

向 HelmCoP 数据库直系同源组添加古菌序列

杨一子¹, 罗大民^{1, 2}

(1. 厦门大学生命科学学院, 福建 厦门 361005)

(2. 厦门大学细胞应激生物学国家重点实验室, 福建 厦门 361005)

[摘要] 含有古菌序列的直系同源组可以在很大程度上便利各种基于 HelmCoP 数据的研究,但是目前 HelmCoP 直系同源组不含有任何古菌序列. 为了向 HelmCoP 直系同源组添加古菌序列,以 1 个 OrthoMCL-DB 网站工具的分配方法为依据,我们将 29 种古菌物种模式菌株的蛋白质组数据,通过 BLAST 比对和 OrthoMCL 分析,逐步添加至各个 HelmCoP 直系同源组中. 结果显示,21 326 条古菌序列被成功地分配给了 1 954 个 HelmCoP 直系同源组中,而其他 2 821 条古菌序列被归类至 NO_GROUP 类别. 这些分配结果得到了良好 BLAST 匹配数据的支持. 一些自由生活物种包括 *Caenorhabditis japonica*、大豆和拟南芥在研究中也受到了关注,因为这些物种相当多的基因在古菌物种内有同源基因.

[关键词] 古菌, OrthoMCL, HelmCoP, BLAST

[中图分类号] TP311.13 [文献标志码] A [文章编号] 1001-4616(2014)03-0089-06

Assigning Archaeal Sequences to HelmCoP Groups

Yang Yizi¹, Luo Damin^{1, 2}

(1. School of Life Sciences, Xiamen University, Xiamen 361005, China)

(2. State Key Laboratory of Cellular Stress Biology, Xiamen University, Xiamen 361005, China)

Abstract: Orthologous groups containing archaeal sequences could facilitate many researches based on HelmCoP data, but unfortunately there were no archaeal sequences in HelmCoP groups. To assign archaeal sequences to HelmCoP groups, we assigned sequences in proteomes of 29 type strains of multiple archaeal species to HelmCoP groups through both BLAST search and OrthoMCL analysis, according to the protocol for assigning sequences by a OrthoMCL-DB web tool. The results showed that 21 326 archaeal sequences were successfully assigned to 1 954 HelmCoP groups and other 2 821 sequences were assigned to NO_GROUP. These assignments gained support from quite good BLAST hits. In addition, some free-living species including *Caenorhabditis japonica*, soybean, and thale cress were also highlighted in this study, because many genes of these species had homologs in archaeal species.

Key words: Archaea, OrthoMCL, HelmCoP, BLAST

古菌作为原核生物的一大类,既有与真细菌相似的地方,又有与真核生物相似的一些其他特征. 在很多情况下古菌序列适合在构建真核生物基因的系统发生树时选作为外类群序列^[1-3]. 此外它们还可用于检测一些古菌和真核生物之间的水平基因转移事件^[4]. 多种真核生物物种的序列在 HelmCoP 数据库^[5]的直系同源组中找到,但遗憾的是其中并没有古菌物种的序列. 甚者,目前这个数据库还没有任何工具可以给它已计算好的直系同源组添加新的序列. 考虑到这些 HelmCoP 数据在线虫及其他生物的研究中有广泛的用途,缺乏古菌序列这个问题肯定会阻碍这些数据在各种研究中的更深入应用.

最新版本的 OrthoMCL-DB 数据库^[6, 7](version 5)是一个与 HelmCoP 相类似的数据库. 该数据库收集了 150 个物种包括古菌的直系同源组. 不仅如此,该数据库还提供了一个以 BLAST 比对^[8]和 OrthoMCL 分析^[9]为基础的网络工具,可以方便快捷地向该数据库的直系同源组添加新的序列. 有了这个工具,研究人员可以很容易地实现向 OrthoMCL-DB 直系同源组添加任何物种序列的目标. BLAST 比对程序和

收稿日期: 2014-03-16.

基金项目: 国家自然科学基金(81171595)、福建省自然科学基金(2010J01229).

通讯联系人: 罗大民 教授, 博导, 研究方向: 寄生虫学. E-mail: dmluo@xmu.edu.cn

OrthoMCL 软件都可以从互联网下载,而那些用于构建 HelmCoP 数据库的数据以及该数据库直系同源组的分组信息,也可以通过与 HelmCoP 数据库开发者^[5]联系而获得。我们希望通过应用 OrthoMCL 数据库网络工具的方法来向 HelmCoP 直系同源组添加古菌序列。本研究结果将利于今后基于 HelmCoP 直系同源组的各类研究。

1 材料和方法

我们从 Ensembl Bacteria 数据库^[10]中下载了 28 种古菌物种模式菌株的蛋白质组的全部数据。这些物种和菌株包括敏捷气热菌模式菌株(*Aeropyrum pernix* K1)、发光古球菌模式菌株(*Archaeoglobus fulgidus* DSM 4304)、排泄矿水嗜酸亚铁微菌模式菌株(*Ferroplasma acidarmanus fer1*)、死海盐盒菌模式菌株(*Haloarcula marismortui* ATCC 43049)、嗜盐杆菌模式菌株(*Halobacterium* sp. NRC-1)、沃氏嗜盐富饶菌模式菌株(*Haloferax volcanii* DS2 (ASM2568v1))、詹氏甲烷球菌模式菌株(*Methanocaldococcus jannaschii* DSM 2661)、布氏拟甲烷球菌模式菌株(*Methanococcoides burtonii* DSM 6242)、海沼甲烷球菌模式菌株(*Methanococcus marisaludis* C5)、万氏甲烷球菌模式菌株(*Methanococcus vannielii* SB)、甲烷嗜高热菌模式菌株(*Methanopyrus kandleri* AV19)、嗜酸甲烷八叠球菌模式菌株(*Methanosarcina acetivorans* C2A)、巴氏甲烷八叠球菌模式菌株(*Methanosarcina barkeri* str. Fusaro)、梅氏产甲烷八叠球菌模式菌株(*Methanosarcina mazei* Go1)、热自养嗜热甲烷杆菌模式菌株(*Methanothermobacter thermautotrophicus* str. Delta H)、骑行纳古菌模式菌株(*Nanoarchaeum equitans* Kin4-M)、干燥嗜酸球菌模式菌株(*Picrophilus torridus* DSM 9790)、耐超高温热棒菌模式菌株(*Pyrobaculum aerophilum* str. IM2)、深海火球菌模式菌株(*Pyrococcus abyssi* GE5)、激烈火球菌模式菌株(*Pyrococcus furiosus* DSM 3638)、掘越氏火球菌模式菌株(*Pyrococcus horikoshii* OT3)、嗜酸热硫化叶菌模式菌株(*Sulfolobus acidocaldarius* DSM 639)、硫磺矿硫化叶菌模式菌株(*Sulfolobus solfataricus* 98/2)、床田硫化叶菌模式菌株(*Sulfolobus tokodaii* str. 7)、伽马耐受超嗜热菌模式菌株(*Thermococcus gammatolerans* EJ3)、十岛村超嗜热菌模式菌株(*Thermococcus kodakarensis* KOD1)、嗜酸热原体模式菌株(*Thermoplasma acidophilum* DSM 1728)、火山热原体模式菌株(*Thermoplasma volcanium* GSS1)。此外,寒冷产甲烷菌模式菌株(*Methanogenium frigidum* Ace-2)^[11]也从 J. Craig Venter Institute 网站被下载下来。这些古菌和菌株的选择是依据以往古菌演化研究经常选择的对象来确定的^[12,13]。

在 HelmCoP 数据库开发者 John Martin 的帮助下,该数据库 18 种物种的蛋白质组也从网络下载下来。随后,这些数据文件通过 orthomclAdjustFasta 程序转换成 OrthoMCL-complaint 格式。它是 2.0 版本 OrthoMCL 软件包含的一个 Perl 程序^[6]。于此同时,所有 29 种古菌物种模式菌株的蛋白质组也被转换成 OrthoMCL-complaint 格式。接着,经过格式转换后的数据需要通过 orthomclFilterFasta 程序分析来移除低质量的数据,而这个分析会得到两个文件: HelmCoP_goodProtein.fasta 和 Archaea_goodProtein.fasta。然后,我们以 HelmCoP_goodProtein.fasta 文件数据为基础,通过 makeblastdb 程序^[14]创建了一个数据库 HelmCoP_goodProtein。这是 BLAST+软件包含的一个程序。我们通过 Archaea_goodProtein.fasta 和 blastp 程序,提交了 HelmCoP_goodProtein 数据库内批量搜索的 BLAST 比对任务。BLAST 分析的 E 值设定在 1×10^{-5} ,而输出结果的格式设定为表格式。

我们通过使用 orthomclBlastParser 程序和 OrthoMCL-complaint 格式的蛋白质组文件,来对 BLAST 分析的输出结果做进一步的解析。在这个分析后,一些指标,例如 percent_match^[6],在分析结果中被计算出来。接下来,我们通过使用 2 个 Perl 程序,即 filter1.pl 和 filter2.pl,来移除那些 percent_match 指标低于 50% 的 BLAST 匹配结果,并提取其余匹配结果中的最佳匹配。我们也通过 John Martin 获取了 HelmCoP 直系同源组的分组信息,该信息保存在 orthomcl_groups.txt 文件中。我们通过使用一个名为 Assigner.pl 程序,将多个古菌蛋白依照其最佳匹配序列所在的直系同源组信息,逐一分配到各个 HelmCoP 直系同源组中。如果某个古菌蛋白的最佳匹配序列并不包含在任何一个 HelmCoP 组中,那么该古菌蛋白则会被分配到 NO_GROUP 类别中。全部古菌蛋白的分配信息写在了文件 Archaea_HelmCoP.xlsx 中。全部 Perl 程序和分析结果都可以通过和我们联系来索取。

2 结果

一共有 24 147 条古菌序列在 BLAST 分析和后继分析中有高于阈值的最佳匹配。其中 21 326 条序列

被分配给了 1 954 个 HelmCoP 直系同源组, 而其余 2 821 条序列被归在了 NO_GROUP 类别. 超过 270 条的古菌序列被分配给了一个直系同源组 ortho17taxa2398. 1 954 个直系同源组中的 154 个组接受了 30 条及更多数量的古菌序列. 相对的 468 个直系同源组只接受了 1 条古菌序列. 平均起来, 每个组能接受到大约 11 条古菌序列. 这些组的组名从 ortho17taxa1003 延伸至 ortho17taxa39470. 前 20 个接受到最多数目古菌序列的 HelmCoP 直系同源组罗列在表 1 中.

表 1 前 20 个接受到最多数目古菌序列的 HelmCoP 直系同源组
Table 1 Top 20 HelmCoP groups receiving largest number of archaeal sequences

名次	组名	古菌序列数量	名次	组名	古菌序列数量
1	ortho17taxa2398	278	11	ortho17taxa1399	103
2	ortho17taxa7962	222	12	ortho17taxa1265	100
3	ortho17taxa12195	144	13	ortho17taxa1311	98
4	ortho17taxa1234	138	14	ortho17taxa1072	95
5	ortho17taxa1026	136	15	ortho17taxa1093	93
6	ortho17taxa1206	133	16	ortho17taxa24900	90
7	ortho17taxa2049	127	17	ortho17taxa1747	88
8	ortho17taxa1246	117	17	ortho17taxa1136	88
9	ortho17taxa1044	115	17	ortho17taxa1046	88
10	ortho17taxa1237	105	20	ortho17taxa4345	83

与此同时 2 821 条被归于 NO_GROUP 类别的古菌序列, 与 564 条 HelmCoP 单一序列有最佳匹配. 其中的 98 条古菌序列共同与一条自由生活线虫 *Caenorhabditis japonica* 的序列, 即 CJA23009 有最佳匹配. 此外, 564 条 HelmCoP 单一序列中的 67 条与 10 条以上的古菌序列有最佳匹配, 而这些古菌序列都被归于 NO_GROUP 类别. 前 20 条与最多数目古菌序列有最佳匹配的 HelmCoP 单一序列罗列在表 2 中.

表 2 前 20 个与最多数目古菌序列有最佳匹配的 HelmCoP 单一序列
Table 2 Top 20 single HelmCoP sequences best matching with largest number of archaeal sequences

名次	序列名	古菌序列数量	名次	序列名	古菌序列数量
1	CJA23009	98	10	YGR210C	31
2	CJA26276	59	12	CJA21595	28
3	Glyma02g44150.1	46	12	CJA19636	28
4	At4g01900.1	43	14	CJA22750	27
5	CJA24156	42	14	YKL211C	27
5	CJA24090	36	14	CJA28534	27
7	CJA20751	35	17	CJA26997	26
7	CJA24369	35	18	YDR538W	24
9	CJA28296	32	19	CJA25043	23
10	CJA24309	31	19	CJA20098	23

被分配古菌序列的物种和菌株信息也被记录下来. 全部 29 种古菌物种模式菌株的蛋白质组中都有序列被分配到 HelmCoP 直系同源组或 NO_GROUP 类别中. 其中, 来自沃氏嗜盐富饶菌模式菌株的序列在所有 29 种模式菌株中数量最多, 而来自寒冷产甲烷菌模式菌株的序列数量最少. 平均起来, 每个模式菌株蛋白质组中有 735 条序列被分配给 HelmCoP 直系同源组, 而另有 97 条序列被归到了 NO_GROUP 类别中. 考虑到整体分配情况, 干燥嗜酸球菌模式菌株蛋白质序列的分配率, 即每种模式菌株被分配蛋白质序列数量占 Archaea_goodProtein.fasta 文件中该菌株全部蛋白质序列数量的百分比率, 在全部 29 种古菌物种模式菌株中是最高的. 其他一些模式菌株, 比如嗜酸热原体模式菌株、火山热原体模式菌株以及深海火球菌模式菌株等等, 也有大于 40% 的序列分配率. 相反的, 寒冷产甲烷菌模式菌株的蛋白质序列的分配率, 明显低于其他模式菌株分配率, 仅有 1.72%. 全部 29 种古菌模式菌株的蛋白质序列数量对比情况显示在图 1 中. 全部类群代码及对应名称罗列在表 3 中.

此外, 我们把分配到 HelmCoP 直系同源组古菌序列数量与归于 NO_GROUP 门类古菌序列数量作比较. 在全部 29 种模式菌株中, 分配到直系同源组的古菌序列数量都远多于归于 NO_GROUP 门类古菌序列的数量. 相对来说, 来自骑行纳古菌模式菌株的蛋白质序列有最大可能性被分配到 HelmCoP 直系同源组中, 而来自嗜酸热原体模式菌株或火山热原体模式菌株的古菌序列也很可能被分配到这些直系同源组中.

与分配序列数量多少类似,对于来自寒冷产甲烷菌模式菌株的蛋白质序列来说,这一可能性仍然是所有模式菌株中最低的。

表 3 类群代码和对应的物种谱系名称

Table 3 List of taxon codes and their corresponding names

类群代码	对应名称	类群代码	对应名称
Aper	Aeropyrum pernix K1	Atha	Arabidopsis thaliana
Aful	Archaeoglobus fulgidus DSM 4304	BMA1	Brugia malayi
Faci	Ferroplasma acidarmanus fer1	Cbre	Caenorhabditis brenneri
Hmar	Haloarcula marismortui ATCC 43049	Cbri	Caenorhabditis briggsae
H_sp	Halobacterium sp. NRC-1	Cjap	Caenorhabditis japonica
Hvol	Haloferax volcanii DS2(ASM2568v1)	Crem	Caenorhabditis remanei
Mjan	Methanocaldococcus jannaschii DSM 2661	Dmel	Drosophila melanogaster
Mbur	Methanococcoides burtonii DSM 6242	Gmax	Glycine max
Mmar	Methanococcus maripaludis C5	Hsap	Homo sapiens
Mvan	Methanococcus vannielii SB	Mmus	Mus musculus
Mkan	Methanopyrus kandleri AV19	M_ha	Meloidogyne hapla
Mace	Methanosarcina acetivorans C2A	M_in	Meloidogyne incognita
Mbar	Methanosarcina barkeri str. Fusaro	Pp_s	Pristionchus pacificus
Mmaz	Methanosarcina mazei Go1	Scer	Saccharomyces cerevisiae
Mthe	Methanothermobacter thermautotrophicus str. Delta H	Sjap	Schistosoma japonicum
Mfri	Methanogenium frigidum Ace-2	Sman	Schistosoma mansoni
Nequ	Nanoarchaeum equitans Kin4-M	Tspv	Trichinella spiralis
Saci	Sulfolobus acidocaldarius DSM 639	Worm	Caenorhabditis elegans
Ssol	Sulfolobus solfataricus 98/2	Paer	Pyrobaculum aerophilum str. IM2
Stok	Sulfolobus tokodaii str. 7	Paby	Pyrococcus abyssi GE5
Tgam	Thermococcus gammatolerans EJ3	Pfur	Pyrococcus furiosus DSM 3638
Tkod	Thermococcus kodakarensis KOD1	Phor	Pyrococcus horikoshii OT3
Taci	Thermoplasma acidophilum DSM 1728	Ptor	Picrophilus torridus DSM 9790
Tvol	Thermoplasma volcanium GSS1		

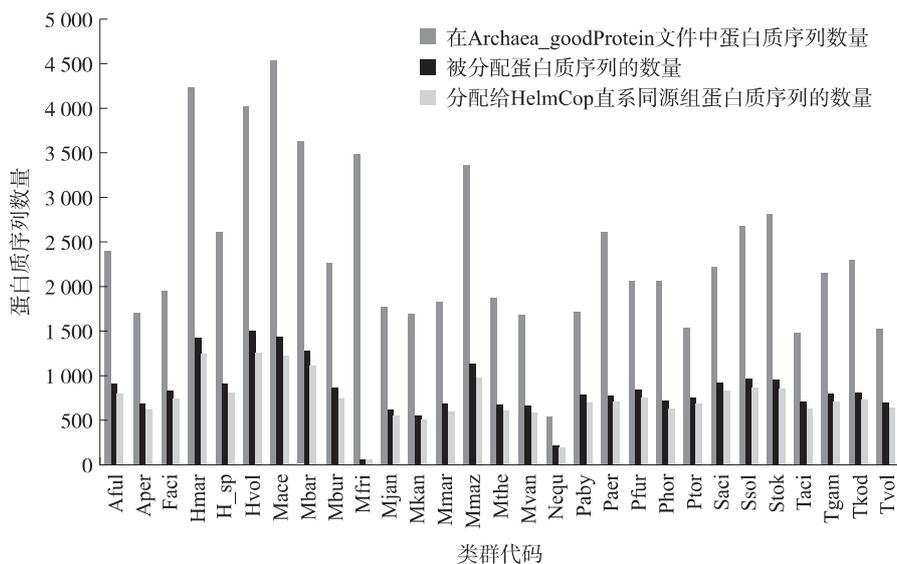


图 1 蛋白质序列数量对比柱状图

Fig.1 Comparison of numbers of protein sequences

HelmCoP 直系同源组和最佳匹配序列的性质受到我们的关注. 总体上看,那些接受到古菌序列的直系同源组含有多种多样生物功能的序列成员. 经过多次搜索查询,我们发现很多接受到最多数量古菌序列的直系同源组,有能够与其他生物分子结合,进而帮助运输这些分子的成员,例如直系同源组 ortho17taxa2398、ortho17taxa7962,以及 ortho17taxa12195 内的序列成员. 此外,还有一些直系同源组有参与新陈代谢过程,合成生物分子的功能的成员. 这些直系同源组中与古菌序列有最佳匹配的序列,很多是来

自自由线虫 *Caenorhabditis japonica*、大豆和拟南芥的蛋白质序列。其他一些物种, 比如人类、小鼠、酿酒酵母以及黑腹果蝇的蛋白质序列也在比较多的情况下与古菌序列有最佳匹配。详细情况见图 2。值得注意的是, 在归至 NO_GROUP 类别的分配过程中, 超过一半数量的最佳匹配序列来自自由线虫 *Caenorhabditis japonica* (图 2(c))。这种情况在其他两张饼图(图 2(a)和(b))中并没有出现。比较起来, 来自自由线虫 *Caenorhabditis remanei* 的最佳匹配序列也比较多, 但是来自其他寄生线虫和自由生活线虫的序列都不多。

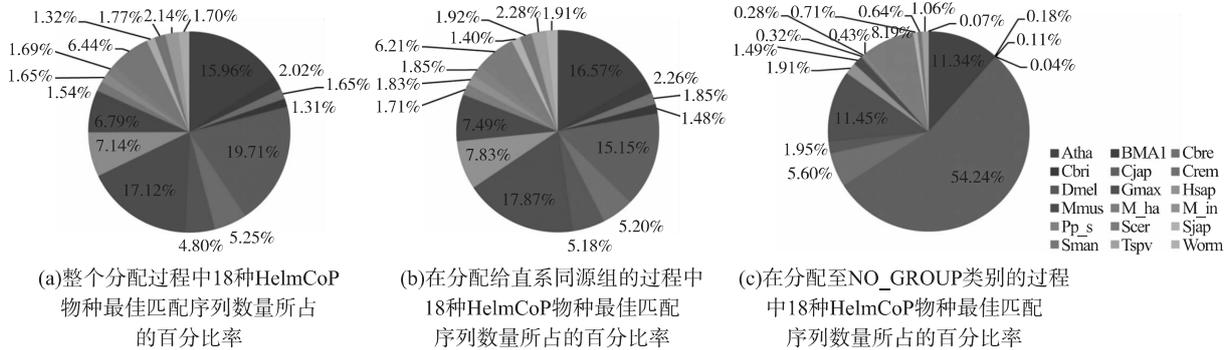


图 2 3 种情况下 18 种 HelmCoP 物种最佳匹配序列数量所占的百分比

Fig. 2 Percents of numbers of best hit sequences from all 18 HelmCoP species

所有的分配结果都经过了阈值筛选。在全部 24 147 个分配结果中, 有 159 个 BLAST 匹配具有数值为 0 的 E 值。这意味着该序列就是提交序列的正确匹配对象。作为比较, 17 个 BLAST 匹配的 E 值指数是 -5。指数 ≤ -50 的 E 值数量是 5 571 个, 占 23 988 (24 147 - 159) 个匹配的 23.22%。当我们考虑指数 ≤ -20 的 E 值数量, 这个百分比增加至 60.99%。如果再进一步考虑指数 ≤ -9 的 E 值数量, 这个百分比则为 88.93%。这显示绝大多数分配结果是正确和可信的, 因为 BLAST 比对结果的显著性高。当然, 根据 OrthoMCL 软件开发者提供的信息, 仍然可能有一小部分分配结果(低于 10%), 与通过 OrthoMCL 软件完全分析而产生的结果有所不同。

3 讨论

本研究方法是依据 OrthoMCL-DB 数据库网站工具的分析方法而设计的, 这个方法在 2011 年 9 月出版的一篇关于 OrthoMCL 软件的文章中有具体描述^[6]。根据实际情况, 本研究进行了一些调整变化, 主要区别之处在于对内旁系同源分组情况的处理。根据 OrthoMCL-DB 开发者的描述, 所有没有高于阈值匹配 (E 值和 percent_match 指标) 的蛋白质序列都会通过分析流程进一步处理。同时, 开发者还说, 如果通过网站工具提交多个蛋白质组加以处理, 将会把这些序列正确分配到各个直系同源组中, 但不能得到可信度高的内旁系同源组分组数据。因此, 在本研究中我们只采用了 OrthoMCL-DB 数据库网站工具分析方法的一部分, 即对直系同源组分配序列的处理来设计本研究的方法。毕竟, 给 HelmCoP 数据库各个直系同源组分配古菌序列是我们最关注的研究目标, 而这个目标理论上是可以准确而高效地实现的。

一些自由生活物种包括线虫 *Caenorhabditis japonica*、大豆和拟南芥在本研究中受到重视。因为相当多的古菌序列与这些物种的蛋白质序列有最佳匹配。来自线虫 *Caenorhabditis japonica* 的序列展现出与其他线虫序列明显不同的特点。以往研究已经显示出 *Caenorhabditis japonica* 和其昆虫携带者之间有独特的共生和腐生关系^[15]。更多的研究还发现, *Caenorhabditis japonica* 在不受到食物匮乏和过高种群密度影响的情况下, 就可以进入“多尔”幼虫状态^[16]。这些情况与其他线虫, 比如秀丽隐杆线虫的已知信息有很大差别。大豆与根瘤菌有共生关系, 而拟南芥可以和一些真菌, 例如 *Piriformospora indica* 有共生关系^[17]。遗憾的是, 这些物种都不属于古菌类别。为什么相当多数量的古菌基因在这几个物种基因组中有同源基因? 在演化历程中, 线虫 *Caenorhabditis japonica*、大豆和拟南芥与其他一些真核生物有怎样的差别? 目前的研究成果还不能很好地回答这些问题。研究人员需要更加深入地研究古菌与这 3 个物种间的演化关系。

与其他模式菌株的蛋白质组数据相比, 寒冷产甲烷菌模式菌株蛋白质组的测序质量可能偏低一些, 其中一些肽链上的部分氨基酸没能完全检测出来。这也许可以部分解释为什么来自这个古菌模式菌株序列的分配率明显低于来自其他古菌序列的分配率。但是, 更加值得关注的是它生长在南极洲的湖泊里, 具有

独特的生活环境和地理位置. 从系统发生角度看, 寒冷产甲烷菌与嗜酸甲烷八叠球菌、巴氏甲烷八叠球菌以及梅氏产甲烷八叠球菌等古菌的亲缘关系较近, 而来自这些古菌的序列均有 30% 以上的分配率. 因此, 寒冷产甲烷菌模式菌株及其蛋白质组值得进一步研究.

总结起来, 本研究提供了古菌序列在 HelmCoP 直系同源组中的分配信息. 此外, 一些自由生活物种包括线虫 *Caenorhabditis japonica*、大豆以及拟南芥, 与古菌之间的演化关系在本研究中显得十分突出. 寒冷产甲烷菌也因其与别的古菌间的差异之处, 而需要得到研究人员的重视.

[参考文献]

- [1] Brinkmann H, Philippe H. Archaea sister group of bacteria? Indications from tree reconstruction artifacts in ancient phylogenies [J]. *Molecular Biology and Evolution*, 1999, 16(6): 817–825.
- [2] Baldauf S L. The deep roots of eukaryotes [J]. *Science*, 2003, 300(5626): 1703–1706.
- [3] Derelle R, Lang B F. Rooting the eukaryotic tree with mitochondrial and bacterial proteins [J]. *Molecular Biology and Evolution*, 2012, 29(4): 1277–1289.
- [4] Brown J R. Ancient horizontal gene transfer [J]. *Nature Review Genetics*, 2003, 4(2): 121–132.
- [5] Abubucker S, Martin J, Taylor C M, et al. HelmCoP: an online resource for helminth functional genomics and drug and vaccine targets prioritization [J]. *PloS One*, 2011, 6(7): e21832.
- [6] Fischer S, Brunk B P, Chen F, et al. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups [J]. *Current Protocols in Bioinformatics*, 2011, 6: 1–19.
- [7] Chen F, Mackey A J, Stoeckert C J, et al. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups [J]. *Nucleic Acids Research*, 2006, 34(Suppl 1): D363–D368.
- [8] Altschul S F, Gish W, Miller W, et al. Basic local alignment search tool [J]. *Journal of Molecular Biology*, 1990, 215(3): 403–410.
- [9] Li L, Stoeckert C J Jr, Roos D S. OrthoMCL: identification of ortholog groups for eukaryotic genomes [J]. *Genome Research*, 2003, 13(9): 2178–2189.
- [10] Kersey P J, Staines D M, Lawson D, et al. Ensembl Genomes: an integrative resource for genome-scale data from non-vertebrate species [J]. *Nucleic Acids Research*, 2012, 40(D1): D91–D97.
- [11] Franzmann P D, Liu Y, Balkwill D L, et al. *Methanogenium frigidum* sp. nov., a psychrophilic, H₂-using methanogen from Ace Lake, Antarctica [J]. *International Journal of Systematic Bacteriology*, 1997, 47(4): 1068–1072.
- [12] Gribaldo S, Brochier-Armanet C. The origin and evolution of Archaea: a state of the art [J]. *Philosophical Transactions of the Royal Society of London Series B Biological Sciences*, 2006, 361(1470): 1007–1022.
- [13] Alsmark C, Foster P G, Sicheritz-Ponten T, et al. Patterns of prokaryotic lateral gene transfers affecting parasitic microbial eukaryotes [J]. *Genome Biology*, 2013, 25, 14(2): R19.
- [14] Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications [J]. *BMC Bioinformatics*, 2009, 10(1): 421.
- [15] Okumura E, Ishikawa Y, Tanaka R, et al. Propagation of *Caenorhabditis japonica* in the nest of its carrier insect, *Parastrachia japonensis* [J]. *Zoological Science*, 2013, 30(3): 174–177.
- [16] Kiontke K, Sudhaus W. Ecology of *Caenorhabditis* species [J]. *WormBook*, 2006: 1–14.
- [17] Nongbri P L, Vahabi K, Mrozinska A, et al. Balancing defense and growth—Analyses of the beneficial symbiosis between *Piriformospora indica* and *Arabidopsis thaliana* [J]. *Symbiosis*, 2013, 58(1/3): 17–28.

[责任编辑: 黄 敏]