

## 规律成簇的间隔短回文重复：结构，功能与应用

崔玉军, 李艳君, 颜焱锋, 杨瑞馥\*

(病原微生物生物安全国家重点实验室, 军事医学科学院微生物流行病学研究所, 北京 100071)

**摘要:** 规律成簇的间隔短回文重复 (Clustered regularly interspaced short palindromic repeats, CRISPRs) 是一类广泛分布于细菌和古菌基因组中的重复结构。最近研究表明, CRISPR 与一系列相关蛋白、前导序列一起, 为原核生物提供对噬菌体等外源基因的获得性免疫能力, 其作用机制可能与真核生物的 RNA 干扰过程类似。作为基因组中高度可变的区域, CRISPR 非常适合成为研究细菌种内分型和微进化的分子靶标。本文综述了 CRISPR 系统的结构、功能及其应用概况, 并对 CRISPR 研究的前景进行了展望。

**关键词:** 规律成簇的间隔短回文重复; CRISPR 相关基因/蛋白; 噬菌体; RNA 干扰

中图分类号: Q933 文献标识码: A 文章编号: 0001-6209 (2008) 11-1549-07

漫长的进化历史中, 细菌与噬菌体病毒从来没有停止过斗争。在生存压力下, 细菌进化出了多种噬菌体抵抗机制, 如吸附抑制或者无效感染 (abortive infection) 等<sup>[1]</sup>。规律成簇的间隔短回文重复 (Clustered regularly interspaced short palindromic repeats, CRISPRs) 是最近发现的针对噬菌体等外源基因的获得性免疫系统<sup>[2-4]</sup>。该结构广泛分布于细菌和古菌, 根据最近更新的 CRISPR 数据库<sup>[5, 6]</sup>, 在已完成全基因组测序的几乎所有古菌 (44/50) 和超过三分之一的细菌中 (250/616) 鉴定出 CRISPR 位点。

CRISPR 的结构特点是包括一段由不连续的同向重复 (direct repeat, DR), 间隔着大小与之相近的非重复间区序列 (spacer) 组成的区域。CRISPR 位点、前导序列 (leader sequence, LS) 以及一系列 CAS (CRISPR-associated) 蛋白<sup>[2,7]</sup>一起被称作 CRISPR 系统 (CRISPR associated system, CASS)<sup>[8]</sup> (图 1)。

CRISPR 重复结构最早在 1987 年大肠杆菌 (*Escherichia coli*) K12 的 iap 基因侧翼序列中被发现<sup>[9]</sup>。随后, 在其他物种也发现了 CRISPR, 并在当时给予了各种不同的命名: DVR (direct variable re-

peat), TREP (tandem repeat), LTRR (Long tandemly repeated repetitive), LCTR (long clusters of tandem repeats), SPIDR (spacers interspersed direct repeats)<sup>[10-14]</sup>。Mojica 等在 2000 年提出把 CRISPR 作为重复序列家族的一类成员, 并将其命名为 Short Regularly Spaced Repeats (SRSRs)<sup>[7]</sup>。2002 年, Mojica 和 Jansen 等最终确定了将该类重复序列命名为 CRISPR, 以反映该类重复的结构特点<sup>[2]</sup>。

对于同一个物种, CRISPR 位点的 DR 区域绝大多数非常保守<sup>[2,15]</sup>; 而间区序列则具有很高的多态性, 即使在同一物种的不同菌株间, 间区序列也往往不同, 这种性质使得 CRISPR 位点可以作为细菌分型鉴定的一个高分辨率分子靶标<sup>[16,17]</sup>。基于 CRISPR 位点多态性建立的间区寡核苷酸分型 (Spoligotyping) 技术已经广泛应用于结核分支杆菌 (*Mycobacterium tuberculosis*)<sup>[18]</sup>。最近有研究者结合白喉棒状杆菌 (*Corynebacterium diphtheriae*) 基因组中两个 CRISPR 位点的多态性, 成功对其进行了鉴定和基因分型<sup>[19]</sup>。

曾经有一些研究通过生物信息学手段预测过 CRISPR 系统的功能。1995 年, Mojica 等根据在沃氏

基金项目: 国家杰出青年科学基金(30525025)

\*通讯作者。Tel: +86-10-66948595; Fax: +86-10-63815689; E-mail: ruifuyang@gmail.com

作者简介: 崔玉军(1978-), 男, 山东东营人, 博士研究生, 研究方向为细菌基因组多态性与微进化。E-mail: yujuncui@sina.com

收稿日期: 2008-07-09; 修回日期: 2008-08-27



源基因组,以及同一细菌基因组的其它部分具有高度同源性<sup>[16,17,21]</sup>,这些同源部分被称作前间区序列(proto-spacer)。嗜热链球菌的 proto-spacer 下游间隔 1~2 个碱基的位置,存在长度为 4 个碱基的保守序列<sup>[24,25]</sup>。这类保守序列可能在间区序列的获取过程中起识别作用。

### 1.3 前导序列

过去的研究中 LS 曾被认为是一种长重复序列(long repeats, LRs)<sup>[27]</sup>。LS 在物种内相对保守,但在物种间存在差异。古菌中 LS 长度在 132~564 碱基之间,其长度和古菌最佳培养温度存在正相关<sup>[22]</sup>。大多数情况下,LS 位于 CRISPR 位点的 5'端,直接与 DR 相连。新“DR-spacer”基序的增加总是发生在 LS 和其相邻的基序之间<sup>[3,17,22,24,25,28]</sup>,这意味着 LS 可能在间区序列增加过程中作为识别位点。有研究认为 LS 处于 CRISPR 位点转录的起始端,并且有可能作为 CRISPR 位点的启动子<sup>[29]</sup>。

### 1.4 CAS 蛋白(基因)

大多数 CRISPR 位点侧翼序列存在一系列与 CRISPR 位点密切相关的基因,称为 CAS 基因,相应的蛋白为 CAS 蛋白<sup>[2]</sup>。这些 CAS 基因/蛋白对于 CASS 的功能实现可能起到重要作用。两个研究组分别对 CAS 蛋白进行了鉴定和分类:Haft 等<sup>[30]</sup>将 CAS 蛋白分为核心蛋白(core cas)、亚型蛋白(subtype cas)和重复相关可疑蛋白(Repeat-Associated Mysterious Protein, RAMP),每个亚型包含 2~6 个不同的 CAS 蛋白,总共鉴定出 45 类 CAS 蛋白家族;Makarova 等<sup>[8]</sup>在此基础上将蛋白家族进行了整合,并增添了一些新发现的 CAS 蛋白,将 CAS 蛋白分成 23 类。

随着 CASS 鉴定方法的改进和功能研究的发展,将来会有更多的 CAS 蛋白/基因被发现。但是,目前几乎所有的 CAS 蛋白功能都是未知的,另外 CAS 蛋白的序列保守性相对较低,因此对新发现的 CAS 蛋白进行分类是比较困难的问题。使用 TIGRFAMs 数据库的隐马氏模型(Hidden Markov Models, HMM)对蛋白序列进行鉴定是目前判断 CAS 蛋白种类的一个比较合理的方法(网址见表 1)。

### 1.5 CASS 可以在细菌间进行水平转移

Makarova 等在 2002 年发现,在亲缘关系很远的原核生物之间(如古菌与细菌)存在序列保守的 CAS 蛋白<sup>[20]</sup>,认为这是 CAS 蛋白在细菌中进行水平基因转移造成的。使用 4 个 CAS 基因序列分别构建的系

统发育树都支持这个结论。Godde 等把来自 52 个菌株的 CAS1-4 蛋白的氨基酸序列分别合并为平均长度 1413 氨基酸的总序列,使用合并后的序列构建系统发育树<sup>[31]</sup>。结果观察到部分亲缘关系很远的物种在同一个分支上存在。同一研究中,他们在大于 40 kb 的质粒上发现了 CASS,为 CASS 系统的水平转移提供了又一项证据。对浸矿菌 II 群进行的群体基因组分析结果显示,在 CRISPR 位点附近长约 8 kb 的一段侧翼序列存在于两群来自不同位置且亲缘关系较远的菌群,证明 CRISPR 位点的水平转移可能伴随着其侧翼序列的转移<sup>[28]</sup>。

## 2 CRISPR 系统的功能

### 2.1 为细胞提供对噬菌体侵染的抵抗能力

由于部分间区序列与染色体外基因序列具有同源性,几个研究组分别提出了 CRISPR 可能属于原核生物抵抗噬菌体免疫系统的假设<sup>[8,16,17,21,22]</sup>。2005 年 Bolotin 等在嗜热链球菌中观察到,CRISPR 位点间区序列的数目与菌株对噬菌体的敏感性成负相关<sup>[21]</sup>。Barrangou 等<sup>[3]</sup>选择分离自乳酪样品的嗜热链球菌噬菌体敏感株和两株噬菌体为研究对象,使用两种噬菌体分别感染敏感株后,获得了 9 株抗噬菌体突变株。对 CRISPR 位点的测序证明,每个突变株都在 LS 一端获得了 1~4 个新的间区序列,这些间区序列均来自感染敏感株的噬菌体基因组。将能够介导对噬菌体抵抗能力的间区序列通过人工插入敏感株的相应位置,敏感株获得对噬菌体的抗性;消除新插入的间区序列,菌株又恢复对噬菌体的敏感性。上述结果最终证实了 CASS 作为原核生物的一种获得性免疫系统的功能。在完成测序的硫磺矿硫化叶菌(*Sulfolobus solfataricus* P2)基因组中发现在原噬菌体中也存在 CRISPR 位点<sup>[10]</sup>。该 CRISPR 系统可能通过调节宿主细菌的基因表达,使噬菌体更容易在与细菌的斗争中存活下来。此外,噬菌体能够以快速突变或重组等基因组进化机制,逃避宿主通过在 CRISPR 位点增加间区序列所获取的免疫能力<sup>[32]</sup>。这些现象体现了细菌与噬菌体之间相互形成强大的选择性压力,导致为适应生存环境而进行的基因组快速进化。

### 2.2 CASS 可能以类 RNA 干扰的机制实现其功能

部分实验证据以及对 CAS 蛋白功能的预测表明,CASS 可能以一种类似于真核生物中 RNA 干扰的形式实现其功能。Tang 等发现在闪烁古生球菌(*Archaeoglobus fulgidus*)和硫磺矿硫化叶菌中

CRISPR 位点可以转录为 RNA 前体, 随后被剪切为一个 DR-spacer 长度的小 RNA 单位<sup>[29]</sup>。剪切位点可能位于 DR 的中间部分, DR 的回文部分通过氢键形成互补结构(颈), 中间的间区序列则形成一个环。这些小 RNA 单位与 CAS 蛋白可能结合在一起共同作用到目标 mRNA 上, 形成 RNA 诱导沉默复合体, 导致 mRNA 的降解, 从而沉默相应的基因。另一种可能的沉默方式考虑到 CASS 中存在聚合酶活性的蛋白, 因此剪切后的小 RNA 单位可以结合到目标 mRNA 上作为引物, 在聚合酶作用下进行延伸, 形成双链产物; 随后再被剪切、沉默。

### 2.3 CRISPR 是原核生物与噬菌体斗争的记录器

综合 CASS 抵抗噬菌体的功能和原核生物基因组的高效(低冗余)特性, 推测 CRISPR 位点有如下性质: 1) 外源基因组序列能够被迅速识别并作为新的间区序列插入到 CRISPR 位点中, 以形成对相应物种(基因)的抵抗能力; 2) 一个或者多个已有的间区序列可以发生随机丢失, 而对物种生存有用的间区序列在自然选择的压力下保存下来。CRISPR 位点的基因序列记录了在原核生物进化中遇到的各类外源基因(主要是噬菌体基因)片段。因而, CRISPR 可以看作是原核生物与噬菌体斗争的记录器。新闻区的

插入是有极性的(总发生在 LS 一侧), 因此间区序列的排列可以反映一个物种的进化方向; 根据 CRISPR 位点间区序列排列的差异, 能够判断不同菌株间的进化关系。

## 3 CRISPR 研究的生物信息学资源

目前已经有多种生物信息学工具可用于 CRISPR 位点搜索及分析, 其中较为快速、高效的包括以下 3 种: PILER-CR<sup>[33]</sup>, CRISPRFinder<sup>[34]</sup> 和 CRT<sup>[35]</sup>。PILER-CR 使用 C++ 编写, 作者提供了源程序下载, 用户需要自行编译后才能使用。CRT 使用 JAVA 编写, 运行前需要用户安装 Java 运行环境(Java Runtime Environment, JRE), 便于在分析单个基因组或者一段 FASTA 格式的序列时使用。以上两个软件可以在本地机上运行。CRISPRFinder 是使用 perl 语言开发的基于网络的软件, 与已完成测序的原核生物的 CRISPR 信息一起存放于 CRISPR DB 网站中<sup>[5, 6, 34, 36]</sup>。其优点是在鉴定间区序列数目少于 3 个的 CRISPR 位点时比前两者准确; 同时, 利用该网站提供的 CRISPRionary、BLAST CRISPRs 等工具, 可以方便的分析不同菌株之间 CRISPR 位点的间区序列排列及其差异, 从而利于进一步的分型、进化分析。其他 CRISPR 研究可能用到的工具(网站)作用和网址见表 1。

表 1 CRISPR 研究相关的生物信息学资源  
Table 1 Bioinformatics resources for CRISPR research

Resources	Description	URL
PILER-CR CRISPRFinder CRT	Tools for detecting CRISPR locus from genomic sequences.	<a href="http://www.drive5.com/pilercr">http://www.drive5.com/pilercr</a> <a href="http://crispr.u-psud.fr/Server/CRISPRfinder.php">http://crispr.u-psud.fr/Server/CRISPRfinder.php</a> <a href="http://www.room220.com/crt">http://www.room220.com/crt</a>
CRISPR DB	Database of CRISPR loci in sequenced prokaryotic genomes, containing series of analysis tools for CRISPR locus detection, comparison and alignment.	<a href="http://crispr.u-psud.fr/crispr/">http://crispr.u-psud.fr/crispr/</a> CRISPRHomePage.php
TIGR CMR	Website for browsing flanking sequences, repeat region and CAS genes of CRISPR loci in sequenced prokaryotic genomes.	<a href="http://cmr.tigr.org/tigr-scripts/CMR/shared/MakeFrontPages.cgi?page=genome_property">http://cmr.tigr.org/tigr-scripts/CMR/shared/MakeFrontPages.cgi?page=genome_property</a>
TIGR HMM search	Website for identifying and classifying CAS proteins by HMM.	<a href="http://tigrblast.tigr.org/web-hmm/">http://tigrblast.tigr.org/web-hmm/</a>

## 4 CRISPR 的应用

### 4.1 间区寡核苷酸分型

以 CRISPR 位点多态性为基础的间区寡核苷酸分型已经成为结核分支杆菌的标准分型鉴定方法。该菌的 CRISPR 位点失去获得新闻区序列的能力, 其多样性表现为已有 43 个间区序列的随机删除。使用生物素标记位于 DR 区域、方向相反的一对引物, 对菌

株核酸进行扩增, 然后将扩增产物与结合在尼龙膜上的间区序列探针进行杂交, X 光照射显影, 观察杂交图谱并使用软件进行分析。与传统的 IS6110 限制性片段长度多态性指纹图相比, 该方法快速、简便, 结果稳定性好, 但分辨率略低<sup>[18]</sup>。国际上已经建立了包含近 40000 株菌、2000 种型别模式的结核分支杆菌数据库, 可以通过接入互联网使用<sup>[37]</sup>。

#### 4.2 利用 CRISPR 位点多态性进行细菌分型

由于其多态性极高, CRISPR 位点可作为细菌分型的理想靶标。使用 CRISPR 位点序列分析的方法对 184 株空肠弯曲杆菌 (*Campylobacter jejuni*) 进行分型<sup>[38]</sup>, 证明 CRISPR 位点的分辨率与 AFLP 和 MLST 相当。澳大利亚的一个研究小组使用高分辨率 DNA 熔解曲线 (high-resolution DNA melt curve, HRM) 的方法对大肠弯曲杆菌 (*Campylobacter coli*) 的 CRISPR 位点进行了分析<sup>[39]</sup>, 结果证明该方法的分型能力不低于 PFGE。鼠疫菌基因组中包含 3 个 CRISPR 位点 (YPa, YPb 和 YPc), 我们在中国、前苏联和外蒙的 26 个鼠疫自然疫源地中, 挑选了 125 株代表性鼠疫菌, 对其 3 个 CRISPR 位点进行了分析。综合 3 个位点的间区序列排列, 可以将鼠疫菌分成 12 个 CRISPR 群, 各 CRISPR 群的分布与疫源地联系紧密。该结果进一步证明 CRISPR 位点在分型中的作用, 并对鼠疫菌的微生物法医学溯源提供了新思路<sup>[26]</sup>。

#### 4.3 细菌进化分析

基于间区序列极性化插入及随机化丢失的特点, Pourcel 等根据 YPa 位点的间区序列排列情况, 推测出鼠疫菌祖先菌株 CRISPR 位点序列和该菌的进化场景<sup>[17]</sup>。在我们的研究中, 挑选了分布和代表性更为广泛的菌株, 并对所有 3 个 CRISPR 位点进行了分析, 根据 CRISPR 位点进化规律 构建出鼠疫菌进化模型。该模型判断鼠疫菌的传播主路线环绕着塔克拉玛干和准葛尔地区, 其余相对年轻的鼠疫菌分支从该主路线上衍生出来, 并由人类活动或其他偶然事件传播到世界各地<sup>[26]</sup>。

## 5 前景展望

对 CASS 的研究还处于起始阶段, 尤其是实验方面的报道还很少。有几方面关键问题尚待解决: 间区序列的获取和丢失机制; CAS 蛋白、LS 与 CRISPR 重复区域之间的相互作用; 以及不同 CAS 蛋白的功能。由于 CRISPR 可能是按照类 RNA 干扰的模式行使功能, 对 CASS 的研究将有利于进一步理解真核生物 RNA 干扰现象。鉴于在已测序原核生物中的广泛分布, CASS 可能是最古老的针对噬菌体的原核生物免疫系统。在漫长的进化过程中, 噬菌体可能也发展出逃避或抑制 CASS 作用的机制。该方面的研究将有利于揭示噬菌体与细菌之间互相斗争的关系。

目前还没有 CASS 对内源基因进行沉默的相关

报道。如果 CASS 确实是以类 RNA 干扰的模式去降解 mRNA, 原核生物中存在的与其基因组序列同源的间区序列转录后将可以作为小 RNA, 造成转录自相应内源基因的 mRNA 降解。这个假设的证实将引发微生物学研究的革命性进展: 在研究某种或某几种基因功能时, 不必再费很大力量去敲除这些基因, 只要选取其特异性序列作为新的间区序列, 与 DR 一起转移到 CRISPR 位点中, 就能沉默相应基因。同样, 也可以通过在携带 CRISPR 位点的大质粒上插入目的基因间区序列, 然后将质粒转导入待研究的细菌中, 从而关闭相应基因表达。

来自噬菌体的间区序列能够在一定程度上反映细菌在进化过程中经历过的生态环境, 而间区序列的极性化插入特性赋予了 CRISPR 位点揭示进化方向的功能。对 CRISPR 位点的研究为我们提供了一个方便的研究进化的工具。CRISPR 位点的高度变异性使其在不同细菌之间 (甚至亲缘关系非常近的物种间), 间区序列几乎都不相同。但是在同一种细菌的不同菌株间, 这种高度变异性则为我们提供了一个合适的鉴定、分型靶标。结核分支杆菌的间区寡核苷酸分型就是一个成功应用的例子。

CASS 提供对噬菌体的特异性抵抗能力使其在生物工业上尤为有用。许多工业化生产如乳业和制酒业等都要依赖细菌, 噬菌体的破坏对于这些产业造成了巨大损失。将针对特定噬菌体的间区序列转移入这些细菌, 能够有效的解决这个问题。已经有研究者申请了此方面的专利。

总之, 对 CASS 功能、机制的深入探讨将在多个方面引发微生物学研究领域的革命性进展。

## 参 考 文 献

- [1] Sturino JM, Klaenhammer TR. Engineered bacteriophage-defence systems in bioprocessing. *Nat Rev Microbiol*, 2006, 4(5): 395-404.
- [2] Jansen R, Embden JD, Gastra W, et al. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*, 2002, 43 (6): 1565-1575.
- [3] Barrangou R, Fremaux C, Deveau H, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 2007, 315 (5819): 1709-1712.
- [4] Sorek R, Kunin V, Hugenholtz P. CRISPR - a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol*, 2008, 6(3): 181-1816.
- [5] Grissa I, Bouchon P, Pourcel C, et al. On-line resources for bacterial micro-evolution studies using MLVA or CRISPR typing.

- Biochimie*, 2008, 90(4): 660–668.
- [6] Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics*, 2007, 8(1): 172.
- [7] Mojica FJ, Diez-Villasenor C, Soria E, *et al.* Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol*, 2000, 36(1): 244–246.
- [8] 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. *Biol Direct*, 2006, 3: 1–7.
- [9] Ishino Y, Shinagawa H, Makino K, *et al.* Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol*, 1987, 169 (12): 5429–5433.
- [10] She Q, Singh RK, Confalonieri F, *et al.* The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci U S A*, 2001, 98 (14): 7835–7840.
- [11] van Embden JD, van Gorkom T, Kremer K, *et al.* Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. *J Bacteriol*, 2000, 182 (9): 2393–2401.
- [12] Masepohl B, Gorlitz K, Bohme H. Long tandemly repeated repetitive (LTR) sequences in the filamentous cyanobacterium *Anabaena* sp. PCC 7120. *Biochim Biophys Acta*, 1996, 1307(1): 26–30.
- [13] Mojica FJ, Ferrer C, Juez G, *et al.* Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Mol Microbiol*, 1995, 17(1): 85–93.
- [14] Groenen PM, Bunschoten AE, van Soolingen D, *et al.* Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol Microbiol*, 1993, 10(5): 1057–1065.
- [15] Kunin V, Sorek R, Hugenholtz P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol*, 2007, 8(4): R61.
- [16] Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, *et al.* Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol*, 2005, 60(2): 174–182.
- [17] Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, 2005, 151(Pt 3): 653–663.
- [18] Kamerbeek J, Schouls L, Kolk A, *et al.* Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol*, 1997, 35(4): 907–914.
- [19] Mokrousov I, Limeschenko E, Vyazovaya A, *et al.* *Corynebacterium diphtheriae* spoligotyping based on combined use of two CRISPR loci. *Biotechnol J*, 2007, 2(7): 901–906.
- [20] Makarova KS, Aravind L, Grishin NV, *et al.* A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res*, 2002, 30(2): 482–496.
- [21] Bolotin A, Quinquis B, Sorokin A, *et al.* Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, 2005, 151(Pt 8): 2551–2561.
- [22] Lillestol RK, Redder P, Garrett RA, *et al.* A putative viral defence mechanism in archaeal cells. *Archaea*, 2006, 2 (1): 59–72.
- [23] Prangishvili D, Forterre P, Garrett RA. Viruses of the Archaea: a unifying view. *Nat Rev Microbiol*, 2006, 4 (11): 837–848.
- [24] Deveau H, Barrangou R, Garneau JE, *et al.* Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol*, 2008, 190(4): 1390–1400.
- [25] Horvath P, Romero DA, Coute-Monvoisin AC, *et al.* Diversity, activity and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol*, 2008, 190(4): 1401–1412.
- [26] Cui Y, Li Y, Gorge O, *et al.* Insight into Microevolution of *Yersinia pestis* by Clustered Regularly Interspaced Short Palindromic Repeats. *PLoS ONE*, 2008, 3(7): e2652.
- [27] Bult CJ, White O, Olsen GJ, *et al.* Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science*, 1996, 273(5278): 1058–1073.
- [28] Tyson GW, Banfield JF. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ Microbiol*, 2008, 10(1): 200–207.
- [29] Tang TH, Bachelier JP, Rozhdestvensky T, *et al.* Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc Natl Acad Sci U S A*, 2002, 99 (11): 7536–7541.
- [30] Haft DH, Selengut J, Mongodin EF, *et al.* A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol*, 2005, 1 (6): e60.
- [31] Godde JS, Bickerton A. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J Mol Evol*, 2006, 62(6): 718–729.
- [32] Andersson AF, Banfield JF. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science*, 2008, 320(5879): 1047–1050.
- [33] Edgar RC. PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics*, 2007, 20: 8–18.
- [34] Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res*, 2007, 35(Web Server issue): W52–7.
- [35] Bland C, Ramsey TL, Sabree F, *et al.* CRISPR Recognition Tool

- (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics*, 2007, 8 (1): 209.
- [36] Grissa I, Vergnaud G, Pourcel C. CRISPRcompar: a website to compare clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res*, 2008, 36(Web Server issue): W145–8.
- [37] Brudey K, Driscoll JR, Rigouts L, et al. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol*, 2006, 3: 6–23.
- [38] Schouls LM, Reulen S, Duim B, et al. Comparative genotyping of *Campylobacter jejuni* by amplified fragment length polymorphism, multilocus sequence typing, and short repeat sequencing: strain diversity, host range, and recombination. *J Clin Microbiol*, 2003, 41 (1): 15–26.
- [39] Price EP, Smith H, Huygens F, et al. High-resolution DNA melt curve analysis of the clustered, regularly interspaced short-palindromic-repeat locus of *Campylobacter jejuni*. *Appl Environ Microbiol*, 2007, 73 (10): 3431–3436.

## Clustered regularly interspaced short palindromic repeats: structure, function and application — A review

Yujun Cui, Yanjun Li, Yanfeng Yan, Ruifu Yang\*

(State Key laboratory of Pathogen and Biosecurity, Institute of Microbiology and Epidemiology, Academy of Military Medical Sciences, Beijing 100071, China)

**Abstract:** CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats), the basis of spoligotyping technology, can provide prokaryotes with heritable adaptive immunity against phages' invasion. Studies on CRISPR loci and their associated elements, including various CAS (CRISPR-associated) proteins and leader sequences, are still in its infant period. We introduce the brief history<sup>1</sup>, structure, function, bioinformatics research and application of this amazing immunity system in prokaryotic organism for inspiring more scientists to find their interest in this developing topic.

**Keywords:** CRISPR; CASS; CAS; phage; evolution; RNAi

Supported by the National Science Fund of China for Distinguished Young Scholars (30525025)

\*Corresponding author. Tel: +86-10-66948595; Fax: +86-10-63815689; E-mail: ruifuyang@gmail.com

Received: 9 July 2008/ Revised: 27 August 2008

### 《微生物学报》答作者问——关于投稿

问: 投稿时都需要哪些手续? 是否还需要纸稿?

答: 从 2006 年起, 本刊开始采用“稿件远程处理系统”。投稿时需要提供:

- (1) 论文研究内容所属单位的介绍信(请注意: 在此强调的是研究内容所属单位, 通常是第一单位), 介绍信主要应证明该文的作者署名无误, 未一稿两投及不涉及保密问题。介绍信模板可从本刊主页“下载专区”或“远程投稿时”下载。
- (2) 在接到经过编辑部内审后 E-mail 发出的“稿件受理通知”后, 需要及时补寄纸样的 1 份稿件和介绍信, 并缴纳 100 元稿件受理费。

问: 审稿费需邮局汇款还是转帐?

答: 邮局汇款! 中科院微生物所共有 4 个期刊编辑部, 因此提醒您在办理汇款时一定要注意以下几点, 否则在登记汇款、办理发票时会造成混乱! 编辑部在收到汇款之后, 将以挂号信形式及时寄回发票。

- (1) 切忌在邮寄来的纸样材料中加入 100 元现金!
- (2) 在收款人一栏填写“微生物学报编辑部”;
- (3) 在备注栏中注明“稿件编号”+“第一作者姓名”;
- (4) 通过邮局汇 100 元审稿费, 汇款后请登陆本刊网站, 填写“汇款时间”、“发票单位”和收“发票地址”等信息。编辑部会在收到后及时登记“收款时间”和“寄发票时间”, 作者可随时查询不必打电话来询问。