



# 创伤弧菌 TF3 全基因组包含的 CRISPR-Cas 系统及 MGIVvuTF3 基因岛分析

云龙<sup>1,2</sup>, 罗鹏<sup>1\*</sup>, 田雨顺<sup>1,2</sup>, 丁雄祺<sup>1,2</sup>, 胡超群<sup>1</sup>

<sup>1</sup> 中国科学院南海海洋研究所, 中国科学院热带海洋生物资源与生态重点实验室, 广东省应用海洋生物学重点实验室, 广东 广州 510301

<sup>2</sup> 中国科学院大学, 北京 101408

**摘要:**【目的】分析创伤弧菌(*Vibrio vulnificus*)全基因组框架序列, 挖掘感兴趣的遗传位点, 探索其是否具有 CRISPR 系统及其特征。【方法】在病虾体内分离获得了一株创伤弧菌 TF3, 通过 Illumina Miseq 测序得到基因组框架序列。经注释分析发现其基因组中存在一个 CRISPR-Cas 系统, 命名为 CRISPR-Cas<sub>TF3</sub>, 进一步分析发现 CRISPR-Cas<sub>TF3</sub> 位于一个基因岛上, 将该基因岛命名为 MGIVvuTF3。对 CRISPR-Cas<sub>TF3</sub> 及 MGIVvuTF3 的特征和来源进行了分析。【结果】CRISPR-Cas<sub>TF3</sub> 属于与大肠杆菌类似的 I-E 型 CRISPR-Cas 系统, CRISPR-Cas<sub>TF3</sub> 包括 8 个 *cas* 基因, 其排列为 *cas3-cas8e-cse2-cas6-cas7-cas5-cas1-cas2*; 具有 50 个重复序列, 每二个重复序列间为一个 Spacer 序列。MGIVvuTF3 具有 attL 和 attR 序列, 含有位点特异性整合、剪切、转移相关的基因。MGIVvuTF3 与霍乱弧菌 O395 基因组中的一个基因岛 MGIVch0395 具较高相似性, 二者最显著的差别在于 Spacer 序列完全不同, 以及各有几个非保守的预测基因。【结论】MGIVvuTF3 及 CRISPR-Cas<sub>TF3</sub> 极有可能通过基因水平转移获得, 并且 CRISPR-Cas<sub>TF3</sub> 系统可以借助 MGIVvuTF3 实现水平转移。

**关键词:** CRISPR, *cas* 基因, 基因岛, 创伤弧菌

创伤弧菌(*Vibrio vulnificus*)是海洋环境中常见的革兰氏阴性弧菌, 它与霍乱弧菌(*V. cholerae*)以及副溶血弧菌(*V. parahaemolyticus*)并称人类三大致病性弧菌<sup>[1]</sup>, 创伤弧菌同时也是多种水产动物的致病菌<sup>[2]</sup>。以往的研究表明, 创伤弧菌具有

高度的遗传多样性, 并且具备多种多样的水平转移遗传元件, 它们赋予宿主菌多样性的环境适应性功能<sup>[2-3]</sup>。CRISPR (Clustered regularly interspaced short palindromic repeats)系统是原核生物抵制噬菌体和质粒等异源核酸入侵、适应环境的功能元

基金项目: 国家自然科学基金(31370149); 广东省渔业科技攻关项目(20170228-9-3)

\*通信作者。E-mail: luopengli@163.com

收稿日期: 2017-08-15; 修回日期: 2017-11-21; 网络出版日期: 2018-01-09

件之一<sup>[4]</sup>, 1987年, Ishino等首次在大肠杆菌中发现了CRISPR存在<sup>[5]</sup>。之后研究人员对其结构和功能进行了深入探究, 发现CRISPR系统主要由以Leader序列(L)、Repeat序列(R)和Spacer序列(S)构成的CRISPR位点和与之相互协同作用的*cas*基因(表达Cas蛋白)构成<sup>[6-7]</sup>。L序列位于CRISPR位点的上游, 是富含AT的不保守序列, 具有RNA聚合酶的识别位点, 与CRISPR位点的转录有关<sup>[8]</sup>; 在多数情况下, 整个CRISPR排列首先转录形成pre-CRISPR RNA, 随后加工成为成熟的crRNAs, 成熟的crRNA与Cas蛋白形成核糖核蛋白效应体, 在crRNA指导下, Cas蛋白特异性锚定和清除异源的病毒(噬菌体)或质粒<sup>[9-12]</sup>; S序列是CRISPR位点的可变功能序列, 其原始来源与入侵的核酸有关, 代表对过去入侵异源核酸的“记忆”<sup>[9-10]</sup>。但是, 在已知的数据库中搜索发现创伤弧菌的全基因组中不存在典型的CRISPR结构, 而我们在患病凡纳滨对虾体内分离到的创伤弧菌TF3, 通过基因组序列分析发现其存在一个CRISPR-Cas系统, 并且CRISPR-Cas系统位于一个基因岛上, 本研究对此CRISPR-Cas系统及其所在的基因岛的基因组学特征进行了详细的分析。

## 1 材料和方法

### 1.1 菌株及全基因组测序

创伤弧菌TF3分离于广东湛江养殖的患病凡纳滨对虾肠道, 并保存于本实验室。采用华大基因Illumina Miseq测序平台进行了全基因组框架图测序。首先对提取的基因组进行机械剪切, 剪切后的DNA进行末端补平, 再通过3'端加碱基A, 使得DNA片段能与3'端带有T碱基的接头连接, 用电泳回收目的片段(400-500 bp)连接产物, 再使用

PCR扩增两端带有接头的DNA片段, 最后用检验合格的文库进行Illumina Miseq上机测序。

### 1.2 创伤弧菌TF3全基因组注释及基因岛分析

采用线上原核生物基因组注释在线工具([http://www.ncbi.nlm.nih.gov/genome/annotation\\_prok](http://www.ncbi.nlm.nih.gov/genome/annotation_prok))对创伤弧菌TF3全基因组进行注释。采用IslandViewer 4进行基因岛预测(<http://www.pathogenomics.sfu.ca/islandviewer/>)。

### 1.3 CRISPR-Cas系统及其所在基因岛基因组特征分析

采用CRISPRFinder<sup>[13]</sup>查找CRISPR位点, 采用Tandem Repeats Finder<sup>[14]</sup>查找CRISPR-Cas上下游attL和attR位点, 利用WebACT<sup>[15]</sup>进行基因组比对。

## 2 结果和分析

### 2.1 创伤弧菌TF3全基因组框架图测序

对创伤弧菌TF3全基因组框架图测序, 91.07%的测序碱基达到Q20值, 平均测序深度为350×, 共获得145条Scaffold序列, 总碱基数达到5077663 bp, 与典型的创伤弧菌株CECT 4999基因组总长(5094689 bp, NZ\_CP014636.1及NZ\_CP014637.1)十分接近, 表明本测序获得的基因组框架序列已经覆盖了基因组的绝大部分区域。创伤弧菌TF3的Scaffold序列最长Scaffold为456071 bp, 最短Scaffold为202 bp。创伤弧菌TF3基因组GC含量为46.6%, 与典型株CECT 4999的GC含量一致。TF3基因组至少包含82个基因岛, 已经探测出的基因岛长度介于4039-103724 bp, 基因岛占整个基因组大小的20.0%。TF3基因组预测编码4564个基因。根据同源蛋白簇(COG)聚类, 共有3492个

基因获得聚类,占比 76.5%。获得 COG 聚类的基因中,38.6%的基因与代谢相关,27.2%的基因与细胞进程与信号相关,16.7%的基因与遗传信息储存与加工相关。获得 COG 聚类的基因的具体功能类别及数量见图 1,由图可见,除了功能未知的基因,最大数量的基因分别被注释到氨基酸转运(E, 307 个)、代谢以及信号转导机制(T, 304 个)相关的同源蛋白簇中。

## 2.2 CRISPR-Cas 系统基因组特征

通过注释创伤弧菌 TF3 全基因组,获得其全基因组基因注释表,分析发现,Scaffold19 框架序列中存在多个 *cas* 基因。进一步采用 CRISPR Finder 查找 Scaffold19 的基因组序列,发现创伤弧菌 TF3 具有一个 CRISPR 位点,由此说明创伤弧菌 TF3 具备一个完整的 CRISPR-Cas 系统,将其命名为 CRISPR-Ca<sub>STF3</sub>。

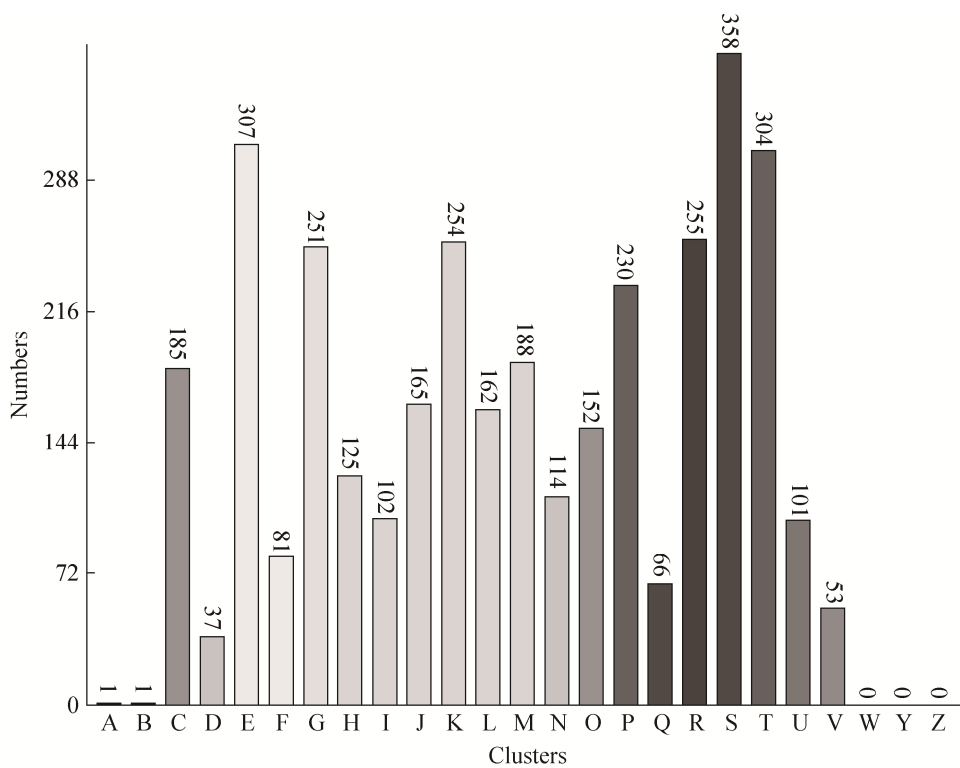


图 1. 同源蛋白簇(COG)聚类统计图

Figure 1. Statistical graph for clusters of orthologous groups of proteins. A: RNA processing and modification; B: Chromatin structure and dynamics; C: Energy production and conversion; D: Cell cycle control, cell division, chromosome partitioning; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; G: Carbohydrate transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; J: Translation, ribosomal structure and biogenesis; K: Transcription; L: Replication, recombination and repair; M: Cell wall/membrane/envelope biogenesis; N: Cell motility; O: Posttranslational modification, protein turnover, chaperones; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport and catabolism; R: General function prediction only; S: Function unknown; T: Signal transduction mechanisms; U: Intracellular trafficking, secretion, and vesicular transport; V: Defense mechanisms; W: Extracellular structures; Y: Nuclear structure; Z: Cytoskeleton.

*cas* 系列基因位于该 CRISPR 位点上游, 包括 8 个串列的 *cas* 基因, 其排列为 *cas3-cas8e-cse2-cas6-cas7-cas5-cas1-cas2* (图 2), *cas* 基因的数量、排列与 I-E 型大肠杆菌的 CRISPR-Cas 系统类似<sup>[16]</sup>, 其中 *cas3* 是此类型 CRISPR 的特征基因, 能编码单链 DNA 核酸酶, 具有切割 DNA 的作用<sup>[17]</sup>。*cas1* 和 *cas2* 存在于所有的 CRISPR 系统中, 且都具有核酸内切酶的活性, 它们参与获取新的间区序列<sup>[17]</sup>, *cas1* 能与反转录酶结合获取 RNA 形式的异源核酸加工成间区序列<sup>[18]</sup>。

CRISPR-Cas<sub>TF3</sub> 的 CRISPR 位点区由 R 序列与 S 序列交替形成。R 序列为 GTCTTCCCCACGCC CGTGGGGGTGTTTC, 大小为 28 bp, 有 50 个重复, 每 2 个 R 序列间为 1 个 S 序列, 因此共有 49 个 S 序列。S 序列长度不一, 其中 44 个长 33 bp, 2 个为 34 bp, 3 个长度为 32 bp。通过对 49 个 S 序列逐个进行 BLASTn 和 BLASTx 分析, 并没有找到与我们搜索的 S 序列同源性很高的序列。整个 CRISPR-Cas<sub>TF3</sub> 系统结构如图 2 所示, CRISPRFinder 数据库显示大多数的创伤弧菌以及其他弧菌菌种并不具有典型的 CRISPR 系统。

### 2.3 基因岛的发现与分析

为了探索 CRISPR-Cas<sub>TF3</sub> 在基因组中的插入位点及其邻近基因的特征, 进一步对 CRISPR-Cas<sub>TF3</sub> 系统所在的 Scaffold19 分析, 发现 CRISPR-Cas<sub>TF3</sub> 上游存在一个类似基因岛(或前噬菌体)的位点特异性整合酶基因 *int*, 通过 Tandem Repeats Finder 对 Scaffold19 是否具有重复的 att 序列进行查找, 发现 Scaffold19 具有 2 个正向重复的 att 序列, 长 16 bp, 其序列为 ATTAGTTCATGCCGTA, 2 个 att 序列所夹区域与采用 IslandViewer 4 预测的其中一个基因岛完全一致, 此区域还包括一些与基因岛

(或前噬菌体)相关的基因<sup>[19]</sup>(表 1), 表明其是一个基因岛, 将其命名为 MGIV<sub>vu</sub>TF3 (GenBank No. MF663762), 二端的 att 位点分别命名为 attL 和 attR。MGIV<sub>vu</sub>TF3 全长 18113 bp, 插入在 *fis* 基因的 3' 末端。MGIV<sub>vu</sub>TF3 包括了 *int* 基因、*orf1* 基因、*cas* 系列基因、*orf2* 基因、*orf3* 基因、*alpA* 基因、*alpB* 基因。*int* 负责基因岛的位点特异性整合; *orf1* 根据其结构域功能预测与基因岛的转移相关, 属于 Fic 蛋白家族; *orf3* 包含一个 Inovirus Gp2 结构, 可能与基因岛复制相关; *alpA* 表达后具有转录调节因子活性; *alpB* 表达后具有转录抑制作用。MGIV<sub>vu</sub>TF3 还包含一个 *orf2* 基因, 编码假想的蛋白, 功能未知。整个基因岛的结构如图 2 所示。

### 2.4 基因岛比较分析

采用 BLASTn 对 MGIV<sub>vu</sub>TF3 的相似序列进行了搜索, 发现仅霍乱弧菌 O395 基因组中存在一个与 MGIV<sub>vu</sub>TF3 类似的基因岛, 命名为 MGIV<sub>ch</sub>0395 (GenBank No: MF663762)。通过 Tandem Repeats Finder 找到 MGIV<sub>ch</sub>0395 具有与 MGIV<sub>vu</sub>TF3 相同的 attL 和 attR 序列。采用 RAST 注释该基因岛序列, 发现 MGIV<sub>ch</sub>0395 和 MGIV<sub>vu</sub>TF3 包含相同的 *cas* 基因及排列, 因此该基因岛也可能存在 CRISPR 位点, 通过 CRISPRFinder 搜索该基因岛, 发现该基因岛的 CRISPR 位点以 GTCTTCCCCACGCC GGTGGGGGTGTTTC 为 R 序列, 长度为 28 bp, 它与 MGIV<sub>vu</sub>TF3 的 R 序列仅有 2 个位点的差异。MGIV<sub>ch</sub>0395 的 CRISPR 系统有 39 个 S 序列, 但是其 S 序列与 MGIV<sub>vu</sub>TF3 序列完全不同。为了更直观地对 2 个基因岛进行比较, 用 WebACT 进行两基因岛序列对比, 其结果如图 3 所示。

由图可见, 这 2 个基因岛之间存在着广泛的相似性和同源性, 基因不存在重排, 表明它们具

有完全相同的结构和基因排列。MGIV*ch0395* 整合在霍乱弧菌 O395 基因组 *fis* 基因 3'末端，因此二者具有完全相同的整合位点。对 2 个基因岛相似

的基因逐个进行了 BLAST 比对，其比对结果见表 1。结果表明 2 个基因岛及其 CRISPR-Cas 虽然保守的基因高度相似，但并非完全一致。

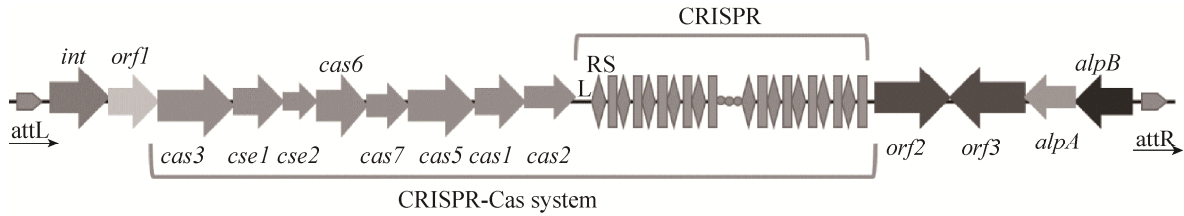


图 2 基因岛 MGIV $\nu$ TF3 结构示意图

Figure 2. Schematic diagrams of genomic island MGIV $\nu$ TF3.

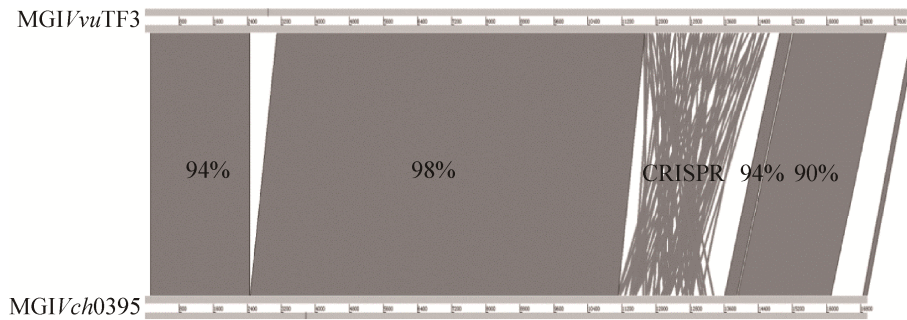


图 3. 基因岛序列对比图

Figure 3. Comparison diagram of genomic island sequence. Similar areas are shown in gray black and numbers indicate similarity.

表 1. MGIV $\nu$ TF3 与 MGIV $ch0395$  所含基因对比

Table 1. Comparison of the genes in MGIV $\nu$ TF3 and MGIV $ch0395$

| Genes        | Locus       | Coding for                 | Similarity with genes in MGIV $ch0395$ /% |
|--------------|-------------|----------------------------|---|
| <i>int</i>   | 179-1438    | Integrase                  | 94  |
| <i>orf1</i>  | 1444-2427   | Fic family protein         | 95  |
| <i>cas3</i>  | 3381-5993   | Cas3                       | 98  |
| <i>cas8e</i> | 5996-7519   | Cas8e                      | 99  |
| <i>cse2</i>  | 7533-8093   | Cse2                       | 99  |
| <i>cas6</i>  | 8086-8733   | Cas6                       | 99  |
| <i>cas7</i>  | 8917-9798   | Cas7                       | 98  |
| <i>cas5</i>  | 9802-10518  | Cas5                       | 95  |
| <i>cas1</i>  | 10521-11396 | Cas1                       | 98  |
| <i>cas2</i>  | 11422-11664 | Cas2                       | 99  |
| <i>orf2</i>  | 14892-15578 | Hypothetical protein       | 91  |
| <i>orf3</i>  | 16423-15773 | InovirusGp2 family protein | 92  |
| <i>alpA</i>  | 16797-16576 | Transcription regulator    | 92  |
| <i>alpB</i>  | 17030-16797 | Transcription repressor    | 93  |

### 3 讨论

CRISPR-Cas 赋予宿主菌对异源入侵 DNA (有些情况是 RNA) 的抵抗力<sup>[4,7,11]</sup>, 关于 CRISPR-Cas 的起源至今不为人知<sup>[20-21]</sup>, 但是研究者已经指出 CRISPR-Cas 很可能通过基因水平转移获得<sup>[2-23]</sup>。本研究中发现 CRISPR-Cas<sub>TF3</sub> 位于基因岛 MGIV<sub>vu</sub>TF3 上。MGIV<sub>vu</sub>TF3 具有 att 序列以及一系列与基因组整合及剪切相关的基因, 并且仅在创伤弧菌 TF3 中发现了 MGIV<sub>vu</sub>TF3 的存在, 与之高度相似的基因岛 MGIV<sub>ch</sub>0395 也仅出现在霍乱弧菌 O395 基因组中, 这些特征均表明 MGIV<sub>vu</sub>TF3 及其所携带的 CRISPR-Cas<sub>TF3</sub> 并非宿主菌固有, 而是通过基因水平转移获得, 进一步验证了上述研究者的推测。然而迄今为止并未发现任何一类细菌或古菌中具有高频率的 CRISPR-Cas 系统存在, 表明 CRISPR-Cas 系统的祖先似乎并不存在于目前已经培养的微生物种类中。

CRISPR-Cas<sub>TF3</sub> 具有多个相同的 S 序列。类似的现象也出现在其他的 CRISPR-Cas 系统中<sup>[24]</sup>, 但目前对于同一个 CRISPR-Cas 系统中具有多个相同的 S 序列的生物学意义尚不清楚。CRISPR-Cas 系统中, S 序列被认为来自于异源入侵的核酸, 包括质粒和噬菌体<sup>[10,12,18]</sup>, CRISPR-Cas<sub>TF3</sub> 具有多达 49 个 S 序列, 然而 BLAST 搜索并没有发现与其序列高度相似的序列, 这一方面表明 CRISPR-Cas<sub>TF3</sub> 的宿主菌可能经历过多次异源 DNA 的入侵, 另一方面表明尽管 NCBI 已经贮备海量的环境微生物核苷酸序列, 但并不足以反映自然界庞大的微生物群体的存在。

早在 2009 年就有研究者对霍乱弧菌 O395 上的 CRISPR 系统进行了报道<sup>[25]</sup>, 但没有对此

CRISPR 系统所处的位置进行分析, 因此没有发现霍乱弧菌 O395 中具有类似 MGIV<sub>vu</sub>TF3 的基因岛。MGIV<sub>ch</sub>0395 和 MGIV<sub>vu</sub>TF3 具有广泛的相似性和同源性, 二者也存在完全一致的基因结构和排列, 因此有理由推测二者来源同一个祖先, 然而 MGIV<sub>ch</sub>0395 和 MGIV<sub>vu</sub>TF3 不尽相同, 特别是 *int*、*orf1*、R 序列, 很可能在漫长的进化过程中, 2 个基因岛已经趋异。此外, 二者的 S 序列完全不同, 表明一系列 DNA 入侵事件很可能发生在宿主菌获得 CRISPR 系统后。另外, CRISPR-Cas 系统是大多数细菌和对抗异源核酸入侵的免疫武器, 我们是否能够借助基因岛转移的方式把此系统转移到一些重要的工业或农业应用菌种中, 使其获得抵制噬菌体侵染的免疫能力, 也需要进一步的探索。

### 参考文献

- [1] 王传得. 致病性弧菌 MPCR-DHPLC 检测和分型方法的建立. 福建农林大学硕士学位论文, 2009.
- [2] 潘军航. 浙江沿海创伤弧菌的分布特征及其种群结构研究. 浙江大学博士学位论文, 2015.
- [3] Yuan M, Yuan YM, Chen HB, Luo JY, Yu MH. Plasmid diversity and multilocus sequencing typing of *Vibrio vulnificus* isolated from the environment of Nanshan District, Shenzhen City. *Practical Preventive Medicine*, 2016, 23(10): 1167-1169, 1175. (in Chinese)  
袁梦, 袁月明, 陈宏彬, 罗锦雁, 俞慕华. 深圳地区环境中创伤弧菌携带质粒分析及多位点序列遗传分型. *实用预防医学*, 2016, 23(10): 1167-1169, 1175.
- [4] Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science*, 2010, 327(5962): 167-170.
- [5] Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology*, 1987, 169(12): 5429-5433.
- [6] Jansen R, van Embden JDA, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in

- prokaryotes. *Molecular Microbiology*, 2002, 43(6): 1565–1575.
- [7] Sorek R, Kunin V, Hugenholtz P. CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Reviews Microbiology*, 2008, 6(3): 181–186.
- [8] Alkhnbashi OS, Shah SA, Garrett RA, Saunders SJ, Costa F, Backofen R. Characterizing leader sequences of CRISPR loci. *Bioinformatics*, 2016, 32(17): i576–i585.
- [9] Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJM, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin EV. An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology*, 2015, 13(11): 722–736.
- [10] Swarts DC, Mosterd C, Van Passel MWJ, Brouns SJJ. CRISPR interference directs strand specific spacer acquisition. *PLoS One*, 2012, 7(4): e35888.
- [11] Rath D, Amlinger L, Rath A, Lundgren M. The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie*, 2015, 117: 119–128.
- [12] Garneau JE, Dupuis MÈ, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 2010, 468(7320): 67–71.
- [13] Grissa I, Vergnaud G, Pourcel C. CRISPRfinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research*, 2007, 35(S2): W52–W57.
- [14] Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, 1999, 27(2): 573–580.
- [15] Abbott JC, Aanensen DM, Rutherford K, Butcher S, Spratt BG. WebACT—an online companion for the Artemis comparison tool. *Bioinformatics*, 2005, 21(18): 3665–3666.
- [16] Haft DH, Selengut J, Mongodin EF, Nelson KE. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Computational Biology*, 2005, 1(6): e60.
- [17] Feng H, Dong G, Yong B, Li W. Function and organization of CRISPR-Cas systems and application in biotechnology. *Journal of Sichuan Normal University (Natural Science)*, 2014, 37(2): 268–281. (in Chinese)
- 冯红, 董阁, 雍彬, 李维. 原核 CRISPR-Cas 系统的结构功能及应用. *四川师范大学学报(自然科学版)*, 2014, 37(2): 268–281.
- [18] Silas S, Mohr G, Sidote DJ, Markham LM, Sanchez-Amat A, Bhaya D, Lambowitz AM, Fire AZ. Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science*, 2016, 351(6276): aad4234.
- [19] Daccord A, Ceccarelli D, Rodrigue S, Burrus V. Comparative analysis of mobilizable genomic islands. *Journal of Bacteriology*, 2013, 195(3): 606–614.
- [20] Makarova KS, Aravind L, Wolf YI, Koonin EV. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology Direct*, 2011, 6: 38.
- [21] Ledford H. Five big mysteries about CRISPR's origins. *Nature*, 2017, 541(7637): 280–282.
- [22] Chakraborty S, Snijders AP, Chakravorty R, Ahmed M, Tarek AM, Hossain MA. Comparative network clustering of direct repeats (DRs) and *cas* genes confirms the possibility of the horizontal transfer of CRISPR locus among bacteria. *Molecular Phylogenetics and Evolution*, 2010, 56(3): 878–887.
- [23] Portillo MC, Gonzalez JM. CRISPR elements in the thermococcales: evidence for associated horizontal gene transfer in *Pyrococcus furiosus*. *Journal of Applied Genetics*, 2009, 50(4): 421–430.
- [24] Ge CY, Lu HX, Qu DF, Zhu JL, Yuan DX, Feng LF. Comparative analysis of clustered regularly interspaced short palindromic repeats (CRISPR) in the genomes of *Vibrio parahaemolyticus* group. *Genomics and Applied Biology*, 2015, 34(8): 1712–1722. (in Chinese)
- 葛彩云, 陆海霞, 曲道峰, 朱军莉, 袁冬霞, 冯立芳. 副溶血性弧菌群中规律成簇间隔短回文重复序列的比较分析. *基因组学与应用生物学*, 2015, 34(8): 1712–1722.
- [25] Chakraborty S, Waise TMZ, Hassan F, Kabir Y, Smith MA, Arif M. Assessment of the evolutionary origin and possibility of CRISPR-Cas (CASS) mediated RNA interference pathway in *Vibrio cholerae* O395. *Silico Biology*, 2009, 9(4): 245–254.

# Analysis of a CRISPR-Cas system and a genomic island, MGIV<sub>vu</sub>TF3, in complete genome of *Vibrio vulnificus* TF3

Long Yun<sup>1,2</sup>, Peng Luo<sup>1\*</sup>, Yushun Tian<sup>1,2</sup>, Xiongqi Ding<sup>1,2</sup>, Chaoqun Hu<sup>1</sup>

<sup>1</sup> CAS Key Laboratory of Tropical Marine Bio-resources and Ecology (LMB), Guangdong Provincial Key Laboratory of Applied Marine Biology (LAMB), South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510301, Guangdong Province, China

<sup>2</sup> University of Chinese Academy of Science, Beijing 101408, China

**Abstract:** **[Objective]** In order to dig interesting genetic regions and explore whether CRISPR-Cas systems exist, the draft genome of *Vibrio vulnificus* TF3 was analyzed. **[Methods]** We isolated *V. vulnificus* strain TF3 from the diseased shrimp *Litopenaeus vannamei*, and then obtained its genome sequence by Illumina Miseq sequencing. The genomic annotation showed that there was a CRISPR-Cas system in *V. vulnificus* TF3, named CRISPR-Cas<sub>TF3</sub>. After further analysis, we found that the CRISPR-Cas<sub>TF3</sub> located in a genomic island, named MGIV<sub>vu</sub>TF3. The characteristics and sources of CRISPR-Cas<sub>TF3</sub> and MGIV<sub>vu</sub>TF3 were analyzed using bioinformatic methods. **[Results]** We found the CRISPR-Cas<sub>TF3</sub> belonged to type I-E CRISPR-Cas system similar to that of *Escherichia coli*. The CRISPR-Cas<sub>TF3</sub> contained eight *cas* genes with the order, *cas3-cas8e-cse2-cas6-cas7-cas5-cas1-cas2*. CRISPR-Cas<sub>TF3</sub> had 50 repeats sequences separated by 49 spacer sequences. MGIV<sub>vu</sub>TF3 featured attL/attR sequences, and key genes participating in site-specific integration, excision, and transfer, and thus it manifested that MGIV<sub>vu</sub>TF3 represents a genomic island. MGIV<sub>vu</sub>TF3 had highly similarity to a genomic island, MGIV<sub>ch</sub>0395, found in *V. cholerae* O395. However, the two genomic islands also had two distinct differences, namely, they had completely different spacer sequences and each of them had several particular predicted genes. **[Conclusion]** MGIV<sub>vu</sub>TF3 harboring CRISPR-Cas<sub>TF3</sub> is most likely obtained through horizontal gene transfer, and therefore the CRISPR-Cas<sub>TF3</sub> system may transfer horizontally by means of MGIV<sub>vu</sub>TF3.

**Keywords:** CRISPR, *cas* genes, genomic island, *Vibrio vulnificus*

(本文责编: 张晓丽)

Supported by the National Natural Science Foundation of China (31370149) and by the Program of Fishery Problem Tackling of Guangdong Province (20170228-9-3)

\*Corresponding author. E-mail: luopengli@163.com

Received: 15 August 2017; Revised: 21 November 2017; Published online: 9 January 2018