

海洋细菌胞外蛋白酶 PPC 结构域序列特性与系统进化分析

黄嘉丰, 刘 丹, 杨兴昊, 吴日帮, 武翠玲, 何海伦

(中南大学 生命科学学院 医学遗传学国家重点实验室, 湖南 长沙 410013)

摘要: PPC(Pre-Peptidase C-terminal)结构域广泛分布于分泌型海洋细菌蛋白酶的 C 末端, 对蛋白酶的分泌, 锚定及与底物相互吸附起到重要作用。对 PPC 结构域序列分析发现, PPC 结构域存在两个较为保守序列区, 且这两个保守区空间结构为 β -折叠。尽管有的 PPC 结构域序列一致性较低, 但不同结构域间存在相同的保守序列, 且其三维结构相似。为了进一步对不同来源的 PPC 结构域的系统进化关系进行研究, 本研究选取公开发表的 PPC 结构域, 通过在 NCBI 数据库中搜索得到 61 条来自于 39 个细菌所分泌的不同蛋白酶的 PPC 的氨基酸序列进行系统发育分析。发现 PPC 结构域在生物进化过程中序列变异性大, 保守性弱。其中, 有些细菌 PPC 结构域可能来源于自身基因重复产生, 有些细菌 PPC 结构域则可能由水平基因转移产生。

关键词: 蛋白酶; PPC 结构域; 保守区; 系统进化树; 基因水平转移

中图分类号: Q93 文献标识码: A 文章编号: 1000-3096(2015)01-0013-11

doi: 10.11759/hyxx20140519002

基因水平转移(Horizontal gene transfer, HGT)是指在具有生殖隔离的不同物种之间, 或单个细胞的叶绿体、线粒体等细胞器之间, 以及细胞器和细胞核之间所进行的 DNA 片段流动^[1-2]。基因交流可以发生在亲缘关系较近的同属不同种之间, 也可以是亲缘关系较远的, 或没有亲缘关系的生物种之间, 有的甚至跨越了界的限制^[3]。基因水平转移对研究在追踪菌株来源, 研究菌株变迁情况以及探讨造成不同地区流行菌株之间的关系上有重要意义。越来越多的证据显示基因水平转移是原核生物进化的重要动力。Santiago 等^[4]对 17 个真细菌和 7 个古生菌的全基因组序列分析发现: 通过水平转移得到的基因可以占到全基因数的 1.5%~14.5%。另有研究发现, 结构域的基因水平转移事件比较容易发生在参与物质和能量代谢的酶蛋白中, 特别是在海洋微生物中, 这种现象普遍存在^[5]。这种多结构域的酶蛋白, 通过结构域基因的重复或是水平转移, 使自身具有了新的功能和催化特性, 是其长期进化和环境适应的结果^[2, 3]。

海洋细菌在自然界氮循环过程中发挥了重要的作用, 近年来不断有新的海洋细菌胞外蛋白酶被发现^[6]。研究发现, PPC(Pre-Peptidase C-terminal)结构域广泛分布于分泌型海洋细菌蛋白酶的 C 末端, 通常以 1 个 PPC 结构域单独存在或 2 个 PPC 结构域串联重复出现, 每个 PPC 结构域包含 70~90 个氨基酸^[7-8]。PPC 结构域多存在于细菌分泌型胞外蛋白酶中, 但

并不存在于酶的催化结构域中。PPC 结构域在细菌和古菌分泌的蛋白酶中都有发现, 特别是在海洋细菌和致病菌蛋白酶中较为多见^[7-8]。如 *Vibrio*, *Pseudoalteromonas*, *Shewanella*, *Alteromonas*, *Aeromonas* 等菌分泌的胞外蛋白酶中很多都含有 PPC 结构域^[9]。目前发现 PPC 结构域存在于至少四个不同家族的蛋白酶中^[10], 金属蛋白的 M4、M9、M28 家族和丝氨酸蛋白酶的 S8 家族。序列比对显示 PPC 结构域与已知的 PKD 结构域没有任何同源性, 但二级结构分析显示 PPC 结构域也是典型的 β 片层结构^[7]。这预示着这些结构域在蛋白酶中可能行使着类似的功能, 在遗传进化上存在着一定的相关性。但目前还没有相关系统研究, 只是分别对个别蛋白酶的 C 端 PPC 结构域进行了初步的功能检测。如 PPC 结构域可能与酶蛋白酶的分泌、锚定或是底物吸附有一定的相关性, 而且来源不同的细菌蛋白酶 PPC 结构域的功能也大不相同^[9, 11], 推测该 PPC 结构域的存在可能与蛋白酶的环境适应和遗传进化密切相关。

收稿日期: 2014-05-19; 修回日期: 2014-07-22

基金项目: 国家自然科学基金项目(31370104, 31070061); 湖南省自然科学基金联合基金项目(13JJ9001); 国家星火计划项目(2013GA770009); 中南大学大学生创新创业训练计划项目(国家级)(201310533050)

作者简介: 黄嘉丰(1993-), 男, 广东汕头人, 大学本科, 专业方向为微生物酶学, E-mail: 1608110217@csu.edu.cn; 何海伦, 通信作者, 博士, 教授, 研究方向为微生物酶学, E-mail: helenhe@csu.edu.cn

目前关于 PPC 结构域进化关系的研究鲜有报道,本研究从数据库中获取 39 种海洋细菌共 61 个 PPC 结构域,实验通过 16S rDNA、PPC 结构域的系统进化分析,结构域和物种基因组的进化距离对比,GC%含量比较等方法,推测不同来源的 PPC 结构域间可能的进化关系,为深入了解不同细菌分泌的多结构域蛋白酶分子进化机制奠定基础,从而有助于进一步了解产带有 PPC 结构域胞外蛋白酶细菌之间的进化关系。

1 材料及方法

1.1 材料

通过在 GenBank、同源蛋白质聚类数据库(COG)、MEROPS 等相关数据库中搜索具有 PPC 结构域的细菌蛋白酶基因序列和氨基酸序列,共得到 61 个 PPC 结构域的氨基酸序列,所得结构域来源于 39 株细菌,其中以 *Pseudoalteromonas*, *Vibrio*, *Shewanella*, *Alteromonas* 种属的细菌占多数。

1.2 方法

1.2.1 PPC 结构域二级结构分析

采用软件 Clustal X, DNAMAN 对选取的 61 个 PPC 结构域进行氨基酸序列比对,标志序列间的保守序列。标记细菌蛋白酶 N 端第 1 个 PPC 结构域为 PPC1,第 2 个 PPC 结构域为 PPC2,依此类推。

利用 ExPASy-Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) 对 PPC 结构域进行同源建模,构建结构域的二级、三级结构。比较结构域间异同,分析保守序列的结构特征^[12]。

1.2.2 系统进化分析

采用 Clustal X 程序对 PPC 结构域的氨基酸序列细菌的 16S rDNA 碱基序列进行比对,个别位点进行人工校正。使用 MEGA 软件最大简约法(Maximum Parsimony, MP) 构建系统发育树,系统树的可靠性经 1000 次重复的自展法进行检验。对得到的进化树进行系统进化分析,选择未能与同一谱系(相同种属)菌株很好聚类的菌株,其 PPC 结构基因作为下一步研究的候选水平转移基因^[13-14]。

1.2.3 菌株 GC 含量分析

通过 NCBI-Genome(<http://www.ncbi.nlm.nih.gov/genome/>)搜索细菌全基因组 GC 含量及 PPC 结构域的核苷酸序列 GC 含量。当 PPC 结构基因 GC 含量与该细菌全基因 GC 含量存在明显差异时,则可判断该结构域基因发生基因水平转移,从而进一步确

认候选菌株是否发生基因转移^[12]。

1.2.4 PPC 结构域序列一致性分析

结合系统发育进化树以及细菌相关基因 GC 含量,利用 Clustal X, DNAMAN 软件对候选水平基因转移的 PPC 结构域氨基酸序列进行序列一致性比对,进一步判断 PPC 结构域的可能来源。

1.2.5 PPC 结构域特征性分析

将 PPC 结构域, 16S rDNA, 结构域所在酶基因及全基因组序列的 GC 含量进行比对,并以细菌全基因 GC 含量为分母,其他三者为分子做 GC 含量比值分布图,分析结构域与全基因组间的差异变化^[2]。推测存在基因水平转移的标准,定为该序列的组成与全基因组序列组成的值间的差异超过 5%^[15]。

2 结果与分析

2.1 PPC 结构域序列分析

对 PPC 结构域氨基酸序列相似度比对结果显示,来源不同的 PPC 结构域同源性较低,物种间的 PPC 结构域序列相似度在 18%~100%(图 1)。从图中可看出,PPC 结构域氨基酸序列的部分片段具有很高的保守性,序列的保守区和可变区间隔排列,其中包括完全保守区和相对保守区(>75%)。序列比对显示,完全保守区的氨基酸序列为 -GG*GDADLYV*-和 -Y*CRPY**GN*E*C-。并利用 ExPASy-Phyre2 程序在线预测 PPC 结构域二级结构及空间构象,同一种属、不同种属之间 PPC 结构域三维空间结构具有明显的相似性,均为 6 个 β -折叠区,间断排列,不具有 α -螺旋结构(图 2),同一种属细菌 PPC 结构域空间结构相似(图 3A, B);不同种属细菌 PPC 结构域空间结构亦相似(图 3A, C, D),都为富含 β -折叠区域,由此推测,PPC 结构域的 β -折叠结构可能与其功能密切相关^[16-18]。此外,氨基酸序列分析发现,PPC 结构域内含有较多酪氨酸,酪氨酸的侧链具有芳香环,常在蛋白相互作用中起到重要作用^[16, 19]。PPC 结构域富含酪氨酸,可能对蛋白质底物起到一定的吸附功能^[19]。PPC 结构域内保守区域的蛋白二级结构基本都为 β -折叠结构区,且芳香族氨基酸大多存在于 PPC 结构域的 β -折叠区,也进一步说明 PPC 结构域内保守序列与蛋白酶结合底物的密切相关。

2.2 基因水平转移分析

2.2.1 系统发育进化分析

通过对 39 个不同细菌来源的 16S rDNA 碱基序

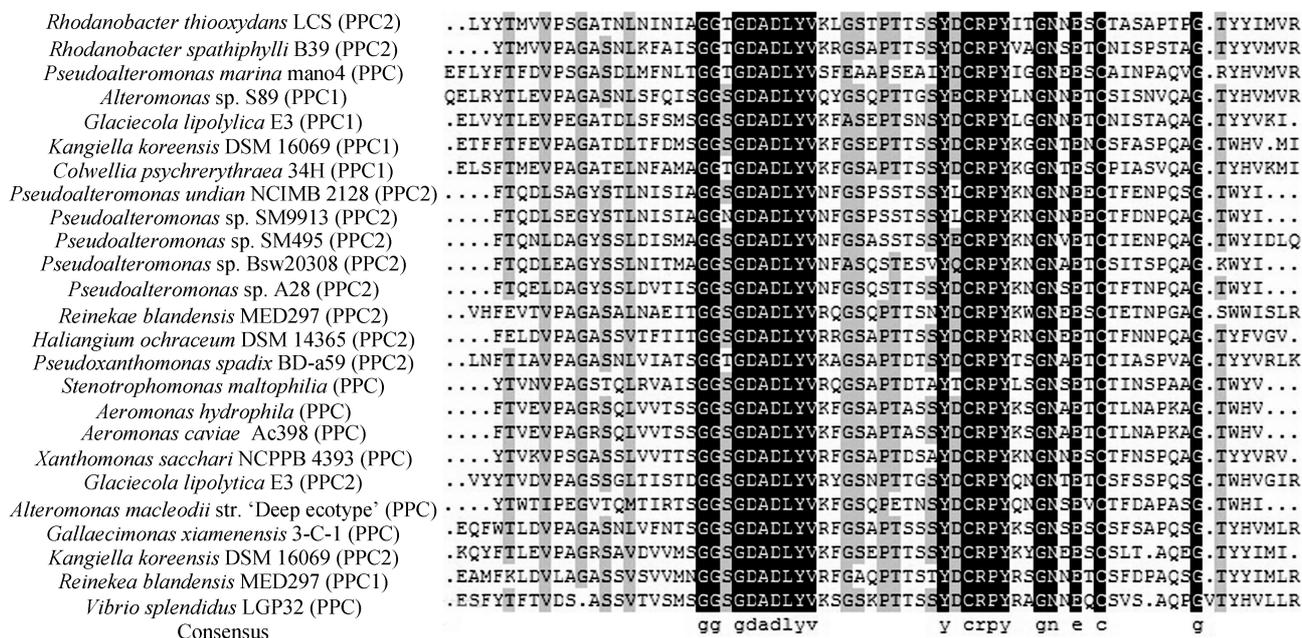


图 1 部分 PPC 结构域序列比对分析

Fig.1 Sequence alignment analysis of partial PPC domain

黑色区域表示 100%同源; 灰色区表示同源性>75%; Consensus 表示 100%同源区域氨基酸序列

The black areas represent 100% sequences homology; Grey areas represent homology more than 75%; Consensus shows the identical sequences

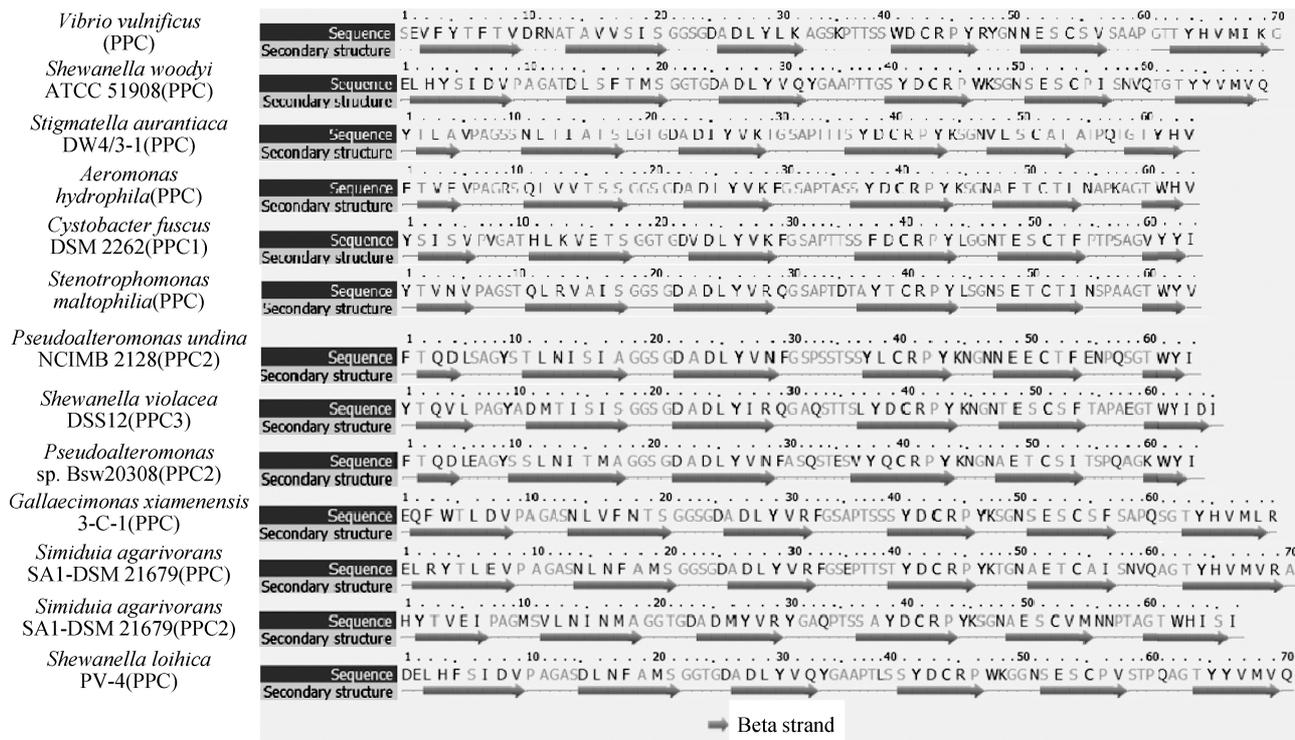


图 2 Phyre2 预测 PPC 结构域的二级结构示意图

Fig.2 Prediction of the secondary structure of PPC domain by Phyre2

➡ 表示该氨基酸序列为 β-折叠区

➡ represent β-sheet of the amino acid sequences

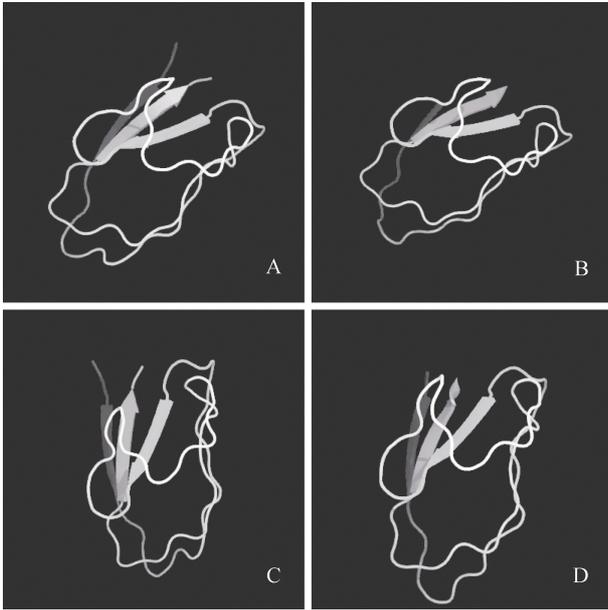


图3 PPC结构域三维结构图

Fig.3 Three dimensional structure of PPC domain

A. *Pseudoalteromonas* sp. SM495 PPC1 结构域三维结构; B. *Pseudoalteromonas* sp. SM495 PPC2 结构域三维结构; C. *Shewanella violacea* DSS12 PPC1 结构域三维结构; D. *Gallaecimonas xiamenensis* 3-C-1 PPC1 结构域三维结构

A. Three-dimensional structure of *Pseudoalteromonas* sp. SM495 PPC1 domain; B. Three-dimensional structure of *Pseudoalteromonas* sp. SM495 PPC2 domain; C. Three-dimensional structure of *Shewanella violacea* DSS12 PPC1 domain; D. Three-dimensional structure of *Gallaecimonas xiamenensis* 3-C-1 PPC1 domain

列进行人工修饰, 最终比对的序列大小在 1 400 bp 左右。使用 MEGA 软件最大简约法(Maximum Parsimony, MP)构建系统发育树, 系统树的可靠性经 1 000 次重复的自展法进行检验, 得细菌 16S rDNA 系统进化树(图 4)。进化分析显示, 39 株细菌分别来源于 21 种不同种属, 通过对细菌间进化距离的分析计算, 得进化树平均进化距离为 0.147 2, 方差 0.003, 最大进化距离为 0.266 5。同一种属细菌聚类明显, 包括 *Pseudoalteromonas*, *Vibrio*, *Shewanella*, *Alteromonas* 等。且各类群之间同源性不高, 具有明显种属差异性。

通过对 61 个 PPC 结构域氨基酸序列进行人工修饰, 最终比对序列大小在 72 个氨基酸左右。使用 MEGA 软件最大简约法(Maximum Parsimony, MP)构建系统发育树, 系统树的可靠性经 1 000 次重复的自展法进行检验, 得 PPC 结构域系统进化树(图 5)。PPC 结构域进化分析结果显示, PPC 结构域进化距离差异较大, 进化距离的分析计算, 发现结构域进化树平均进化距离为 0.517 4, 方差 0.027 6, 最大进化距离为 1.563 1。表明该结构域在进化过程中变异大,

保守性弱。PPC 结构域系统进化树显示, *Pseudoalteromonas* 结构域 PPC1 和 PPC2 各聚成一类, 且 *Pseudoalteromonas* PPC1 聚类中有 *Shewanella* PPC1 与其同源类聚, 表明菌 *Shewanella* 中 PPC 基因可能由菌 *Pseudoalteromonas* PPC 基因水平转移而来。由图 5 可知, 同一聚类的结构域同源性较高, 但两聚类之间则存在明显差异; *Vibrio* 的 PPC 结构域独自构成一聚类, 无其他菌属混杂, 表明弧菌的 PPC 结构域在进化过程中较为稳定; 与 *Vibrio* 相反, *Shewanella* 的 PPC 结构域则相对分散, 结构域间差异性较大, 表明该菌 PPC 结构域在进化过程中可能发生结构域的水平基因转移, 如 *Reinekea blandensis* MED297(PPC2) 与 *Shewanella violacea* DSS12 (PPC3), *Pseudoalteromonas* sp. SM495(PPC1) 与 *Shewanella violacea* DSS12(PPC1)。在比对结果中, 除同一谱系外, 有些基因序列与其他谱系的菌株有较高的相似性, 具有明显的水平转移的迹象, 因此, 初步判断有 12 个细菌的 PPC 结构域可能发生了基因转移(图 5)。另外, *Xanthomonas campestris* 自身 PPC1 与 PPC2 进化分析类聚成一束, 推测可能发生了基因内重复。

2.2.2 基因组和结构域中 GC 含量分析

不同细菌物种之间基因组 GC 含量是不同的, 每个细菌物种基因组的 GC 含量相对来说是比较稳定的, 而且在不同基因间是相对一致的, 它们不受外界因素的影响^[20-21]。如果某菌株某段特定 DNA 序列的 GC 含量明显高于或低于其基因组的其他部分, 就暗示着该特定 DNA 序列是通过水平转移从外源的细菌或其他物种中得到的。如 Karaolis 等对霍乱弧菌 (*Vibrio cholerae*) 毒力基因亦即致病性岛 (Pathogenicity Island) 研究发现, 霍乱弧菌毒力基因的 GC 含量是 35%, 低于全基因组的 GC 含量(47%~49%), 并且与霍乱弧菌以外的肠致病大肠杆菌 (*Enteropathogenic Escherichia coli*) 的毒力岛的结构相似, 也就暗示霍乱弧菌毒力岛基因可能是由肠致病大肠杆菌发生基因水平转移得来的^[12]。

实验通过 NCBI-Genome(<http://www.ncbi.nlm.nih.gov/genome/>)搜索 39 株细菌全基因组 GC 及 PPC 结构域的核苷酸序列, GC 含量计算结果见表 1。

比较表 1 中 PPC 结构域基因 GC 含量与细菌基因组 GC 含量存在的差异, 序号 3(PPC2), 8(PPC1), 9(PPC1), 11(PPC1), 13(PPC1), 15(PPC1), 15(PPC2), 17(PPC2), 24, 25, 28, 30, 36(PPC2), 38 以及 39 的细菌

PPC 结构域的 GC 含量与基因组 GC 含量差异较大, 其中, 有 4 个结构域基因为本次实验的候选水平转移基因, 分别是 9. *Alteromonas* sp. S89 (PPC1), 11. *Kangiella koreensis* DSM 16069 (PPC1), 15. *Colwellia*

psychrerythraea 34H (PPC1), 17. *Glaciecola lipolytica* E3 (PPC2)。其他非候选基因在进化树中或自成一束, 或与同属细菌聚集, 因此判断其是否发生水平转移事件有待数据量的进一步扩大。

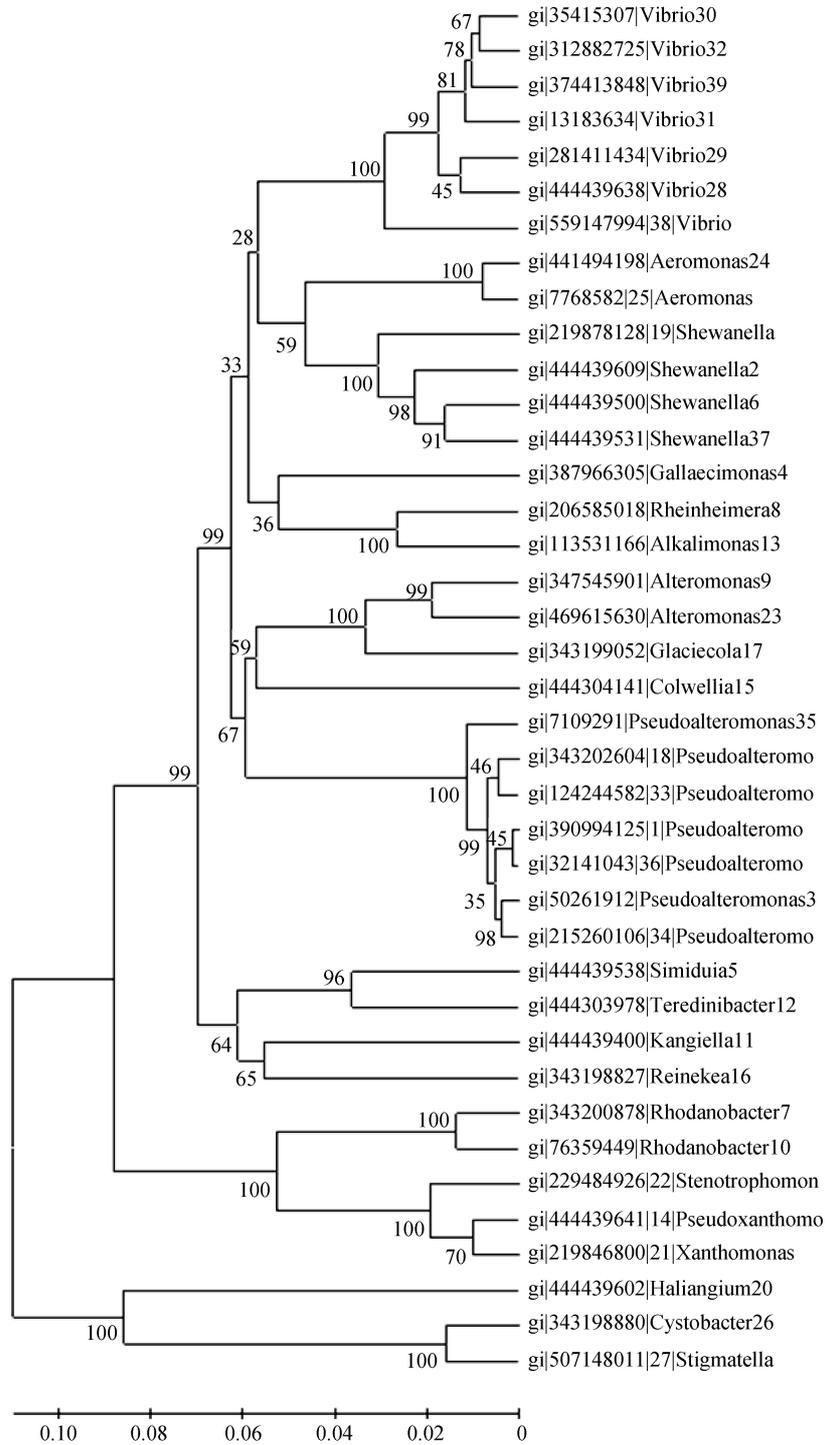


图 4 基于细菌 16S rDNA 序列使用最大简约法构建的系统进化树

Fig.4 The phylogenetic tree, using maximum parsimony, of bacterium based by 16S rDNA sequences
分枝长度表示进化距离
the branch length represents the evolutionary distance

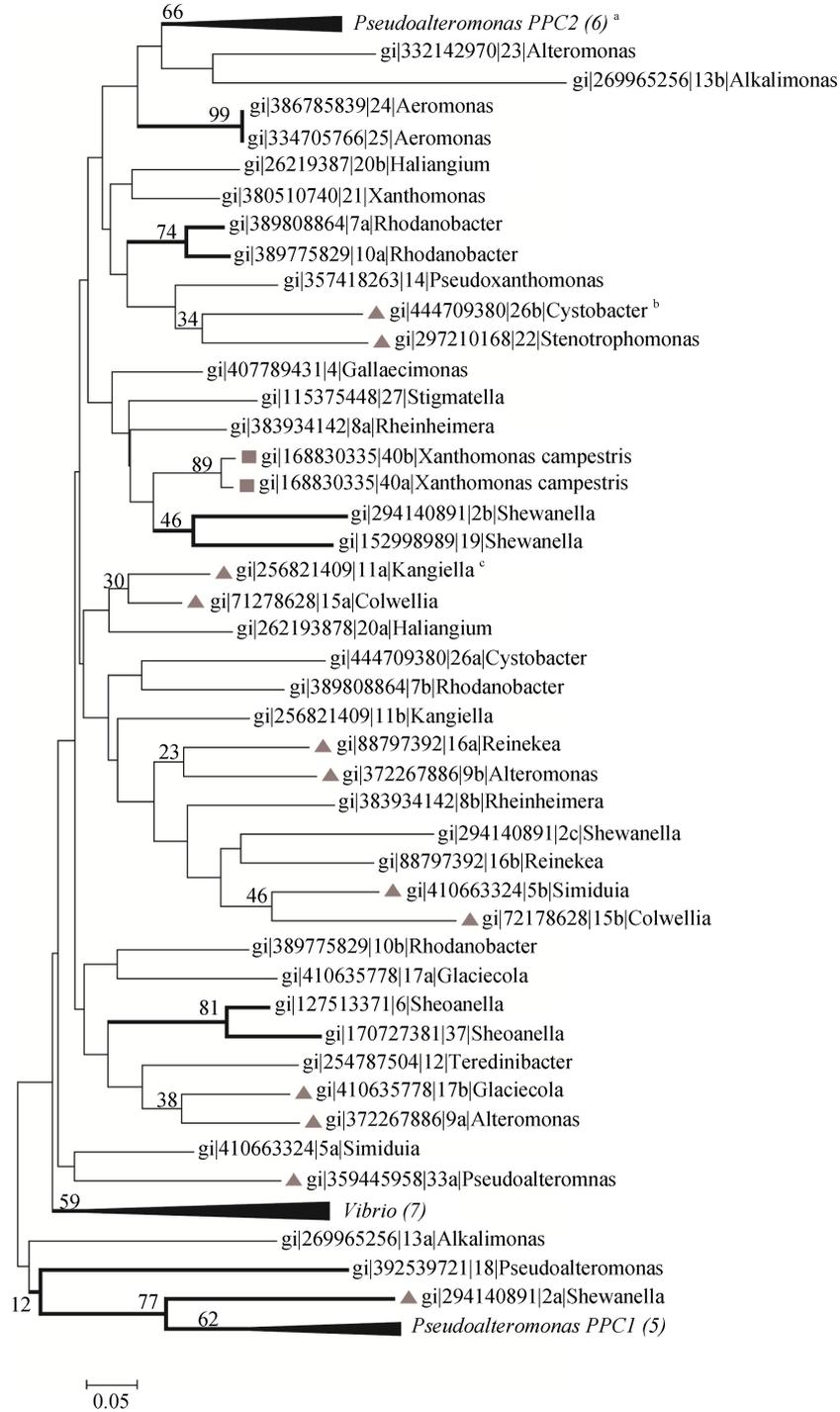


图 5 基于 PPC 结构域氨基酸序列使用最大简约法构建的系统进化树

Fig. 5 The phylogenetic tree of bacterium based on amino acid sequence of PPC domain, using maximum parsimony

- a. 表示有 6 个 *Pseudoalteromonas* 的第二个(靠近 C 端)PPC 结构域聚成一树;
 - b. ▲ 表示可能发生水平基因转移; ■ 表示该菌可能发生基因内结构域重复;
 - c. gi 表示该结构域在 NCBI 上的搜索号; 11a|*Kangiella* 表示编号为 11 的细菌 *Kangiella* 蛋白酶 N 端数第 1 个 PPC 结构域信息
- a. There are 6 PPC2 domains (close to the C-terminal) of *Pseudoalteromonas* gathering into a bunch;
- b. ▲ indicates it may occur horizontal gene transfer; ■ indicates the potential duplication of the PPC domain gene;
- c. gi represents the serial number of the domain on the NCBI; 11a|*Kangiella* is the PPC domain (close to the N-terminal) of the bacterium *Kangiella* numbered 11

表 1 实验菌株相关结构基因 GC 的含量

Tab. 1 The percentage of GC in the strain related genes

菌株 编号	菌株名称	GC (%)					
		全基因组	酶基因	16S rDNA	PPC1 ^a	PPC2	PPC3
1	<i>Pseudoalteromonas undina</i> NCIMB 2128	39.9	39.7	52.2	40.7	39.1	
2	<i>Shewanella violacea</i> DSS12	44.7	41.0	53.7	42.3	46.3	44.6
3	<i>Pseudoalteromonas</i> sp. Bsw20308	38.9	37.9	52.9	35.6	33.6^c	
4	<i>Gallaecimonas xiamenensis</i> 3-C-1	NULL	NULL ^b	54.2	NULL		
5	<i>Simiduia agarivorans</i> SA1 = DSM 21679	55.8	55.7	54.7	52.5	53.0	
6	<i>Shewanella loihica</i> PV-4	53.7	52.4	53.9	57.3		
7	<i>Rhodanobacter thiooxydans</i> LCS2	67.2	67.9	56.7	65.9	65.7	
8	<i>Rheinheimera nanhaiensis</i> E407-8	51.3	54.5	53.4	59.8	51	
9	<i>Alteromonas</i> sp. S89	56.8	63.3	53.3	61	60.4	
10	<i>Rhodanobacter spathiphylli</i> B39	66.5	67.8	56.3	64.4	67.8	
11	<i>Kangiella koreensis</i> DSM 16069	43.7	44.1	52.3	49.3	46.7	
12	<i>Teredinibacter turnerae</i> T7901	50.9	52.6	53.5	52.7		
13	<i>Alkalimonas collagenimarina</i>	NULL	50.7	53.6	45.3	50.9	
14	<i>Pseudoxanthomonas spadix</i> BD-a59	66.7	67.4	54.8	68.7		
15	<i>Colwellia psychrerythraea</i> 34H	38	39.5	53.0	42.3	43.2	
16	<i>Reinekea blandensis</i> MED297	52.4	53.3	53.9	53.2	52.5	
17	<i>Glaciecola lipolytica</i> E3	40.8	45.0	52.4	42.8	44.9	
18	<i>Pseudoalteromonas marina</i> mano4	39.7	40.8	52.9	37.9		
19	<i>Shewanella baltica</i> OS185	46.3	47.2	47.5	50.2		
20	<i>Haliangium ochraceum</i> DSM 14365	69.5	68.0	55.6	69.7	71.1	
21	<i>Xanthomonas sacchari</i> NCPPB 4393	69	NULL	55.6	NULL		
22	<i>Stenotrophomonas maltophilia</i>	66.8	68.4	69.5	67.8		
23	<i>Alteromonas macleodii</i> str. 'Deep ecotype'	44.9	48.5	55.4	43.9		
24	<i>Aeromonas hydrophila</i>	61.5	64.9	52.7	68.3		
25	<i>Aeromonas caviae</i> Ae398	61.4	64.9	55.3	68.3		
26	<i>Cystobacter fuscus</i> DSM 2262	68.6	67.8	57.2	72.2	70.4	
27	<i>Stigmatella aurantiaca</i> DW4/3-1	67.5	66.6	56.6	65.3		
28	<i>Vibrio splendidus</i> LGP32	43.9	43.7	68.4	50.7		
29	<i>Vibrio</i> sp. RC586	46.4	47.5	53.5	44.7		
30	<i>Vibrio coralliilyticus</i> ATCC BAA-450	45.7	48.8	48.2	51.2		
31	<i>Vibrio</i> sp. Ex25	44.9	46.8	54.0	47.6		
32	<i>Vibrio caribbenthicus</i> ATCC BAA-2122	41.6	45.7	53.9	45.2		
33	<i>Pseudoalteromonas</i> sp. BSi20439	40.2	44.5	52.6	44.1	41.1	
34	<i>Pseudoalteromonas</i> sp. SM495	NULL	39.1	53.2	41.6	38.6	
35	<i>Pseudoalteromonas</i> sp. A28	NULL	40.9	53.2	39.7	43.5	
36	<i>Pseudoalteromonas</i> sp. SM9913	40.3	37.2	52.9	38.5	35.3	
37	<i>Shewanella woodyi</i> ATCC 51908	43.7	43.5	53.8	47.8		
38	<i>Vibrio cholerae</i> PS15	47.5	49.7	39.1	54		
39	<i>Vibrio vulnificus</i>	46.7	49.3	54.2	52.1		

注: a. 有些蛋白酶的 C 末端有多个 PPC 结构域, 记靠近 N 端的第一个结构域为 PPC1, 依此类推;

b. NULL 表示在数据库内尚无该结构信息;

c. 数字加粗表示此 PPC 结构域 GC% 与基因组(或酶基因)GC% 差异较大, 表明该结构域可能发生水平基因转移

2.3 PPC 结构域特征性分析

将 PPC 结构域, PPC 所在蛋白酶基因, 16S rDNA 及全基因组序列的 GC 含量进行比对, 以细菌全基因组 GC 作为分母, 分别做出 PPC, 16S rDNA 以及酶基因 GC% 比值散点图(图 6)。对所得比值进行 F-检验, 检验结果显示, PPC 结构基因与全基因 GC 含量比值在 0.8683~1.1675 之间, 平均值为 1.0323, 方差 0.0051; 16S rDNA 基因与全基因 GC 含量比值在 0.7996~1.3957 之间, 平均值为 1.0939, 方差为 0.0371; 酶基因与全基因 GC 含量比值在 0.9181~1.1148 之间, 平均值为 1.0201, 方差为 0.0027。且 F-检验三组数据 P 值均小于 0.01, 说明差异具有统计学意义。

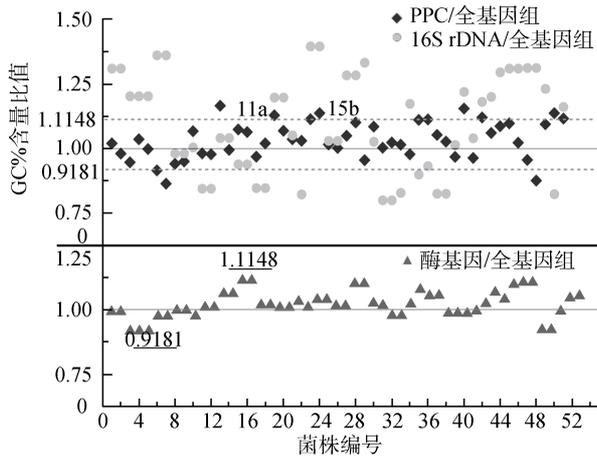


图 6 菌株 PPC 结构域、16S rDNA、结构域所在酶基因与全基因组的 GC 含量比值

Fig.6 The GC% content ratios of PPC domain/Whole genome, 16S rDNA/Whole genome and protease/Whole genome

可以看出, 各单元 GC 含量与全基因 GC 含量拟合程度比较: 酶基因>PPC 结构域>16S rDNA。16S rDNA 与细菌基因组的 GC 含量差异明显, 蛋白酶基因与全基因组的 GC 含量高度一致, PPC 结构域基因与全基因组 GC 含量的一致性介于上述两者之间。以酶基因与全基因的 GC 含量比值 0.9181~1.1148 为界限, PPC 结构域与全基因的 GC 含量比值散点图超过

此界限则定位偏离点, 则 PPC 结构域与全基因的 GC 含量比值散点图存在 10 个偏离点(图 6), 这暗示了这 10 个结构域可能来源于水平基因转移。

PPC 结构域基因与全基因组的 GC% 含量比值存在的 10 个偏离点, 其中 11a 和 15b 为本研究的候选转移菌, 而其余 8 个偏离点细菌在进化树上与同一菌属细菌聚成一束(图 4), 是否发生转移事件有待进一步研究。

2.4 PPC 结构域可能的基因水平转移方式

2.4.1 酶基因内 PPC 结构域基因重复

Xanthomonas campestris 的 MprB 蛋白酶基因编码 2 个结构相似的 PPC 结构蛋白。通过对 PPC 结构域构建的系统进化树进一步分析(图 5), 可以发现细菌 *Xanthomonas campestris* 的 MprB 蛋白酶基因内的两个 PPC 结构域在进化分支上类聚成一束, 相似度高。通过序列一致性分析发现相似度达 94%, 且 PPC 结构域位于蛋白酶的 C 末端, 重复出现, 呈串联排列(图 7A, B), 可初步推测 MprB 蛋白酶内两个 PPC 结构域是基因内结构域重复产生。Jenke-Kodama H 等对蓝细菌 Polyketide synthases (PKS) 基因中的 KS-AT 结构域系统分析发现, 蓝细菌 PKS 基因内的两个 KS-AT 结构域聚成一束, 且同源蛋白结构域氨基酸序列相似度达 70% 以上, 在同一 PKS 基因内的 KS-AT 结构域发生了基因内重复^[11, 14]。另外, Sarah A Teichman 等对大肠杆菌和酿酒酵母基因调节网络研究发现, 大肠杆菌和酿酒酵母基因组中转录因子基因存在基因内重复现象: 大肠杆菌同源转录因子蛋白相似度达 77%, 酿酒酵母同源转录因子蛋白相似度达 69%^[22]。本工作关于基因重复研究方面与 Jenke-Kodama H 等所做的实验基本相同, 因此, 通过比对各个模块中结构域间序列的相似度, 可以初步判断结构域间是否有重复发生。

2.4.2 跨种属基因水平转移可能性分析

由 PPC 结构域构建的系统进化树(图 5)以及基因组和结构域中 GC 含量分析(表 1)可知, 共 4 个 PPC

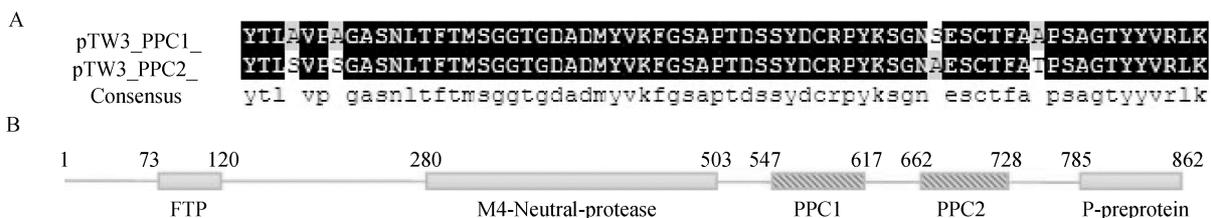


图 7 细菌 *Xanthomonas campestris* 基因内 PPC 结构域同源比对及结构域所在 MprB 蛋白酶结构图

Fig.7 Homologous comparison between two PPC domains in *Xanthomonas campestris* and the structure diagram of the protease

结构域可能是由基因水平转移而来, 分别是 9a. *Alteromonas* sp. S89 (PPC1), 11a. *Kangiella koreensis* DSM 16069 (PPC1), 15a. *Colwellia psychrerythraea* 34H (PPC1)和 17b. *Glaciecola lipolytica* E3 (PPC2)。其中, 值得注意的是, 11a 与 15a 在进化树上距离接近; 9a 与 17b 在进化树上距离接近, 说明这两类 PPC

结构域之间发生水平基因转移可能性较大。序列相似性比较显示, 11a 与 15a 序列相似度为 73%(图 8A); 而 9a 与 17b 序列相似度为 68%(图 8B)。由于现在发现的 PPC 结构域数量有限, 暂无法确定两者 PPC 结构域的进化方向, 因此, 关于它们的进化方向还有待进一步研究。

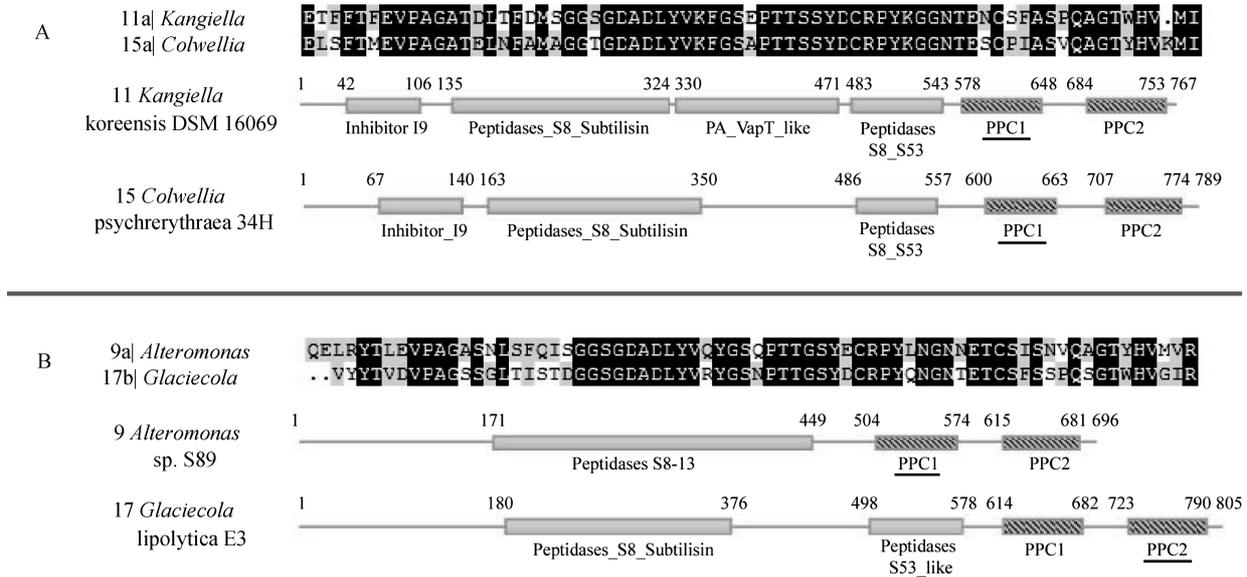


图 8 细菌 *Alteromonas* sp. S89 (PPC1), *Kangiella koreensis* DSM 16069 (PPC1), *Colwellia psychrerythraea* 34H (PPC1)和 *Glaciecola lipolytica* E3 (PPC2) PPC 结构域同源比对及结构域所在 MprB 蛋白酶结构图

Fig.8 Homologous comparison between four PPC domains in *Alteromonas* sp. S89 (PPC1), *Kangiella koreensis* DSM 16069 (PPC1), *Colwellia psychrerythraea* 34H (PPC1) and *Glaciecola lipolytica* E3 (PPC2) and the Structure diagram of those protease

3 讨论

很多研究已经证明, 结构域的水平转移事件比较容易发生在参与物质和能量代谢的酶蛋白中^[21]。因此, 转移事件的发生对生物, 尤其是指微生物, 适应环境具有重要意义, 特别是在生命发生早期地球海洋环境相对恶劣且环境变化较为频繁的情况下, 是导致生物趋同进化的一大自然动力^[22-24]。通过对 PPC 结构域的进化及功能分析, 不难发现, PPC 结构域基因在进化过程中受到基因转移的影响, 在细菌基因内, 细菌与细菌间发生过水平转移事件。新的结构域的加入, 赋予了酶蛋白新的功能, 可以更好的在吸附底物蛋白的过程中发挥作用^[2, 25]。

已有研究发现, 转移事件发生过程中, 亲缘关系和地理位置相比, 相同环境生态位更能决定一个转移基因是否能融入新的基因组中^[19]。Smillie 等研究发现, 寄生在人体内的人类相关细菌发生基因水平转移的概率要比普通细菌发生转移的概率高 25 倍,

从而使人类相关细菌更好的适应其所生存的环境条件^[19]。PPC 结构域在海洋细菌和致病菌蛋白酶中较为多见^[7], 如 *Vibrio*, *Pseudoalteromonas*, *Shewanella*, *Alteromonas*, *Aeromonas* 等菌分泌的胞外蛋白酶中很多都含有 PPC 结构域^[7-8]。He 等通过对南极嗜冷菌 *Pseudoalteromonas* sp. SM495 分泌的胞外蛋白酶 E495 的 PPC 结构域的功能检测, 发现 PPC 结构域可以特异性的吸附海洋中富含的藻类蛋白, 利于蛋白酶 E495 对周围海洋环境中蛋白底物的富集^[8]。而对于弧菌等伤口致病菌, Miyoshi 等则发现海洋致病菌 *Vibro vulnificus* 产的金属蛋白酶 VVP 的 PPC 结构域对酶与不溶性蛋白底物和红血球细胞膜的结合是必需的^[25]。初步推测, 蛋白酶 C 末端 PPC 结构域的存在可能是为了细菌适应生存环境, 满足自身生长的需要。因此, 深入研究 PPC 结构域的进化关系和规律, 可为研究海洋细菌环境适应与结构域进化关系奠定理论基础。

基因重复也是原核生物进化的主要动力之一^[23-24], 基因内结构域重复在酶功能的进化上也具有重要意义。Selvaraj 等^[26]研究发现, 木聚糖酶 C 端的两个重复的碳水化合物结合结构域, 和只有一个结合结构域的酶相比具有更高的酶活性。Jenke-Kodama H 等^[11, 20]研究发现, 蓝细菌中 KS-AT 结构域发生了基因内重复的 I 型 PKS, 在长链脂肪酸合成初期发挥作用。然而, 无 KS-AT 结构域重复的 II 型 PKS 则只在长链脂肪酸合成过程中起作用。在对 PPC 结构域的研究过程中, 我们也发现了海洋细菌 *Xanthomonas campestris* 具有典型的重复现象的发生, 其具体的作用和功能还有待进一步探讨。

参考文献:

- [1] Smillie C S, Smith M B, Friedman J, et al. Ecology drives a global network of gene exchange connecting the human microbiome [J]. *Nature*, 2011, 480(7376): 241-244.
- [2] 刘庆辉, 郭振国, 任嘉红. 原核生物 *eno* 基因在系统进化中应用及水平转移分析[J]. *遗传*, 2012, 34(7): 907-918.
- [3] 李志江, 海权, 刁现民. 基因水平转移的评判方法和转移方式研究进展[J]. *遗传*, 2008, 30(9): 1108-1114.
- [4] Garcia-Vallve S, Romeu A, Palau J. Horizontal gene transfer in bacterial and archaeal complete genomes[J]. *Genome Res*, 2000, 10(11): 1719-1725.
- [5] Gogarten J P, Doolittle W F, Lawrence J G. Prokaryotic evolution in light of gene transfer[J]. *Mol Biol Evol*, 2002, 19: 2226-2238.
- [6] 王静, 党宏月, 杨官品, 等. 南冲绳海槽深海沉积物中度嗜盐菌系统进化及多样性分析[J]. *海洋科学*, 2009, 33(5): 17-22.
- [7] Yeats C, Bentley S, Bateman A. New knowledge from old: in silico discovery of novel protein domains in *Streptomyces coelicolor*[J]. *BMC Microbiol*, 2003, 3: 1-20.
- [8] Penny Coggill, Alex Bateman. The YARHG Domain: An Extracellular Domain in Search of a Function[J]. *PLoS One*, 2012, 7(5): e35575.
- [9] He Hai-Lun, Guo Jun (co-first author), Chen Xiu-Lan, et al. Structural and Functional Characterization of Mature Forms of Metalloprotease E495 from Arctic Sea-Ice Bacterium *Pseudoalteromonas* sp. SM495 [J]. *PLoS ONE*, 2012, 7(4): e35442.
- [10] Rawlings N D, Morton F R, Kok C Y, et al. MEROPS: the peptidase database [J]. *Nucleic Acids Res*, 2008, 36: D320-325.
- [11] Rawlings N D, O'Brien E, Barrett A J. MEROPS: the protease database [J]. *Nucleic Acids Res*, 2002, 30: 343-346.
- [12] Gasteiger E, Gattiker A, Hoogland C, et al. ExpASY: The proteomics server for in-depth protein knowledge and analysis [J]. *Nucleic Acids Res*. 2003, 31(13): 3784-3788.
- [13] 焦豫良, 王淑军, 吕明生, 等. 多结构域酶的结构域进化关系[J]. *生命的化学*, 2012, 32(1): 39-43.
- [14] Dehal P S, Joachimiak M P, Price M N, et al. Microbes Online: an integrated portal for comparative and functional genomics [J]. *Nucleic Acids Res*, 2010, 38(Suppl.1): D396-D400.
- [15] 吕林峰, 岳建宇, 费忠安, 等. 油菜菌核病菌 (*Sclerotinia sclerotiorum*)与细菌间的基因水平转移[J]. *应用与环境生物学报*, 2012, 18(1): 42-48.
- [16] Fears K P, Petrovykh D Y, Clark T D. Evaluating protocols and analytical methods for peptide adsorption experiments [J]. *Biointerphases*. 2013, 8(1): 20-33.
- [17] Quang Gao, Wujun Xu, Yao Xu, et al. Amino acid adsorption on mesoporous materials: influence of types of amino acids, modification of mesoporous materials, and solution conditions [J]. *J PhysChem B*, 2008, 112(7): 2261-2267.
- [18] Nonaka T, Fujihashi M, Kita A, et al. mcrystral structure of an oxidatively stable subtilisin-like alkaline serine protease, KP-43, with a C-terminal beta-barrel domain [J]. *Bi01 Chem*, 2004, 279: 47344-47351.
- [19] Mathé C, Devineau S, Aude J C, et al. Structural determinants for protein adsorption/non-adsorption to silica surface [J]. *PLoS One*, 2013, 8(11): e81346.
- [20] Mizi A, Zouros E, Moschonas N, et al. The complete maternal and paternal mitochondrial genomes of the mediterranean mussel *Mytilus galloprovincialis*: implications for the doubly uniparental inheritance mode of mtDNA [J]. *Mol Biol Evol*, 2005, 22(4):

- 952-967.
- [21] Yukawa H, Omumasaba C A, Nonaka H, et al. Comparative analysis of the *Corynebacterium glutamicum* group and complete genome sequence of strain R [J]. *Microbiology*, 2007, 153(4): 1042-1058.
- [22] Jenke-Kodama H, Dittmann E. Evolution of metabolic diversity: insights from microbial polyketide synthases [J]. *Phytochemistry*, 2009, 70: 1858-1866.
- [23] Teichmann S A, Park J, Chothia C. Structural assignments to the *Mycoplasma genitalium* proteins show extensive gene duplications and domain rearrangements [J]. *Proc Natl Acad Sci*, 1998, 95: 14658-14663.
- [24] Selvaraj T, Kim S K, Kim Y H, et al. The role of carbohydrate-binding module(CBM) repeat of a multimodular xylanase(XynX) from *Clostridium thermocellum* in cellulose and Xylan binding [J]. *Microbiol*, 2010, 48: 856-861.
- [25] Brenner S E, Hubbard T, Murzin A, et al. Gene duplications in *H. influenzae* [J]. *Nature*, 1995, 378: 140.
- [26] Gross F, Luniak N, Perlova O, et al. Bacterial type III polyketide synthases: phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expression in pseudomonads [J]. *Arch Microbiol*. 2006; 185(1): 28-38.

Sequence characteristics and system evolution of PPC domain from marine bacterial extracellular protease

HUANG Jia-feng, LIU Dan, YANG Xing-hao, WU Ri-bang, WU Cui-ling, HE Hai-lun
(School of Life Science, Central South University, Changsha 410013, China)

Received: May, 19, 2014

Key words: Protease; PPC domain; Conservative region; Phylogenetic analysis; Horizontal gene transfer

Abstract: PPC (Pre-Peptidase C-terminal) domain is widely distributed in the C-terminal region of protease secreted from marine bacterial, and it plays an important role in secretion and localization of protease, as well as adsorption of reaction substrate. The PPC domain sequence analysis showed that there are two conserved sequences in PPC domain amino acid sequence and β -sheet was characterized as spatial structure of the regions. Although some PPC domain sequences show lower consistency, there are conserved sequences and similar three-dimensional structure of PPC domain. In order to analyze the evolutionary relationship of the PPC domain between different strains, this paper did phylogenetic analysis on the amino acid sequences of 61 PPC domains from 39 different stains searched from the NCBI database. The study found that PPC domain sequences were variability and nonconservative in the biological evolution process. According to the results, some PPC domain may be derived from its gene duplication, while some may happen due to horizontal gene transfer event.

(本文编辑: 康亦兼)