1 显示了读出序列长度的分布, 其平均值和中值分别为 406 和 442 bp。测序准确率 99.5%, 单核苷酸同聚体错误是主要的错误来源。读出序列组装成 157 个大的重叠群, N50 大小为 61174 bp, 重叠群之间的空缺有 9 个大于 2 kb。Sanger 测序补洞平均读长 485 bp。测序基因组覆盖度为 12x, 覆盖了全基因组的 99.4%(图 2)。高覆盖区(>60x)包括了基因组中除 rRNA 操作子外主要的转座酶和多拷贝的假设蛋白, 其中有 11 个可移动遗传因子相关基因和插入序列(insertion sequence, IS)IS4 基因家族中的 6 个基因。覆盖图中缺失位置的基因主要包括 ABC 转运相关蛋白、DNA 结合的转录激活因子、响应调控和类多肽酶等功能基因。

### 2.2 基因组特征

*Gmn* 环形基因组长度为 5144318 bp(图 3)。包括 4528 个预测蛋白编码基因、59 个 tRNA 基因和 15 个 rRNA 基因。其中, 4303 个注释基因包括 3220 个有已知功能的同源基因、1043 个编码假设蛋白基因和 40 个可移动的遗传因子(包括 8 个整合酶、26 个转座酶

和6个噬菌体相关因子)。基因编码密度为86%，基因间区域平均长度166 bp，GC含量44.6%。

*Gmn* 的注释基因中包括一系列糖类代谢基因，如葡萄糖、山梨醛、 $\alpha$ -L-岩藻糖苷、半乳糖苷、海藻糖、甘露糖、N-乙酰半乳糖胺等。本研究同时预测到1个编码 $\beta$ -1,4-木聚糖酶的基因，作为可抵抗渗透压的基因与细菌的冷适应性有关。ABC类糖转运相关基因和TRAP类C4-二羧酸转运相关基因各自相邻，可能都以操纵子实现表达调控。基因组中还包括大

量的双组分信号转导蛋白、转录调控子、热休克响应蛋白、鞭毛/纤毛装配蛋白、伴侣蛋白、分泌蛋白等，从而使菌株有效地适应环境的多样变化。

### 2.3 比较基因组分析

在基因组水平上，比较了 *Gmn* 与 *Pseudoalteromonas atlantica* T6c 和 *Alteromonas macleodii* ‘Deep ecotype’。结果发现，根据序列相似性，*Gmn* 在进化上与 *P. atlantica* T6c 更为接近(图4)，这与对16S rRNA

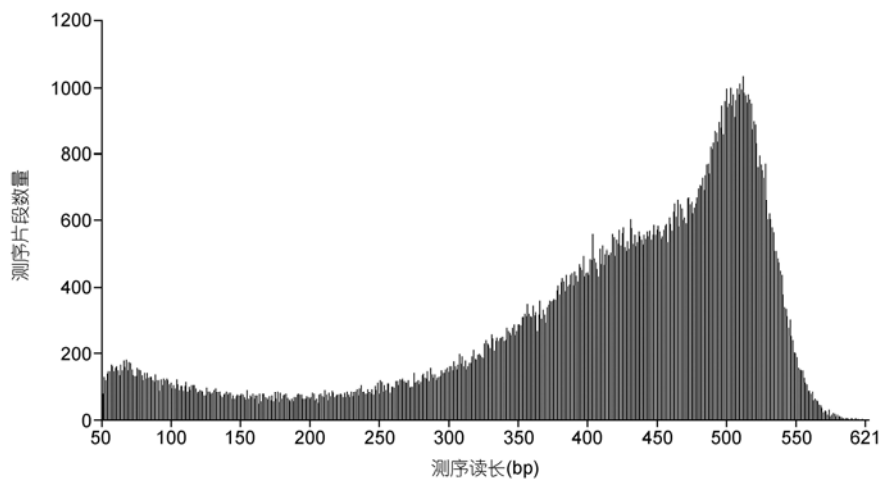


图1 BIGIS-4 单反应模块测序读长分布

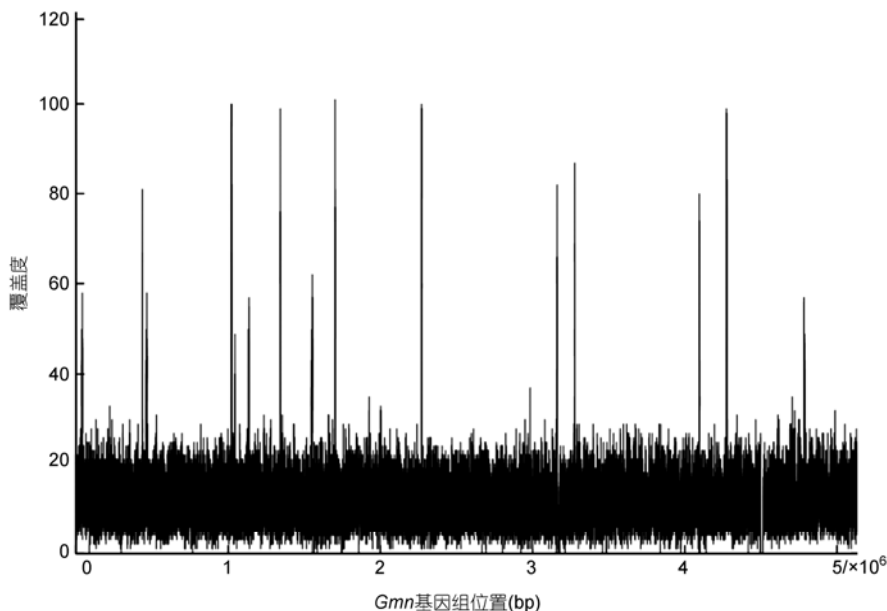


图2 BIGIS-4 读长序列基因组覆盖图<sup>[18]</sup>

在 *Gmn* 基因组中，发现了10个IS元件的完整拷贝，可以分为IS4和IS110两个家族，其中IS4家族(7个拷贝)占主导地位

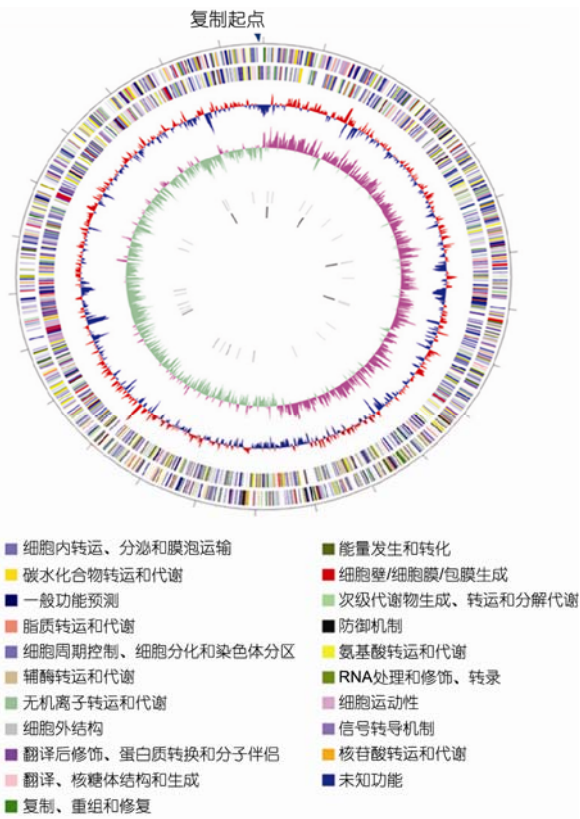


图3 *Gmn* 基因组圈图

自外向内, 环 1: 基因组位置标尺(kb); 环 2 和 3: CDS 描述(颜色标识见图例)。环 4: GC 含量(10 kb 窗口, 1 kb 步长), 红色标识 GC>44.6%, 蓝色标识 GC≤44.6%; 环 5: GC 偏移(G-C/G+C, 10 kb 窗口, 1 kb 步长), 玫红色标识>0, 绿色标识<0; 环 6: tRNA; 环 7: rRNA。基因组复制起始位点通过 GC 偏移预测, 复制终点和复制起点处有明显的 GC 偏移过渡

的分析结果一致(图 5)。

*Gmn* 和 *P. atlantica* T6c 的基因组分布具有很强的共线性, 非共线性序列仅有零散星点分布。*Gmn* 基因组的 4303 个注释基因中有 3968 个与 *P. atlantica* T6c 基因序列同源匹配度最高, 而与 *A. macleodii* ‘Deep ecotype’相匹配的只有 13 个。*Gmn* 与 *P. atlantica* T6c 不相匹配的基因中, 有 17 个是能行使基因水平迁移功能的整合酶或转座酶, 62 个是假设蛋白且其中 8 个在基因组中有多个拷贝。相比之下, *A. macleodii* ‘Deep ecotype’和 *Gmn* 的匹配基因多为行使不同功能的保守基因, *Gmn* 中与 *A. macleodii* ‘Deep ecotype’不匹配的基因主要与代谢有关, 而对于 *A. macleodii* ‘Deep ecotype’, 大部分独有基因(69.8%)是假设蛋白, 其余独有基因也未显示出任何偏向性分

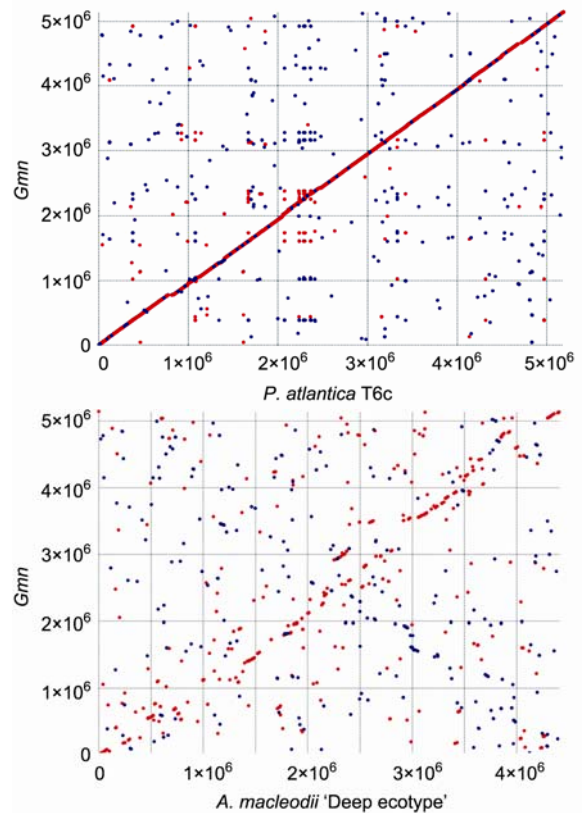


图4 *Gmn* 与 *P. atlantica* T6c 和 *Gmn* 与 *A. macleodii* ‘Deep ecotype’的全基因组比对

利用 mummerplot 作图, reference 序列为 X 轴, query 序列为 Y 轴。线和点标示了两个序列相匹配的位置, 正向匹配标识为红色, 反向匹配标识为蓝色。当两个序列匹配无完全一致时将出现一条贯穿左下角到右上角的红色线

布。此外, *A. macleodii* ‘Deep ecotype’同时包含了很多转座元件和 IS 元件<sup>[19]</sup>。

## 2.4 环境适应性分析

在 *Gmn* 基因组中, 发现一个编码冷适应活性和耐盐性的β-木聚糖酶基因, 可以将β-1,4-木聚糖降解成木糖<sup>[20]</sup>。木聚糖酶基因在 *P. atlantica* T6c 中也有发现, 但在 *A. macleodii* ‘Deep ecotype’中未找到编码序列。此外, 通过氨基酸序列同源分析发现, *Gmn* 与 *P. atlantica* T6c 还共有 GroE 伴侣蛋白(一致性为 100%), 而与 *A. macleodii* ‘Deep ecotype’一致性为 89%。GroE 伴侣蛋白由 2 个协作蛋白 GroES 和 GroEL 构成, 属于一类在进化中高度保守的伴侣蛋白。GroE 参与蛋白折叠, 帮助微生物应对包括热激、NaCl 和 SDS 等各种环境压力<sup>[21]</sup>。当暴露于压力环境中, GroE 就会高表

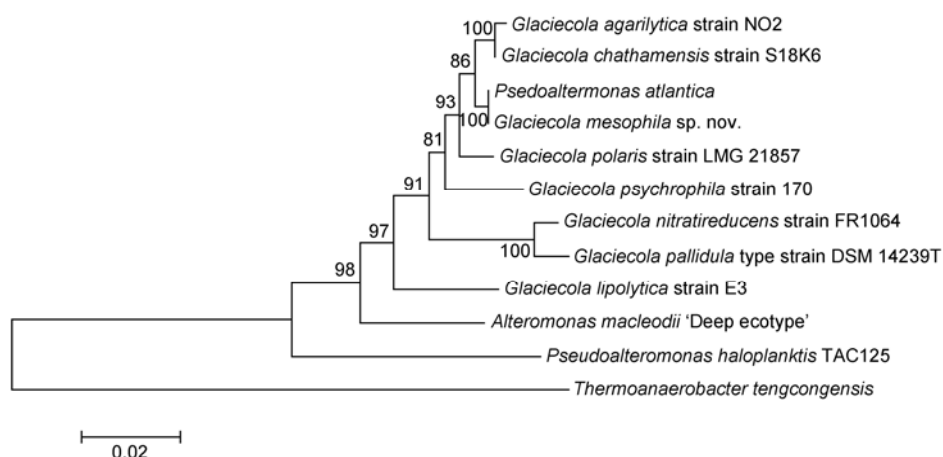


图 5 *Glaciecola* 属细菌和相关菌种的进化树构建

*Glaciecola* 属及相关种属的 16S rDNA 序列来源于 GenBank 数据库, 多序列通过 ClustalW 程序比对, Neighbor-joining 算法构建。进化树节点标记数值为步长值。 *Thermoanaerobacter tengcongens* 为外类群对照

达<sup>[22]</sup>, 尤其在深海微生物基因组中<sup>[23]</sup>。这两个基因在 *G. mesophila* sp. nov. 基因组上呈 *groES-groEL* 顺序排列, 中间有 35 bp 的间隔。GroEL 在嗜冷、嗜温、嗜热以及各类细菌中均保持高度相似性(覆盖率>95%, 一致性大于 80%)。

### 3 讨论

本文以中温菌 *Gmn* 为研究对象, 应用 BIGIS-4 测序系统对其进行全基因组测序分析。 *Gmn* 是从海洋无脊椎动物体内分离到的一株新型海洋细菌。基于表型、生化特性和进化分析结果, *Gmn* 是 *Glaciecola* 属的一个新种。 *A. macleodii* 'Deep ecotype' 是从南亚得里亚海底盆地 1000 米深处分离得到的一个常见深海分离株<sup>[24]</sup>, 与 *Glaciecola* 亲缘关系很近<sup>[1]</sup>。 *Pseudoalteromonas* 属是最早由 Gauthier 等人<sup>[25]</sup>从 *Alteromonas* 属分离建立的新属, *P. atlantica* T6c 菌株从加利福尼亚圣地亚哥湾的生物膜分离得到。对 *Gmn* 与 *P. atlantica* T6c 和 *A. macleodii* 'Deep ecotype' 的比较基因组分析显示, *P. atlantica* T6c 是与 *Gmn* 亲缘关系最近的菌株。基于 16S rRNA 序列进化分析同样证实, *P. atlantica* T6c 和 *Gmn* 进化距离最近。多于 90% 的 *Gmn* 基因与 *P. atlantica* T6c 中的同源基因具有较高的相似性。这两株细菌可能代表了海洋无脊椎动物的一个重要的微生物区系。相对而言, *A. macleodii* 'Deep ecotype' 则是深海生态型的代表。与

预期一致, *A. macleodii* 'Deep ecotype' 与 *Gmn* 同源基因的数量和相似度都很低, *Gmn* 有更多的代谢基因, 能降解更多的糖和氨基酸, 可能比深海细菌 *A. macleodii* 'Deep ecotype' 具有更广泛的营养来源。

在 *Gmn* 和 *P. atlantica* T6c 基因组中找到了编码在低温和高盐环境下仍然具有较高活性的内切- $\beta$ -1,4-木聚糖酶的基因。Collins 等人<sup>[20]</sup>发现, 木聚糖酶嗜冷型的最适温度低于其嗜温型, 进一步为木聚糖酶在低温环境中的适应性提供了依据。此外, 基于生理生化实验结果, *A. macleodii* 个别菌株能在 37°C 甚至 40°C 高温存活, 耐受 10% 的高盐, 而在同样条件下, *Gmn* 和 *P. atlantica* T6c 都不能存活<sup>[1]</sup>。在 *A. macleodii* 中没有找到编码木聚糖酶的基因, 这在一定程度上反映了海洋不同生境微生物对于压力的不同应对机制。另外, *Gmn* 中丰富的双组分信号转导蛋白和转录调控子, 相关的细胞过程和信号系统协同作用也能帮助细菌有效地应对胞内和胞外的压力刺激。

*Glaciecola* 属囊括了能在海洋多样的温度、盐度和营养环境中生长繁殖的细菌种类, 包括海冰、海水、海洋沉积物和无脊椎动物等, 涵盖嗜冷菌、耐冷菌和中温菌种。相应地, 形成多样的适应机制。与不同属微生物相比, 同属微生物具有更相似的遗传背景, 更有利于研究由环境驱动的微生物进化过程。期待获得更多的 *Glaciecola* 基因组信息, 从而能更精确地了解微生物冷适应进化过程, 进而建立其冷适应进化模型。

## 参考文献

- 1 Romanenko L A, Zhukova N V, Rohde M, et al. *Glaciecola mesophila* sp. nov., a novel marine agar-digesting bacterium. *Int J Syst Evol Microbiol*, 2003, 53: 647–651
- 2 Zhang D C, Yu Y, Chen B, et al. *Glaciecola psychrophila* sp. nov., a novel psychrophilic bacterium isolated from the Arctic. *Int J Syst Evol Microbiol*, 2006, 56: 2867–2869
- 3 Van Trappen S, Tan T L, Yang J, et al. *Glaciecola polaris* sp. nov., a novel budding and prosthecate bacterium from the Arctic Ocean, and emended description of the genus *Glaciecola*. *Int J Syst Evol Microbiol*, 2004, 54: 1765–1771
- 4 Baik K S, Park Y D, Seong C N, et al. *Glaciecola nitratireducens* sp. nov., isolated from seawater. *Int J Syst Evol Microbiol*, 2006, 56: 2185–2188
- 5 Chen L P, Xu H Y, Fu S Z, et al. *Glaciecola lipolytica* sp. nov., isolated from seawater near Tianjin city, China. *Int J Syst Evol Microbiol*, 2009, 59: 73–76
- 6 Matsuyama H, Hirabayashi T, Kasahara H, et al. *Glaciecola chathamensis* sp. nov., a novel marine polysaccharide-producing bacterium. *Int J Syst Evol Microbiol*, 2006, 56: 2883–2886
- 7 Zhang Y J, Zhang X Y, Mi Z H, et al. *Glaciecola arctica* sp. nov., isolated from Arctic marine sediment. *Int J Syst Evol Microbiol*, 2010
- 8 Yong J J, Park S J, Kim H J, et al. *Glaciecola agarilytica* sp. nov., an agar-digesting marine bacterium from the East Sea, Korea. *Int J Syst Evol Microbiol*, 2007, 57: 951–953
- 9 Prabakaran S R, Manorama R, Delille D, et al. Predominance of *Roseobacter*, *Sulfitobacter*, *Glaciecola* and *Psychrobacter* in seawater collected off Ushuaia, Argentina, Sub-Antarctica. *FEMS Microbiol Ecol*, 2007, 59: 342–355
- 10 Guo B, Chen X L, Sun C Y, et al. Gene cloning, expression and characterization of a new cold-active and salt-tolerant endo-beta-1,4-xylanase from marine *Glaciecola mesophila* KMM 241. *Appl Microbiol Biotechnol*, 2009, 84: 1107–1115
- 11 Ronaghi M, Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate. *Science*, 1998, 281: 363–365
- 12 Gordon D. Viewing and editing assembled sequences using Consed. *Curr Protoc Bioinformatics*, 2003, Chapter 11: Unit11.2
- 13 Delcher A L, Bratke K A, Powers E C, et al. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*, 2007, 23: 673–679
- 14 Zdobnov E M, Apweiler R. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics(Oxford, England)*, 2001, 17: 847–848
- 15 Schattner P, Brooks A N, Lowe T M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res*, 2005, 33: W686–W689
- 16 Siguiet P, Perochon J, Lestrade L, et al. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res*, 2006, 34: D32–D36
- 17 Kurtz S, Phillippy A, Delcher A L, et al. Versatile and open software for comparing large genomes. *Genome Biol*, 2004, 5: R12
- 18 Bott O J, Bergmann J, Hoffmann I, et al. Analysis and specification of telemedical systems using modelling and simulation: the MOSAIK-M approach. *Stud Health Technol Inform*, 2005, 116: 503–508
- 19 Ivars-Martinez E, Martin-Cuadrado A B, D'Auria G, et al. Comparative genomics of two ecotypes of the marine planktonic copiotroph *Alteromonas macleodii* suggests alternative lifestyles associated with different kinds of particulate organic matter. *ISME J*, 2008, 2: 1194–1212
- 20 Collins T, Gerday C, Feller G. Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiol Rev*, 2005, 29: 3–23
- 21 Tosco A, Birolo L, Madonna S, et al. GroEL from the psychrophilic bacterium *Pseudoalteromonas haloplanktis* TAC 125: molecular characterization and gene cloning. *Extremophiles*, 2003, 7: 17–28
- 22 Yamauchi S, Okuyama H, Morita E H, et al. Gene structure and transcriptional regulation specific to the groESL operon from the psychrophilic bacterium *Colwellia maris*. *Arch Microbiol*, 2003, 180: 272–278
- 23 Xu K, Ma B G. Comparative analysis of predicted gene expression among deep-sea genomes. *Gene*, 2007, 397: 136–142
- 24 Lopez-Lopez A, Bartual S G, Stal L, et al. Genetic analysis of housekeeping genes reveals a deep-sea ecotype of *Alteromonas macleodii* in the Mediterranean Sea. *Environ Microbiol*, 2005, 7: 649–659
- 25 Gauthier G, Gauthier M, Christen R. Phylogenetic analysis of the genera *Alteromonas*, *Shewanella*, and *Moritella* using genes coding for small-subunit rRNA sequences and division of the genus *Alteromonas* into two genera, *Alteromonas* (emended) and *Pseudoalteromonas* gen. nov., and proposal of twelve new species combinations. *Int J Syst Bacteriol*, 1995, 45: 755–761