微生物学通报
Microbiology China
tongbao@im.ac.cn

Nov. 20, 2016, 43(11): 2353−2365
http://journals.im.ac.cn/wswxtbcn
DOI: 10.13344/j.microbiol.china.150981

研究报告

产甲烷古菌中 CRISPR 簇的研究

方静　侯佳林　张宇　王风平　何莹*

(上海交通大学生命科学技术学院　上海　200240)

摘　要:【目的】通过对 51 个产甲烷古菌基因组中成簇的规律间隔短回文重复序列(Clustered regularly interspaced short palindromic repeats，CRISPR)的组成和来源进行研究，推测产甲烷古菌与环境中其他微生物的物质交换和相互作用，在基因组水平上阐述产甲烷古菌之间的遗传差异。【方法】利用 CRISPRdb 和 CRISPRFinder，找出产甲烷古菌基因组中所有潜在的 CRISPR 簇。对 CRISPR 簇的基本组成部分进行分析: 利用 BLASTCLUST 对重复序列(Repeat)进行分类; 分别将间隔序列(Spacer)与 Refseq 病毒基因组、Refseq 质粒基因组和 Refseq 产甲烷古菌基因组进行比对，从而获得间隔序列的物种来源和功能信息的注释。【结果】在 51 个产甲烷古菌中共找到了 196 个 CRISPR 簇，这些 CRISPR 簇中包含了总共 4 355 条间隔序列。在这些产甲烷古菌中，CRISPR 簇的分布是不均匀的，且每个物种的间隔序列数量与其 CRISPR 簇数量是不成正比的。在对重复序列进行分类之后，发现 Mclu1 是分布最广且最具代表性的一类重复序列。在 4 355 条间隔序列中有 388 条具有物种注释信息，266 条具有功能注释信息。从 CRISPR 簇间隔序列的来源来看，产甲烷古菌曾受到来自 Poxiviridae、Siphoviridae 以及 Myoviridae 属病毒的攻击，并且产甲烷古菌之间存在比较广泛的遗传物质交换。【结论】产甲烷古菌基因组中的 CRISPR 簇在组成和来源上存在较大的差异，这些差异与它们的生存环境有较大的关系。从 CRISPR 簇的角度阐述了产甲烷古菌之间基因组序列的差异。

关键词: 产甲烷古菌，CRISPR 簇，分布，生存环境

Distribution of CRISPR arrays in methanogenic archaea

FANG Jing　HOU Jia-Lin　ZHANG Yu　WANG Feng-Ping　HE Ying*

(School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: [Objective] By studying the distribution and origin of clustered regularly interspaced short palindromic repeats (CRISPR) in 51 methanogenic archaea genomes, material exchanges and interaction within methanogenic archaea and other microorganism were inferred, and the genomic divergence among the genomes of methanogenic archaea was characterized. **[Methods]** We identified all potential CRISPR arrays in the methanogenic archaea by applying the CRISPRdb and CRISPRFinder, and then the components of each CRISPR array were analyzed, where repeats were classified by BLASTCLUST and spacers were aligned to Refseq viral genome, Refseq plasmid

***Corresponding author:** E-mail: heying1982@sjtu.edu.cn
Received: November 27, 2015; **Accepted:** April 27, 2016; **Published online** (www.cnki.net)**:** June 16, 2016
*通讯作者：E-mail：heying1982@sjtu.edu.cn
收稿日期：2015-11-27；接受日期：2016-04-27；优先数字出版日期(www.cnki.net)：2016-06-16

genome and Refseq methanogenic genome respectively to retrieve both taxonomic and functional annotation. **[Results]** Among the 51 methanogenic archaea, in total 196 CRISPR arrays with 4 355 spacers were identified. The distribution of CRISPR arrays in methanogenic archaea was not even, and the number of spacers in one strain was not proportional to the number of CRISPR arrays in the strain. After clustering on repeat sequences of CRISPR arrays, we found that Mclu1 was the most diverse and representative repeat sequence in methanogenic archaea. Among 4 355 spacers, 388 spacers were assigned with taxonomic information and 266 spacers were assigned with functional annotation. By inferring the origin of each spacer sequence, we found that methanogenic archaea might be attacked by viruses belonging to Poxviridae, Siphoviridae or Myoviridae family. Moreover, and t exchanges of genetic materials among these archaea were observed. **[Conclusion]** Differences in the distribution and origin of CRISPR arrays in methanogenic archaea genomes were observed and characterized, and these differences might result from the interactions and conditions of their living environment. In this study, we could also infer from CRISPR arrays characterize the genomic divergence among methanogenic archaea.

Keywords: Methanogenic archaea, CRISPR array, Distribution, Environment

产甲烷古菌是一类严格厌氧的广古菌(Euryarchaeota)，它们具有一个共同的特点：能够进行甲烷生成过程(Methanogenesis)。甲烷生成过程是一个能量代谢过程，产甲烷古菌能够利用这个过程将一系列底物，包括 CO_2 和 H_2、甲酸盐、甲醇、甲胺或/和乙酸盐转化为甲烷[1-2]。据估计，每年全球大约有 10 亿 t 的甲烷由产甲烷古菌产生[2]，因此产甲烷古菌在全球碳循环中起到非常重要的作用。产甲烷古菌由以下 5 个目组成，它们是 Methanobacteriales 、Methanococcales 、Methanomicrobiales 、Methanosarcinales 和 Methanopyrales。不同目中产甲烷古菌的 16S rRNA 基因序列相似度至少为 82%[3]，它们之间的差异主要体现在细胞膜结构、脂质组成以及产甲烷过程中使用的底物等性质上。产甲烷古菌主要生存在厌氧环境中，如海洋沉积物、淡水沉积物、淹水土壤、人和动物的胃肠道以及垃圾场等环境[3]。

成簇的规律间隔的短回文重复序列(Clustered regularly interspaced short palindromic repeats，CRISPR)是微生物防御来自噬菌体、质粒或其他来源核酸序列入侵的特殊组成部分[4-5]。据估计，大约有 40%的细菌基因组和 90%的古菌基因组具有 CRISPR 簇[6]。这些微生物能够利用一种称为 CRISPR/Cas (CRISPR-associated)的系统对外源核酸序列进行识别并将其降解[4-5]。CRISPR/Cas 系统是一种适应性免疫系统，其机理类似于真核生物中的 RNA 干涉(RNA interference，RNAi)。

CRISPR 簇由前导序列(Leader sequence)、重复序列(Repeat)和间隔序列(Spacer)组成(图 1)。前导序列位于 CRISPR 簇的上游，是一段富含 AT 且长度大约有几百个碱基的非保守区域，其功能可能是作为启动子[7]。重复序列的数量在不同基因组中存在着很大的差异，从最少的 2 个到几百个不等，而长度一般在 23–50 个碱基。重复序列的一部分功能已经被发现：在降解外源核酸时，为了避免 CRISPR 序列被一同降解掉，CRISPR 序列中的重复序列会与 crRNA (CRISPR RNA)中的重复序列进行互补配对[8]。重复序列被组成不同的间隔序列分开，这些间隔序列的长度从 17–84 个碱基不等。目前已经证实了间隔序列就是外源核酸被微生物识别并被整合到 CRISPR 簇中的序列[9]。在 CRISPR 簇的周围存在着 CRISPR 相关(CRISPR-associated，Cas)蛋白，这些蛋白在整个免疫过程中起着至关重要的作用。在外源核酸初次入侵时，Cas 蛋白会识别并裂解外源核酸序列并将其整合到 CRISPR 簇中[10]，最近已有文章对这一过程的详细机制进行了报道[11-12]；当外源核酸再次入侵时，Cas 蛋白协助 CRISPR 序列转录成 crRNA，并与 crRNA 一起

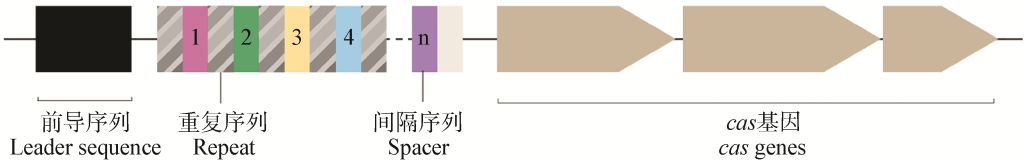


图 1　CRISPR 簇的特征图

Figure 1　Features of CRISPR array

注：CRISPR 簇由前导序列(黑色盒)、重复序列(灰色斜纹盒)、间隔序列(彩色盒)和 CRISPR 相关基因构成.

Note: CRISPR array consists of leader sequence (black box), repeats (gray boxes with stripes), spacers (color boxes) and CRISPR-associated genes.

识别并降解外源核酸序列[10]。另外，Cas 蛋白的结构特点和具体作用也正在被大量研究，目前已有对 Cas1 和 Cas2 蛋白结构和功能特点的报道[13-15]。通过对产甲烷古菌基因组中 CRISPR 簇的识别和鉴定，研究者能够从中了解产甲烷古菌中 CRISPR 簇的分布规律和特点，以及这些古菌之间的分子差异。然后通过找到 CRISPR 簇间隔序列的物种信息，可以帮助增进对产甲烷古菌与环境中病毒或噬菌体的相互作用以及产甲烷古菌之间物质交换的认知。

1　材料与方法

1.1　产甲烷古菌基因组中 CRISPR 簇的鉴定

大多数产甲烷古菌 CRISPR 簇的数据来自于 CRISPRdb (http://crispr.u-psud.fr/crispr/)[6]。由于这个数据库中没有包含 *Methanocella arvoryzae* MRE50，所以通过 CRISPRFinder[16](默认参数)对其进行 CRISPR 簇的预测，最后只保留了 CRISPRFinder 认为可信的 CRISPR 簇。

1.2　对重复序列进行分类

利用 BLASTCLUST[17]对所有重复序列进行聚类(-L0.5-S50-eF-pF-W15)[18]，然后根据类别中序列数量的多少依次进行编号，最后利用 ClustalW[19]对每个类别的序列进行序列比对，将比对后的结果用 WebLogo[20]显示，即得到每个类别的代表序列。

将本研究中得到的重复序列的类别与 Kunin 等[21]得到的结果进行比较。首先，利用 HMMBUILD (默认参数)对 Kunin 研究中的 12 个重复序列类别分别构建隐式马尔可夫模型(Hidden Markov Model，HMM)，然后通过 HMMSEARCH (-E1e−07-Z1)将本研究中的序列与这 12 个类别的 HMM 模型进行比较。

1.3　Cas 基因的鉴定

将产甲烷古菌的序列分别与 Pfam 数据库[22] (e-value: 0.001)、TIGRFAM 数据库[23](e-value: 0.001)和 COG 数据库[24](e-value: 0.001)进行比对注释，并整合各个来源的注释结果。

1.4　16S rRNA 基因进化树的构建

所有相关的序列都下载自 NCBI (National Center for Biotechnology Information)数据库，若 NCBI 中没有对应的序列信息，则通过 WebMGA[25]中的 HMMER 3.0[26]对基因组进行 16S rRNA 基因的预测。接着，通过 ClustalW 进行多重序列比对，比对后的结果会被去除缺口(Gaps)。然后通过 FastTree2[27]基于最大似然法构建进化树，其中使用的模型为 GTR (Generalized Time Reversible)，与 CAT 近似。最后通过 MEGA[28]对进化树构建结果进行可视化。

1.5　间隔序列的注释

利用 BLASTn (-gapopen 10-gapextend 2-penalty-1-reward 1-word_size 7-evalue 0.01)[29]将间隔序列与如下数据库进行比对：Refseq 病毒基因组(共 5 301 个)、Refseq 质粒基因组(4 402 个)和 Refseq 产甲烷古菌基因组(51 个)。剔除了与自身相匹配的情况，并且只检查符合如下条件的比对结果：覆盖率≥0.5，相似率≥0.85，e-value≤0.01。

每个间隔序列可能有多个满足上述标准的匹

配结果(Hits)。为了筛选出可信的匹配结果，对每个匹配结果根据如下公式进行打分[30]：得分=3×覆盖率+相似率。每个间隔序列的注释信息(包括功能注释和物种来源注释，Function and organism of the best hit)会取各自匹配结果中得分最高的。即便如此，还是会有一些间隔序列有多个满足条件的匹配结果。因此，规定在统计物种或功能信息时，当某条间隔序列拥有多个匹配结果且这些匹配结果的注释信息属于不同的统计单元时，进行多次统计。对于功能注释，要检查间隔序列是否匹配到了基因的编码区域(Coding regions)。如果匹配到了基因的编码区域，那么相应的基因将会与 COG 数据库(e-value: 0.001)进行重新注释，以获得更准确的功能注释信息。

2 结果与分析

2.1 产甲烷古菌中 CRISPR 簇的分布

如表 1 所示，在 51 个产甲烷古菌的基因组中总共找到了 196 个 CRISPR 簇。在这些 CRISPR 簇中，有一个簇是属于 *Methanocaldococcus* sp. FS406-22 的质粒 pFS01 的。有 4 个产甲烷古菌中发现了超过 20 个 CRISPR 簇，然而超过 50% (28/51)的产甲烷古菌含有的 CRISPR 簇的数量不超过 3 个，其中还有 10 个菌种没有发现任何 CRISPR 簇。这些发现表明产甲烷古菌中 CRISPR 簇的分布是不均匀的，尤其是一些 Methanococcales 古菌中含有大量的 CRISPR 簇。

在这 196 个 CRISPR 簇中存在着 4 355 条间隔序列。如图 2 所示，每个菌种的 CRISPR 簇数和间隔序列数被其相应的基因组大小所归一化。在这 51 个产甲烷古菌中，有 6 个菌种拥有超过 200 条间隔序列，它们是：*Methanospirillum hungatei* JF-1、*Methanothermococcus okinawensis* IH1、*Methanobrevibacter* sp. AbM4、*Methanococcus voltae* A3、*Methanocaldococcus* sp. FS406-22 和 *Methanocaldococcus vulcanius* M7。表明这 6 种产甲烷古菌很可能与外界发生过较为频繁的遗传物质交换，此外这些物种在基因组水平上可能存在更大的差异。

值得注意的是，每个菌种间隔序列的数量与其 CRISPR 簇的数量不成正比(图 2)。比如，*Methanobrevibacter* sp. AbM4 只拥有一个含有 246 条间隔序列的 CRISPR 簇，然而 *Methanocaldococcus vulcanius* M7 的 215 条间隔序列却分布于 20 个 CRISPR 簇中。更有趣的是，大部分具有大量间隔序列的产甲烷古菌属于 Methanococcales 和 Methanobacterials，尤其是 Methanococcales (图 2)。另外，在产甲烷古菌中大多数 CRISPR 簇含有较少的间隔序列(<10 条)，而少数的 CRISPR 簇却拥有大量的间隔序列。比如，接近 50% (96/196)的 CRISPR 簇含有少于 10 条间隔序列，但是 9 个 CRISPR 簇却拥有超过 100 条间隔序列。不同产甲烷古菌中 CRISPR 簇的组成和分布的差异，很有可能与物种所在生存环境的差异有关。

2.2 对重复序列进行分类

196 条 CRISPR 簇的重复序列可分成 51 个类别，其中 27 个类别只有一条序列，表明产甲烷古菌 CRISPR 簇的类别存在很高的特异性。而剩下的 24 个类别总共含有 169 条重复序列，其中最大的类别(Mclu1)拥有 43 条序列(表 2)。在过去的研究中，Kunin 等在 195 个微生物基因组中发现了 561 个 CRISPR 簇，并把这些 CRISPR 簇的重复序列分成了 12 个类别(命名为 Cluster1–12)。为了与 Kunin 的结果进行比较，对这 12 个类别分别构建了 HMM 模型，并将本研究中得到的重复序列与这些模型进行对比(如材料和方法中的描述)。通过比对，发现本研究中的 196 条重复序列中有 67 条(来自于 17 个类别)与这 6 个 HMM 模型相匹配。而许多 Kunin 等认为是相同类别的序列在本研究中却被分到了不同类别。另外，在这 6 个 Kunin 研究的类别中有 3 个(Cluster1，2，3)被鉴定是属于细菌的。这

表 1 产甲烷古菌中鉴定出的 CRISPR 簇的基本信息

Table 1 General statistics of CRISPR array identification in methanogens

菌种 Species	CRISPR 簇数 CRISPR array numbers	间隔序列数 Spacers numbers	基因组大小 Genome size (bp)	Contig 数 Contig numbers	*cas* 基因数 *cas* gene numbers
Candidatus Methanomassiliicoccus intestinalis Issoire-Mx1	2	112	1 931 651	1	9
Methanocorpusculum labreanum Z	1	18	1 804 962	1	12
Methanosphaerula palustris E1-9c	2	104	2 922 917	1	17
Methanoregula formicica SMSP	2	143	2 820 858	1	22
Methanoregula boonei 6A8	0	0	2 542 943	1	3
Methanospirillum hungatei JF-1	7	264	3 544 738	1	32
Methanoculleus bourgensis MS2	1	144	2 789 774	1	8
Methanoculleus marisnigri JR1	1	6	2 478 101	1	5
Methanoplanus petrolearius DSM 11571	0	0	2 843 290	1	2
Methanocella conradii HZ254	1	81	2 378 438	1	11
Methanocella paludicola SANAE	0	0	2 957 635	1	4
Methanocella arvoryzae MRE50	1	114	3 179 916	1	15
Methanosaeta concilii GP6	6	95	3 026 645	2	17
Methanosaeta thermophila PT	2	156	1 879 471	1	26
Methanosaeta harundinacea 6Ac	1	39	2 571 034	2	14
Methanosarcina mazei Go1	4	128	4 096 345	1	24
Methanosarcina mazei Tuc01	3	5	3 427 949	1	11
Methanosarcina barkeri str. Fusaro	4	95	4 873 766	2	26
Methanosarcina acetivorans C2A	5	69	5 751 492	1	19
Methanolobus psychrophilus R15	0	0	3 072 769	1	9
Methanomethylovorans hollandica DSM 15978	3	25	2 714 013	2	12
Methanococcoides burtonii DSM 6242	2	84	2 575 032	1	21
Methanohalophilus mahii DSM 5219	0	0	2 012 424	1	5
Methanohalobium evestigatum Z-7303	0	0	2 406 232	2	5
Methanosalsum zhilinae DSM 4017	1	27	2 138 444	1	13
Methanocaldococcus infernus ME	14	144	1 328 194	1	19
Methanocaldococcus vulcanius M7	20	215	1 761 737	3	25
Methanocaldococcus fervens AG86	8	79	1 507 251	2	25
Methanocaldococcus sp. FS406-22*	23 (1)	231 (8)	1 773 136	2	22

(待续)

(续表)

Methanocaldococcus jannaschii DSM 2661	20	178	1 739 927	3	21
Methanotorris igneus Kol 5	25	178	1 854 197	1	21
Methanococcus aeolicus Nankai-3	1	18	1 569 500	1	14
Methanothermococcus okinawensis IH1	8	247	1 677 465	2	16
Methanococcus maripaludis X1	1	55	1 746 697	1	11
Methanococcus maripaludis S2	0	0	1 661 137	1	4
Methanococcus maripaludis C6	0	0	1 744 193	1	2
Methanococcus maripaludis C5	1	27	1 789 046	2	9
Methanococcus maripaludis C7	0	0	1 772 694	1	3
Methanococcus vannielii SB	2	103	1 720 048	1	18
Methanococcus voltae A3	3	233	1 936 387	1	32
Methanothermobacter marburgensis str. Marburg	2	39	1 639 135	2	4
Methanothermobacter thermautotrophicus str. Delta H	2	169	1 751 377	1	23
Methanobacterium sp. SWAN-1	1	79	2 546 541	1	15
Methanobacterium sp. AL-21	1	7	2 583 753	1	7
Methanobacterium sp. MB1	2	88	2 029 766	1	11
Methanosphaera stadtmanae DSM 3091	3	121	1 767 403	1	16
Methanobrevibacter ruminantium M1	3	110	2 937 203	1	20
Methanobrevibacter smithii ATCC 35061	1	43	1 853 160	1	12
Methanobrevibacter sp. AbM4	1	246	1 998 189	1	14
Methanothermus fervidus DSM 2088	0	0	1 243 342	1	3
Methanopyrus kandleri AV19	5	36	1 694 969	1	13

注：*标记菌种的质粒中发现了 CRISPR 簇. 在对应位置处用括号标明了质粒中 CRISPR 簇的数据，括号外为该菌种所有 CRISPR 簇的数据(包括质粒中的 CRISPR 簇). 表 1 中所有菌种的基因组皆为完整的基因组.

Note: The star sign indicated that there were CRISPR arrays in the plasmid of this archaea. The information of CRISPR array in plasmid was in the brackets, and the total information of CRISPR array in the archaea was outside of the brackets. The all genomes in the table 1 were complete.

表明Kunin的分类方法并没有很好地将这些古菌的重复序列从细菌序列中分离出来，而我们的方法则具有较高的分辨度。在本研究的 24 个含有多条重复序列的类别中，有9个类别的重复序列来自于不同的菌种。其中物种多样性最丰富的是Mclu1，这个类别中的重复序列来自于 13 个不同的物种，而排在第二位的 Mclu2 的物种种类却只有 2 种。因此，可认定Mclu1 是在产甲烷古菌中分布最广的一类 CRISPR 簇的重复序列，而其他的重复序列类别具有较高的物种特异性，反映了产甲烷古菌间的分子差异。

2.3 间隔序列的物种注释

将得到的 4 355 条间隔序列与已知的病毒、质粒和 51 个产甲烷古菌的基因组对比之后，发现有388 (8.9%)条间隔序列获得了物种的注释信息。

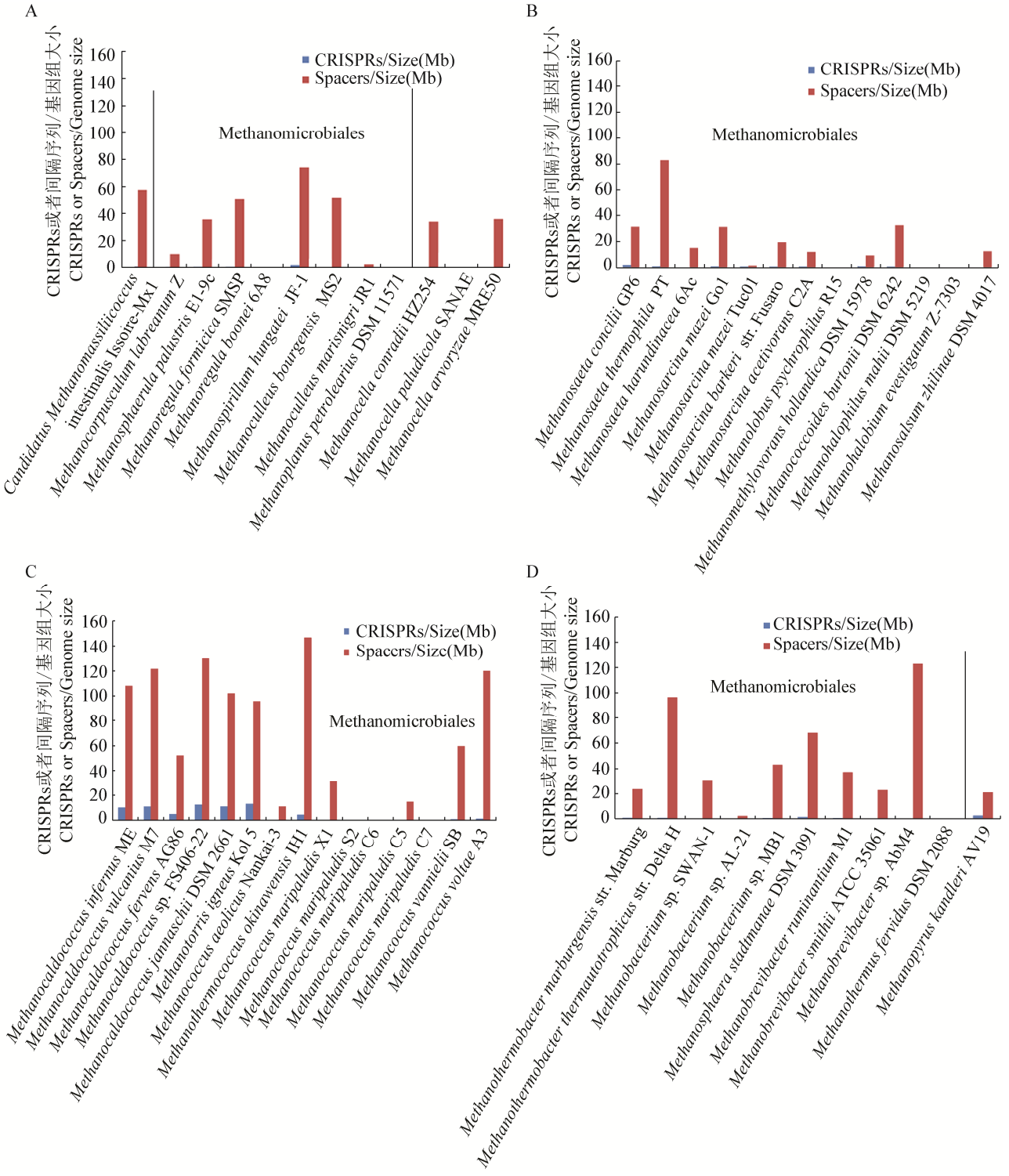


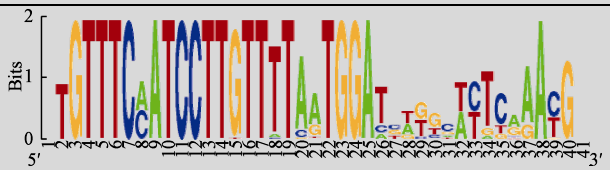
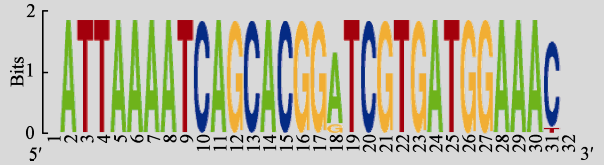
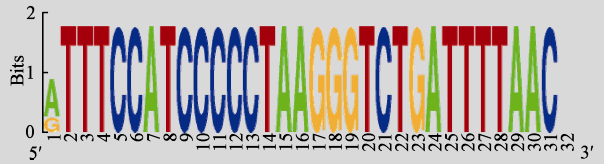
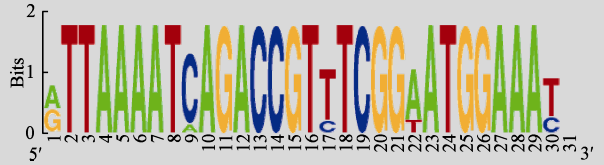
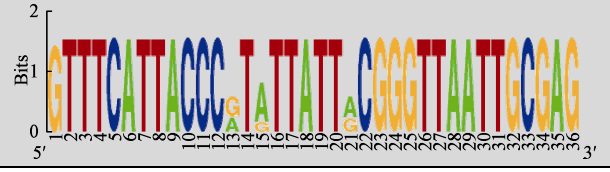
图 2　产甲烷古菌中 CRISPR 簇和间隔序列的分布图

Figure 2　The distribution of CRISPR array and spacers in methanogens

注：分布图被分为 4 个部分：A、B、C、D. 图中纵坐标为被基因组大小(Mb)归一化后的 CRISPR 数量或间隔序列数量.

Note: The distribution was divided into four pictures: A, B, C and D. In these pictures, the Y axis is number of CRISPRs or spacers normalized by the size (Mb) of genome.

表 2 CRISPR 重复序列类别的前 9 个类别
Table 2 Top9 cluster in CRISPR repeat clusters

类别 Cluster	重复序列数(间隔序列数，菌种数) Repeats numbers (Spacers numbers, Species numbers)	序列标志 Sequence logo	文献中的类别(门) Cluster in the reference (Phylum)
Mclu 1	43 (567, 13)		
Mclu 2	16 (105, 2)		Cluster 1 (Bacteria)
Mclu 3	16 (182, 1)		
Mclu 4	15 (185, 2)		Cluster 1 (Bacteria)
Mclu 5	10 (91, 1)		
Mclu 6	10 (111, 2)		Cluster 1 (Bacteria)
Mclu 7	8 (70, 1)		
Mclu 8	8 (247, 1)		Cluster 1 (Bacteria)
Mclu 9	5 (36, 1)		

2.3.1 与病毒基因组比对的结果：通过比对，发现有 86 条间隔序列的匹配结果符合筛选标准。根据 ICTV (International Committee on Taxonomy of Viruses Classification)，有 75 条间隔序列能够被分配到特定病毒的属(图 3)，其主要组成为：Poxviridae，19 条；Siphoviridae，16 条；Myoviridae，14 条。在物种信息注释为 Spihoviridae 属的 16 条间隔序列中有 10 条来自于 *Methanothermobacter thermautotrophicus* strain Delta H 的 CRISPR 簇，并且这 10 条间隔序列的匹配结果满足了一个非常严格的标准(覆盖率≥0.9，相似率≥0.9，e-value≤1e−05)，这其中还有 6 个完全匹配的结果(覆盖率=100%，相似率=100%)。由于这 10 条间隔序列的匹配结果非常好，并且考虑到病毒序列的高度可变性，认为这些间隔序列是相应菌种在最近受到外源核酸序列攻击而留下的痕迹。另外，这 10 条间隔序列来自于同一个 CRISPR 簇，其中 4 条序列的物种注释为 Methanothermobacter prophage psiM100，而其余 6 条的物种注释为 Methanobacterium phage psiM2。这些发现可能表明 *Methanothermobacter thermautotrophicus* strain Delta H 曾成功地抵御了来自于这两种噬菌体的攻击。因此推断这两种噬菌体攻击 *Methanothermobacter thermautotrophicus* strain Delta H 的机制可能是非常相似的，并且它们可能对产甲烷古菌的攻击比较频繁。

2.3.2 与质粒基因组比对的结果：总共有 66 条间隔序列的物种信息注释为质粒，其中 58 条属于细菌的质粒，7 条属于古菌的质粒，还有 1 条属于真核生物的质粒。58 个细菌质粒的寄主分别属于：Bacillales，9 条；Spirochaetales，9 条；Clostridiales，7 条；古菌质粒的寄主信息为：Methanococcales，4 条；Halobacteriales，2 条；Methanosarinales，1 条(图 4)。这些结果表明，产甲烷古菌很可能遭受更多来自于细菌质粒的攻击。

2.3.3 与产甲烷古菌基因组比对的结果：有 240 条间隔序列与产甲烷古菌基因组成功匹配，并且间隔序列的物种注释信息都基于古菌目的分类单元进行统计。通过与质粒的结果相结合，可以推测产甲烷古菌之间的相互作用关系(图 5)。通过图 5 能够发现产甲烷菌之间存在着核酸序列的相互作用，比如水平基因转移(Horizontal gene transfer，HGT)。比如 *Methanocaldococcus jannaschii* DSM 2661 曾遭受来自于 Methanosarcinales 古菌、Methanococcales 古菌和 Methanobacteriales 古菌的

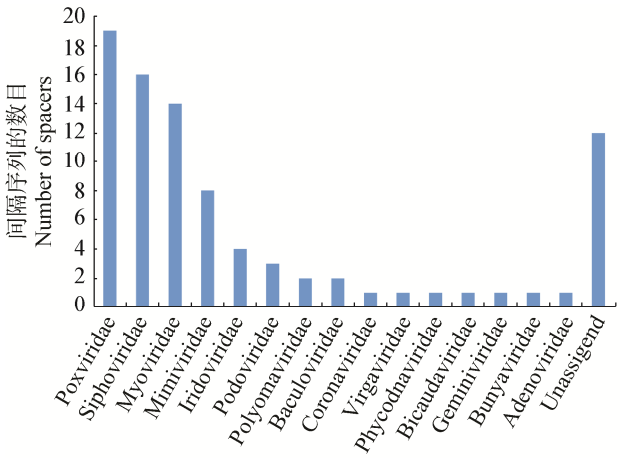


图 3 来自于病毒的间隔序列物种分布图

Figure 3 The distribution of spacers from virus

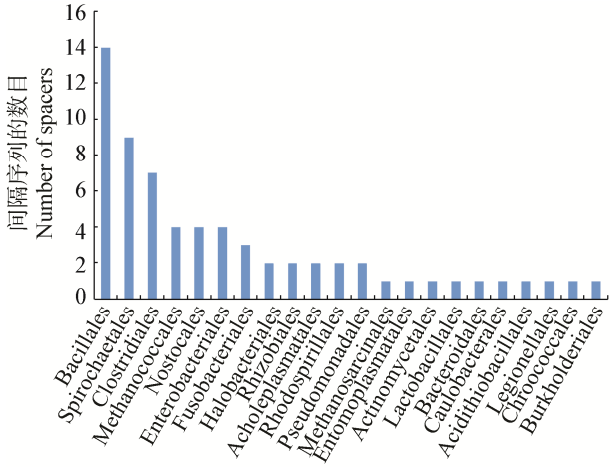


图 4 来自于质粒的间隔序列物种分布图

Figure 4 The distribution of spacers from plasmid

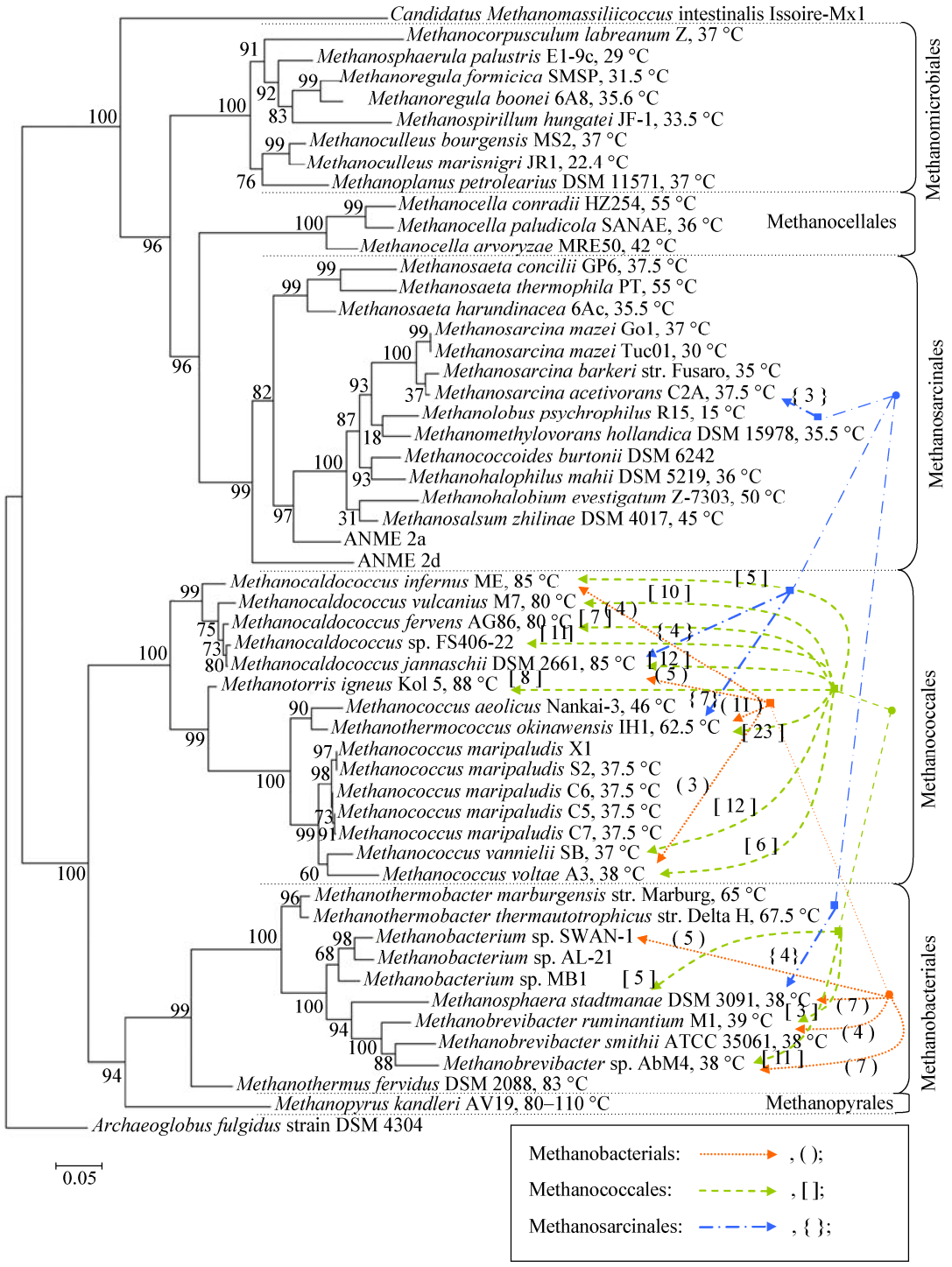


图 5 产甲烷古菌 16S rRNA 基因进化树

Figure 5 16S rRNA gene phylogenetic tree of all methanogens investigated in this study

注：图中用带箭头的曲线表示起始端的物种的核酸序列会攻击箭头端的物种并留下间隔序列，而不同括号中的数字表示鉴定出的间隔序列的数量.

Note: Matches between spacers and their targets were displayed by directed arrows that from spacers to targets.

序列“攻击”。由于 *Methanocaldococcus jannaschii* DSM 2661 的最适温度为 85 °C，而图 5 中 Methanosarcinales 古菌的已知最适温度都比这个菌要低得多。因此，这个菌遭受 Methanosarcinales 古菌的攻击很可能是处在其演化的早期阶段。推测在演化过程中，微生物通过诸如水平基因转移的方法来实现共演化，同时，为了保持自身遗传物质的稳定性，微生物演化并保留了 CRISPR/Cas 系统来抵御这些外源核酸序列。另外，我们还发现这些微生物曾遭受到来自于相同目的菌种的攻击。比如，*Methanococcus vannielli* strain SB 的 CRISPR 簇中有 12 条间隔序列的物种注释信息为 Mehthanococcales 目。当匹配结果满足一个非常严格的条件时(覆盖率≥0.9，相似率≥0.9，e-value≤1e−05)，我们认为这些间隔序列来自于自身的基因组。比如，3 条来自于 *Methanococcus voltae* A3 的间隔序列就是自我匹配的，其中 2 条的功能注释为假设的 ORFs (Hypothetical ORFs)，而剩余的 1 条则为与运动相关的基因(Mobility related gene)。这表明 CRISPR/Cas 系统能够将自身的核酸序列整合进间隔序列中，因此进一步说明 CRISPR/Cas 系统不仅能够抵御外源核酸的攻击还可以进行“自免疫”(Autoimmunity)。

2.4 间隔序列的功能注释

将得到功能注释信息的间隔序列根据以下功能类别进行分类：病毒或者前病毒基因(Virus or provirus gene)、持家基因(Housekeeping gene)、质粒的基因(Plasmid gene)、整合酶(Integrase)、转座酶(Transposases)、与运动相关的基因、*cas* 基因和假设的 ORFs。最终，在 388 条成功取得物种注释的间隔序列中，有 266 条具有功能注释(由于有 2 条间隔序列有 2 个符合条件的匹配结果，因此最终的匹配结果总共有 268 对)。如图 6 所示，超过 40%的间隔序列与质粒或病毒的基因相匹配。接近 24%的间隔序列为持家基因，这表明 CRISPR/Cas 系统经常抵御来自于染色体相关的核酸序列攻击，而这些序列可能通过水平基因转移获得。另外，来自于 *Methanocaldococcus infernus* strain ME 的 CRISPR 簇的 2 条间隔序列与来自于 Methanococcales 古菌的 *cas* 基因相匹配，但是由于这 2 个匹配结果没有达到非常高的标准(覆盖率≥0.9，相似率≥0.9，e-value≤1e−05)，所以无法确定 *cas* 基因的确切来源。

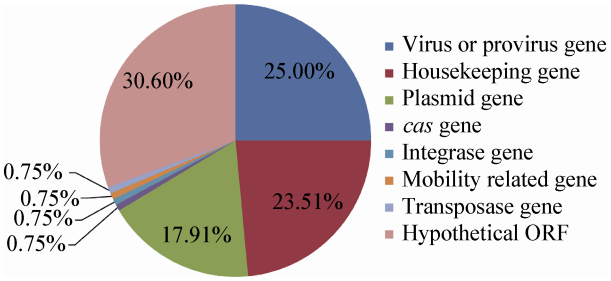


图 6 间隔序列的功能注释分布图

Figure 6 The function distribution of the spacers

3 讨论

在 51 个产甲烷古菌中总共找到了 196 个 CRISPR 簇，在这些 CRISPR 簇中包含了 4 355 条间隔序列。在产甲烷古菌中，CRISPR 簇的分布是不均匀的，并且每个菌种间隔序列的数量与其 CRISPR 簇的数量不成正比。在对重复序列进行分类之后，我们发现 Mclu1 是分布最广且最具代表性的重复序列。在 4 355 条间隔序列中有 388 条具有物种注释信息，266 条具有功能注释信息。在这 388 条有物种注释信息的间隔序列中大多数是与产甲烷古菌有关的。病毒的核酸序列是高度可变的且间隔序列又过于短小，因此有理由相信大部分未得到物种信息注释的间隔序列很可能来自病毒。但确切的结论还需进一步研究。通过本研究，我们得出如下结论：产甲烷古菌不仅受到来自外界的病毒(Poxiviridae、Siphoviridae 以及 Myoviridae 属)、细菌或其它种类古菌的攻击，而且产甲烷古菌之间也存在较频繁的遗传物质交换。

如表 1 和图 2 所示，CRISPR 簇的数量较多的菌种为 Methanococcales 目的嗜热菌种，而间隔序列的数量较多的菌种为 Methanococcales 目和

Methanobacteriales 目的部分菌种。我们知道，间隔序列是外源核酸序列进攻留下的痕迹，间隔序列数量的多少从某种程度上反映了该菌种遭受外源核酸攻击的次数的多少。因此，具有以下两种特征的 CRISPR 簇的菌种会受到较多的外源核酸序列攻击：大量间隔序列分布于较多的 CRISPR 簇(8−25 个)中，比如 Methanococcales 目的嗜热菌种；大量间隔序列分布于较少的 CRISPR 簇(1−4 个)中，比如 Methanobacteriales 目的部分菌种。通过结合菌种自身的生长温度以及自身基因组的稳定性，我们发现这两者之间可能存在着某种联系。Methanococcales 目的嗜热菌种多来源于高温环境(最适生长温度大于 80 °C)中，在高温条件下遗传物质交换更加频繁，因此很可能会遭受更频繁的外源序列的攻击。并且，这些菌种基因组的热稳定性较差(GC 含量为 30%−35%)，这样，外源序列进攻更容易成功。因此，这些菌种演化并保留下来的具有多个 CRISPR 簇(8−25 个)的 CRISPR/Cas 系统，可以有效地阻止外源序列的进攻并且维持基因组的稳定性。而另一些嗜热菌，如 *Methanothermobacter thermautotrophicus* str. Delta H，生存温度相对较低(最适生长温度 60−70 °C)，其自身基因组的热稳定性较强(GC 含量为 50%左右)，其 CRISPR/Cas 系统可能不需要太多的 CRISPR 簇就能够应付外源序列的攻击。而那些嗜温或嗜冷产甲烷菌的生存环境并不会出现频繁的外源序列攻击，因此，它们也不需要保留太多的 CRISPR 簇。从这个角度出发，我们可以推断生存温度越高的菌种，遭到外源序列的攻击越频繁，遗传物质的交换也就越频繁，这类物种间基因组序列上差异会随着演化过程而逐渐增大，相应地，它们之间的遗传差异也就越大。

综上所述，产甲烷古菌基因组水平上的差异，与其所生存的环境具有较大的关系。高温环境，或者类似的极端环境，有可能在某种程度上推动了产甲烷古菌之间的遗传物质交换，而 CRISPR/Cas 系统的存在则在一定程度上保持住了菌种内遗传物质的稳定性，对不同物种间的遗传差异起到了一定的贡献。

参 考 文 献

[1] Garcia JL, Patel BKC, Ollivier B. Taxonomic, phylogenetic, and ecological diversity of methanogenic Archaea[J]. Anaerobe, 2000, 6(4): 205-226

[2] Thauer RK. Biochemistry of methanogenesis: a tribute to Marjory Stephenson[J]. Microbiology, 1998, 144(9): 2377-2406

[3] Liu YC, Whitman WB. Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea[J]. Annals of the New York Academy of Sciences, 2008, 1125: 171-189

[4] Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea[J]. Nature Reviews: Genetics, 2010, 11(3): 181-190

[5] Horvath P, Barrangou R. CRISPR/Cas, the immune system of Bacteria and Archaea[J]. Science, 2010, 327(5962): 167-170

[6] Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats[J]. BMC Bioinformatics, 2007, 8: 172

[7] Makarova KS, Grishin NV, Shabalina SA, et al. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action[J]. Biology Direct, 2006, 1: 7

[8] Marraffini LA, Sontheimer EJ. Self versus non-self discrimination during CRISPR RNA-directed immunity[J]. Nature, 2010, 463(7280): 568-571

[9] Barrangou R, Fremaux C, Deveau H, et al. CRISPR provides acquired resistance against viruses in prokaryotes[J]. Science, 2007, 315(5819): 1709-1712

[10] Makarova KS, Haft DH, Barrangou R, et al. Evolution and classification of the CRISPR-Cas systems[J]. Nature Reviews: Microbiology, 2011, 9(6): 467-477

[11] Wang JY, Li JZ, Zhao HT, et al. Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems[J]. Cell, 2015, 163(4): 840-853

[12] Arslan Z, Hermanns V, Wurm R, et al. Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system[J]. Nucleic Acids Research, 2014, 42(12): 7884-7893

[13] Ka D, Kim D, Baek G, et al. Structural and functional characterization of *Streptococcus pyogenes* Cas2 protein under different pH conditions[J]. Biochemical & Biophysical Research Communications, 2014, 451(1): 152-157

[14] Kim TY, Shin M, Yen LHT, et al. Crystal structure of Cas1 from *Archaeoglobus fulgidus* and characterization of its nucleolytic activity[J]. Biochemical and Biophysical Research Communications, 2013, 441(4): 720-725

[15] Gunderson FF, Mallama CA, Fairbairn SG, et al. Nuclease activity of *Legionella pneumophila* Cas2 promotes intracellular infection of amoebal host cells[J]. Infection & Immunity, 2014, 83(3): 1008-1018

[16] Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: A web tool to identify clustered regularly interspaced short palindromic repeats[J]. Nucleic Acids Research, 2007, 35: W52-W57

[17] Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool[J]. Journal of Molecular Biology, 1990, 215(3): 403-410

[18] Sorokin VA, Gelfand MS, Artamonova II. Evolutionary dynamics of clustered irregularly interspaced short palindromic repeat systems in the ocean metagenome[J]. Applied and Environmental Microbiology, 2010, 76(7): 2136-2144

[19] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice[J]. Nucleic Acids Research, 1994, 22(22):

4673-4680
[20] Crooks GE, Hon G, Chandonia JM, et al. WebLogo: a sequence logo generator[J]. Genome Research, 2004, 14(6): 1188-1190
[21] Kunin V, Sorek R, Hugenholtz P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats[J]. Genome Biology, 2007, 8(4): R61
[22] Finn RD, Mistry J, Tate J, et al. The Pfam protein families database[J]. Nucleic Acids Research, 2010, 40(Database issue): 263-266
[23] Haft DH, Selengut JD, White O. The TIGRFAMs database of protein families[J]. Nucleic Acids Research, 2003, 31(1): 371-373
[24] Tatusov RL, Fedorova ND, Jackson JD, et al. The COG database: an updated version includes eukaryotes[J]. BMC Bioinformatics, 2003, 4: 41
[25] Wu ST, Zhu ZW, Fu LM, et al. WebMGA: a customizable web server for fast metagenomic sequence analysis[J]. BMC Genomics, 2011, 12: 444
[26] 'HMMER', http://hmmer.janelia.org/
[27] Price MN, Dehal PS, Arkin AP. FastTree 2-approximately maximum-likelihood trees for large alignments[J]. PLoS One, 2010, 5(3): e9490
[28] Huson DH, Mitra S, Ruscheweyh HJ, et al. Integrative analysis of environmental sequences using MEGAN4[J]. Genome Research, 2011, 21(9): 1552-1560
[29] Biswas A, Gagnon JN, Brouns SJ, et al. CRISPRTarget: bioinformatic prediction and analysis of crRNA targets[J]. RNA Biology, 2013, 10(5): 817-827
[30] Brodt A, Lurie-Weinberger MN, Gophna U. CRISPR loci reveal networks of gene exchange in archaea[J]. Biology Direct, 2011, 6: 65

征 稿 简 则

1 刊物简介与栏目设置

《微生物学通报》是由中国科学院微生物研究所和中国微生物学会主办的，以微生物学应用基础研究及技术创新与应用为主的综合性学术期刊。刊登内容包括：工业微生物学、海洋微生物学、环境微生物学、基础微生物学、农业微生物学、食品微生物学、兽医微生物学、药物微生物学、医学微生物学、病毒学、酶工程、发酵工程、代谢工程等领域的最新研究成果，产业化新技术和新进展，以及微生物学教学研究和改革等。设置的栏目有：研究报告、专论与综述、生物实验室、高校教改纵横、显微世界、专栏、书讯、会讯等。

2 投稿方式

投稿时请登陆我刊主页 http://journals.im.ac.cn/wswxtbcn，点击作者投稿区，第一次投稿请先注册，获得用户名和密码，然后依照提示提交稿件，详见主页"投稿须知"。

3 写作要求

来稿要求论点明确，数据可靠，简明通顺，重点突出。

3.1 参考文献

参考文献按文内引用的先后顺序排序编码，未公开发表的资料请勿引用。我刊参考文献需要注明著者(文献作者不超过 3 人时全部列出，多于 3 人时列出前 3 人，后加"等"或"et al."，作者姓前、名后，名字之间用逗号隔开)、文献名、刊名、年卷期及页码。国外期刊名必须写完整，不用缩写，不用斜体。参考文献数量不限。

参考文献格式举例:

[1] Marcella C, Claudia E, Pier GR, et al. Oxidation of cystine to cysteic acid in proteins by peroyacids as monitored by immobilized pH gradients[J]. Electrophoresis, 1991, 12(5): 376-377
[2] Wang BJ, Liu SJ. Perspectives on the cultivability of environmental microorganisms[J]. Microbiology China, 2013, 40(1): 6-17 (in Chinese)
王保军, 刘双江. 环境微生物培养新技术的研究进展[J]. 微生物学通报, 2013, 40(1): 6-17
[3] Shen T, Wang JY. Biochemistry[M]. Beijing: Higher Education Press, 1990: 87 (in Chinese)
沈同, 王镜岩. 生物化学[M]. 北京: 高等教育出版社, 1990: 87
[4] Liu X. Diversity and temporal-spatial variability of sediment bacterial communities in Jiaozhou Bay[D]. Qingdao: Doctoral Dissertation of Institute of Oceanology, Chinese Academy of Sciences, 2010 (in Chinese)
刘欣. 胶州湾沉积物细菌多样性及菌群时空分布规律[D]. 青岛: 中国科学院海洋研究所博士学位论文, 2010

3.2 脚注(正文首页下方)

Foundation item:
*Corresponding author: Tel: ; Fax: ; E-mail:
Received: January 01, 20xx; Accepted: March 01, 20xx; Published online (www.cnki.net): March 31, 20xx
基金项目： 基金项目(No.)
*通讯作者：Tel： ；Fax： ；E-mail:
收稿日期：20xx-00-00；接受日期：20xx-00-00；优先数字出版日期(www.cnki.net)：20xx-00-00

(下转 p.2373)