

中国地区幽门螺杆菌 cagY 序列特征分析

王海滨 孙灵利^{通信作者}

北京市朝阳区疾病预防控制中心 北京 100021

〔摘要〕目的 分析中国地区幽门螺杆菌 (*H. pylori*) cagY 基因的核苷酸序列特征, 为 CagY 的功能学研究提供基础。方法 PCR 方法扩增 77 株中国地区 *H. pylori* 的 cagY 基因并进行测序, 使用 Vector NTI Suite 6 软件进行序列比对及多态性分析; 使用 Tandem Repeats Finder 软件分析短片段重复序列。结果 77 株中国地区 *H. pylori* 的 cagY 长度范围是 3534bp ~ 5394bp; 77 株中国 *H. pylori* 的 cagY 基因共携带有 6 类重复片段, 不同菌株中的重复片段的分布、重复片段的类型及排列顺序各异。结论 中国地区 *H. pylori* 的 cagY 基因携带重复片段, 使 cagY 基因序列表现出显著的多态性。cagY 的多态性可能参与了 *H. pylori* 的免疫逃避等致病机制。

〔关键词〕幽门螺杆菌; cagY 基因; 重复片段

〔中图分类号〕R378 〔文献标识码〕A 〔文章编号〕2095-7165 (2020) 07-009-04

Characteristic of cagY gene of *Helicobacter pylori* in China

WANG Hai-bin*, SUN Lin-li* Chaoyang Center for Disease Control and Prevention, Beijing 100021, China

〔Abstract〕Objective To analysis the characteristics of cagY genes of *Helicobacter pylori* (*H. pylori*), then provide data for the functional research of CagY. Methods cagY genes from 77 *H. pylori* which were isolated in China were amplified. PCR products were sequenced and analyzed using the Vector NTI Suite6 software for sequence alignment and polymorphism analysis; Tandem Repeats Finder software was used to analyze the short repeat segments. Results The range of cagY length is 3534bp ~ 5394bp. There are six types of repeats involved in these 77 strains. Different distributions, modes and orders of repeat segments were observed in different strains. The characteristic of the repeat segments are described as follows. (1) Repeat segments of 73 strains distributes in only one region, while the other 4 strains in two regions. (2) Each strain contains different types of repeat segments, and the repeat segments with the same length are different in nucleotide sequences. (3) The long repeat segments contains the short repeat segments. We also find that the core sequence of the repeat segments are two short segments with the lengths of 114-117bp and 93bp. The numbers and the distribution of the repeat segments are significant different among different strains. Conclusion Based on the different types of the short repeat segments, the cagY genes from 77 *H. pylori* which were isolated in China showed significant polymorphism. And the polymorphism may be associated with the immune escape or other pathogenic mechanisms.

〔Key words〕*Helicobacter pylori*; cagY gene; repeat segments

幽门螺杆菌 (*Helicobacter pylori*, *H. pylori*) 是一种革兰氏染色阴性, 微需氧、弯曲状细菌, 主要寄居在人胃部。大部分 *H. pylori* 感染者终生无症状, 部分感染者可发展成为慢性胃炎、消化道溃疡、粘膜相关淋巴组织淋巴瘤甚至胃腺癌。1994 年世界卫生组织国际癌症研究机构 (IARC) 正式将 *H. pylori* 感染列为第 I 类 (即肯定的) 人类致癌因子, *H. pylori* 是目前为止唯一被列为致癌因子的原核生物性病原微生物^[1]。

H. pylori 基因组包含一个 cag 致病岛 (cag pathogenicity island, cagPAI), 该致病岛包含 28-31 个 ORF, 与 *H. pylori* 的毒力密切相关。cagPAI 编码了细胞毒素相关蛋白 (cytotoxin-associated gene A protein, CagA) 和 IV 型分泌系统 (Type IV secretory system, T4SS) 的组分蛋白^[2]。T4SS 可介导 CagA 转运进入宿主细胞内, 进而诱发一系列病理性的信号通路。此外, 研究发现 *H. pylori* 感染人体后可以诱发强烈的宿主免疫反应, 但是 *H. pylori* 可以逃避这种免疫反应而不被清除^[3]。*H. pylori* 的这种免疫逃避机制依然未完全阐明。近年来研究发现, cagPAI 编码的蛋白 CagY 可能参与了 CagA 转运的早期阶段、*H. pylori* 的免疫逃避及介导 IL-8 的分泌。

国外学者通过对 *H. pylori* 菌株 26695、J99 和 NCTC11638 等的研究发现, 这些菌株的 cagY 序列中携带大量的重复序列, 而且这些重复序列在菌株之间表现为高度多态性。目前尚无针对中国菌株 cagY 的序列特征分析, 本研究选取了 77 株分离自不同患者的 *H. pylori*, 对其 cagY 进行了核苷酸测序, 并对其基因特征进行了描述, 为 CagY 的功能学研究提供基础。

1 材料和方法

1.1 实验菌株

本研究中选取的 77 株 *H. pylori* 分别来自中国的西部内陆

地区西安 (17 株)、东部沿海地区浙江 (18 株) 和东南沿海地区福建 (32 株) 及南部边疆地区云南 (10 株)。下文中分别以 XA、ZJ、FJ 和 YN 代表分离自四个地区的菌株。菌株由中国疾病预防控制中心传染病预防控制所诊断室分离。NCBI 公布的 23 株 cagPAI 测序菌株作为本研究的参考菌株 (见表 1)。

1.2 *H. pylori* 复苏培养及基因组 DNA 提取

取 -80℃ 冻存的临床分离 *H. pylori* 菌株涂布于绵羊全血哥伦比亚琼脂平板培养基上, 37℃ 微需氧气体环境 (5%O₂, 10%CO₂ 和 85%N₂) 培养 48h。观察生长情况, 再次传菌培养 48h, 经菌落形态、菌落涂片及生化鉴定 (快速尿素酶试验、氧化酶试验、触酶试验), 显微镜下确定为弯曲状的革兰氏阴性菌且三酶反应皆为阳性结果后收集菌苔。按 TIANamp Bacteria DNA Kit (天根公司) 操作说明提取 *H. pylori* 基因组 DNA。

1.3 扩增 cagY 基因

第一轮扩增: 将 cagY 分为 3 部分, 按照 NCBI 上提供的 23 株 cagPAI 序列, 使用 Primer premier 5.0 软件设计第一组引物 (见表 2)。第二轮扩增: 参照第一组引物的两端测序结果优化引物设计, 设计第二组引物, 共计 3 对 (见表 2); 将第一轮扩增失败的菌株进行第二轮扩增。第三轮, 针对依然无法扩增的片段采取小片段 PCR 扩增。依据已测序序列设计引物 (见表 2)。PCR 反应条件: 预变性 95℃ /5 min, 变性 95℃ /5 min, 退火条件见表 1, 延伸 72℃ /1.0 min, 循环 35 次, 最后延伸 72℃ /5min。

1.4 测序及序列分析

PCR 扩增产物由北京华大和上海生工测序公司进行双向测序。使用 Vector NTI Suite 6 软件进行序列比对、序列多态性分析及测序结果的拼接, 使用默认参数; 使用 Tandem Repeats Finder

分析短片段重复序列, 使用默认参数。统计学分析使用两独立样本 *t* 检验, *P* 值取 0.05。

2 结果

2.1 77 株中国 *H. pylori* 的 *cagY* 基因有显著的核苷酸序列多态性

77 株中国 *H. pylori* 的 *cagY* 长度范围是 3534 ~ 5394bp, 平均长度为 4467bp (见表 3); 23 株 NCBI 提交菌株的 *cagY* 长度在 4635 ~ 6009bp 之间, 平均长度为 5468bp, 中国四个地区的菌株与 NCBI 菌株相比有着较明显的差异 ($P < 0.05$)。

77 株中国 *H. pylori* 的 *cagY* 基因长度区间分布是: 小于 4000bp 的有 4 株 (5.2%), 4000 ~ 4500bp 之间的有 50 株 (64.9%), 4500bp ~ 5000bp 之间的有 19 株 (24.6%), 大于 5000bp 的有 4 株 (5.2%), 89.6% 的 *cagY* 基因长度位于 4500bp ~ 5000bp 之间。*cagY* 的长度多态性主要是因为不同的菌株携带了不同拷贝数及不同长度的重复片段引起的。Vector NTI Suite6 多序列比对显示, 在中国菌株的 *cagY* 基因内存在一段长度多态性区域, 以 0K129 为参照片段, 该区域位于 0K129 的 *cagY* 基因的 1986 ~ 3924bp 之间; 而 *cagY* 的两侧翼序列相对保守 (见图 1)。

2.2 77 株中国 *H. pylori* 的 *cagY* 基因携带有重复片段

使用重复片段分析软件共找到 9 类长度不同的重复片段: 114-117bp、207bp、231bp、324bp、435-438bp、93bp、106bp、390bp 和 424bp, 在中国菌株中前 6 类片段较为常见, 后 3 类片段较少见。

这些重复片段有以下特点: (1) 73 株 (约占 94.8%) 中国菌株的重复片段连续集中在 *cagY* 的一个区域内, 有 4 株菌株的重复片段分布于 *cagY* 的两个区域, 这部分菌株的第二个区域含有 390bp 的重复序列 (见图 2)。(2) 每株菌株的重复片段区域可包含一种或几种重复片段。例如: XA125 菌株 *cagY* 中含有 3 个拷贝的 117bp、2.5 个拷贝的 207bp、1.9 个拷贝的 117bp、2 个拷贝的 324bp 和 2.4 个拷贝的 390bp 等 4 类重复序列, 且两种 117bp 重复序列的核苷酸序列有显著差异。XA139 菌株的 *cagY* 中只含有 114-117bp 一类重复片段, 分别有 3 个和 4 个拷贝。这两种长度一致的重复序列也有差异。

(3) 较长的重复片段包含较短的重复序列组合。93bp、106bp 与 114-117bp 序列是所有重复片段的核心序列, 其中以 93bp 和 114-117bp 最为普遍, 在所有的 77 株中国菌株及 NCBI 的测序菌株都含有该片段, 而 106bp 只存在与个别菌株 (见图 3)。

2.3 不同菌株中的重复片段排列顺序不同

不同长度的重复片段的核心序列是 114-117bp 和 93bp, 为了便于描述, 定义 114-117bp 序列为 A 序列, 93bp 为 B 序列。分析不同菌株的重复序列区域发现: 不同的菌株携带重复片段的数目及排列方式有着明显的差异, 且重复片段内部的 A 序列与 B 序列的排列也是有差异的 (见图 4)。

本研究在不同长度的 *cagY* 序列中共选取了 10 条重复序列片段区进行阐述。重复序列模式图中显示, FJ2 和 ZJ371 只含有 3 个 114-117bp 重复片段核心序列, 而在 XA221 中有 15 个 114-117bp 序列和 5 个 93bp 序列共计 20 个重复片段核心序列。其他菌株也都含有 8-16 个数目不等的重复片段核心序列。以这些重复片段核心序列为基础构成了 438bp、207bp 等较长片段的重复序列。

从图 4 中可以看到, A 序列和 B 序列的排列方式有一定的规律性。重复序列区域一般以 A 序列起始, 一般每 1-5 个 A 序列后紧随一个 B 序列。以 3A+B 形式多见, 4A+B 和 5A+B 较少见。438bp 重复片段是由 3 个 A 片段和 1 个 B 片段组成的, A 片段和 B 片段在 438bp 内部的排列结构是不同的。FJ17 中的 438bp 序列的结构是 AAAB, FJ21 中的 435bp 是 AABA, FJ4、YN140、YN335 中的 435bp 是 ABAA, 而在 XA221 和 YN123 中有两种结构模式: AAAB 和 AABA。

3 讨论

cagY 位于 *cagPAI* 上游区域, *cagY* 长度变异较大, 中国菌株较 NCBI 测序株短 1kb 左右, 有一定的地域特点。*cagY* 最大的特点是其基因片段上含有大量的长短各异的重复片段, 包括 114-117bp、207bp、231bp、324bp、438bp、93bp、106bp、390bp 和 424bp 等类别, 中国菌株和 NCBI 测序株都含有这些重复片段, 在拷贝数方面差异较大。此外

由于同样长度的重复片段在不同的菌株内有着不同的序列, 使得 *cagY* 的核苷酸序列显得较为杂乱, 促使了 *CagY* 在大小及表型方面的差异。

Aras 等^[4]报道了当时仅有的 26695、J99 和 NCTC11638 等 3 株菌株的 *cagY* 基因重复序列特征。Aras 分析发现, 这三株菌株的 *cagY* 基因分为 5 个区域, 依次为: 5' 端重复序列区域 (FRR)、5' 端保守区 (FCR)、中间端重复序列区域 (MRR)、3' 端保守区 (TCR) 和 VirB10 同源序列区域 (VHR)。从目前的测序结果来看, 大部分中国菌株 (73/77) 及部分日韩菌株不含有 5' 端重复序列区域 (FRR)。基于目前更多的 *cagPAI* 测序, 经过本研究对目前的 *cagY* 测序株的比对发现, FRR 确实多见于西方型菌株, 而当时 Aras 分析的这三株菌株恰好是西方型菌株。据此推测, FRR 可能是与东西方地域性特征相关的一条序列。FRR 是否与 *H. pylori* 的毒力有关, 甚至是否与亚洲的高胃癌发病率有关, 值得进一步研究。

不同研究者对 *cagY* 上的重复片段的分析结果是不同的。2003 年 Liu 等^[5]在 1999 年较早的报道了 26695 的 *CagY* 存在重复的氨基酸序列, 发现重复的氨基酸序列主要分布在 2 个区域内, 分别命名为 Repeat I 区和 Repeat II 区。Repeat I 区含有 2 个紧密相连的 130aa 重复序列片段及一条 57aa 的片段, 这条 57aa 的片段恰好与 130aa 重复片段的前 57aa 相似。Repeat II 区的重复序列分为 5 类: δ (10aa)、 μ (13aa)、 α (14aa)、 ϵ (5aa)、 λ (13aa) 和 β (14aa), 这 5 类重复序列又组成了 3 种基序: $\alpha - \epsilon - \lambda$ (32aa, 96bp)、 $\beta - \delta - \mu$ (37aa, 111bp) 和 $\alpha - \delta - \mu$ (37aa, 111bp)。2008 年, Delahay 等^[6]对 NCBI 上的 14 条 *cagY* 序列进行了分析, 他将 Liu 等确定 Repeat II 区的重复基序进行了调整, 将 Repeat II 区中的重复氨基酸序列分为 2 种: A 基序 ($\delta \mu \alpha$) 和 B 基序 ($\epsilon \lambda \beta$)。A 基序有 38-39 个氨基酸残基, 约 114-117bp, B 基序有 31 个氨基酸残基, 约 93bp。针对 Repeat II 区的分析结果, 本研究与 Delahay 的分析结果基本一致, 同样发现了核心重复序列 114-117bp 和 93bp。与 Delahay 等分析结果不同的是, 本研究进一步分析了由核心重复序列 114-117bp 和 93bp 还组成了其他一些较长的重复序列, 并且本研究还发现了这两种核心重复序列在不同的菌株中的排列顺序是不同的。这种重复序列在拷贝数、长度及排列方式方面的差异导致了 *cagY* 序列长度的多态性及 *CagY* 表型的多态性。

重复序列的存在使 *cagY* 可以方便地接受外来基因片段并发生同源重组, 或者在基因组复制的过程中 *cagY* 的自身的基因片段发生同源重组^[7]。*H. pylori* 的这种特点为 *CagY* 提供了大量的表型多态性。基于 *CagY* 的多态性, 研究人员对 *CagY* 的功能做了如下推测: ① *H. pylori* 感染宿主后需要调整 *CagY* 的空间表达构象, 使之适应宿主细胞的受体, 这个过程是通过对重复序列的拷贝数及排列顺序的调整来实现的^[8]。② *CagY* 可能参与了 *H. pylori* 与宿主免疫系统的相互作用, 即 *CagY* 通过不断地变化表型来避免机体免疫系统的监视与识别, 从而导致了免疫逃避。③这种现象可以部分解释“患者根除 *H. pylori* 后容易发生重复感染”的问题。人体的体液免疫系统一般都是针对细菌外膜蛋白产生有保护作用的抗体, 而 *CagY* 正好位于 *H. pylori* 的细胞外膜。基于 *CagY* 的高度多态性, 当一种 *H. pylori* 被根除后, 针对此种 *H. pylori* 的 *CagY* 抗体无法抵御另一种 *H. pylori* 的感染^[4-6,9]。

CagY 的确切功能仍然没有完全阐明, 但是到目前为止, 在其他菌株的 T4SS 中尚未发现类似的重复片段。所以这种仅存在于 *H. pylori* 的重复序列的功能值得进一步的研究。

[参考文献]

- [1] Odenbreit S, Puls J, Sedlmaier B, et al. Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. *Science*, 2000, 287(5457):1497-1500.
- [2] Akopyants NS, Clifton SW, Kersulyte D et al. Analyses of the *cag* pathogenicity island of *Helicobacter pylori*. *Mol Microbiol*, 1998, 28(1):37-53.
- [3] Blanchard TG, Czinn SJ, Nedrud JG. Host response and vaccine development to *Helicobacter pylori* infection. *Curr Top Microbiol Immunol*, 1999, 241:181-213.

[4] Aras RA, Fischer W, Perez-Perez GI et al. Plasticity of repetitive DNA sequences within a bacterial (Type IV) secretion system component. J Exp Med, 2003, 198(9):1349-1360.

[5] Liu G, McDaniel TK, Falkow S, Karlin S. Sequence anomalies in the Cag7 gene of the Helicobacter pylori pathogenicity island. Proc Natl Acad Sci U S A, 1999, 96(12):7011-7016.

[6] Delahay RM, Balkwill GD, Bunting KA, Edwards W, Atherton JC, Searle MS. The highly repetitive region of the Helicobacter pylori CagY protein comprises tandem arrays of an alpha-helical repeat module. J Mol Biol, 2008, 377(3):956-971.

[7] Krinos CM, Coyne MJ, Weinacht KG, Tzianabos AO, Kasper DL,

表 1 NCBI 公布的 23 株 cagPAI 测序菌株

Table 1. The strains published on NCBI of which the cagPAI had been sequenced

菌株编号	NCBI 编号	cagY 长度 (bp)	菌株编号	NCBI 编号	cagY 长度 (bp)
26695	NC000915	5784	F16	AB120416	5391
J99	NC000921	5460	F17	AB120417	5463
AG1	NC008086	3887	F28	AB120418	5388
Shi470	NC010698	5565	F32	AB120419	6009
P12	NC011498	5706	F79	AB120420	5442
51	CP000012	4635	F80	AB120421	5685
52	CP001680	5697	OK101	AB120422	5433
3K	DQ985738	5784	OK107	AB120423	5394
908	EF195721	3480	OK109	AB120424	5307
2017	EF195722	3637	OK112	AB120425	5409
2018	EF195723	3202	OK129	AB120426	5784
G27	NC011333	5711			

Comstock LE. Extensive surface diversity of a commensal microorganism by multiple DNA inversions. Nature, 2001, 414(6863):555-558.

[8] Alvi A, Devi SM, Ahmed I et al. Microevolution of Helicobacter pylori type IV secretion systems in an ulcer disease patient over a ten-year period. J Clin Microbiol, 2007, 45(12):4039-4043.

[9] Jimenez-Soto LF, Kutter S, Sewald X et al. Helicobacter pylori type IV secretion apparatus exploits beta1 integrin in a novel RGD-independent manner. PLoS Pathog, 2009, 5(12):e1000684.

表 2 cagY 扩增引物及 PCR 反应条件

Table 2. The primers used for PCR of cagY and the PCR conditions

引物名称	引物序列 (5' —3')	退火条件
第一轮 1F	AGTTGATCCCGCTTGCCATAGAAC	63 °C /1 min
第一轮 1R	GACAGAGCGGCTATCATGAAGTGT	
第一轮 2F	GGATAACCTTTAGCCGCCATGT	60.2 °C /1 min
第一轮 2R	CTATGGTGAATTGGAGCGTGTG	
第一轮 3F	GTGGGGTTGTCAAGATGATGATCTG	57.7 °C /1 min
第一轮 3R	TCTTTGCTCCCTAAACTCC	
第二轮 1F	TAGGCGTGTGTCAATTTGAGAGA	62 °C /1min
第二轮 1R	GTGATAGCAAGCGTGGTTAATAGC	
第二轮 2F	GGATAACCTTTAGCCGCCATGT	61 °C /1min
第二轮 2R	CTATGGTGAATTGGAGCGTGTG	
第二轮 3F	TATTTATCTCTGACAAGAGGAGCT	55 °C /1min
第二轮 3R	TGTATCGTTCATGAACTCCTTTC	
第三轮 1F	GTCATCTAAGTTATCCAAGCACTCG	58 °C /1 min
第三轮 1R	AACCGATGAAGAACGAAACGAGT	
第三轮 2F	GCGGTTTTTGATACAATCCTTATAC	54 °C /1 min
第三轮 2R	TTTTGATGATGGCACATTCACCT	
第三轮 3F	GCTTTATCTTTTGTGGAGTTCTT	55 °C /1 min
第三轮 3R	AACTACTAGCTGATATGAGCGTCAA	

表 3 77 株中国 H.pylori 菌株的 cagY 基因长度

Table 3. The lengths of cagY of 77 H.pylori in China

菌株编号	长度 (bp)	菌株编号	长度 (bp)	菌株编号	长度 (bp)	菌株编号	长度 (bp)	菌株编号	长度 (bp)	菌株编号	长度 (bp)
FJ2	3534	FJ15	4407	FJ29	4413	XA167	4830	YN140	4407	ZJ460	4197
FJ3	3964	FJ16	4518	FJ31	3963	XA190	4407	YN156	4845	ZJ511	4131
FJ4	4404	FJ17	4644	FJ32	4407	XA194	4413	YN188	4521	ZJ546	4404
FJ5	4404	FJ18	4407	FJ33	4521	XA195	4404	YN333	4842	ZJ691	4404
FJ6	4407	FJ19	4404	FJ34	4407	XA203	4398	YN335	4407	ZJ715	4203
FJ7	4407	FJ20	4401	FJ35	4845	XA221	5394	YN377	4407	ZJ733	4635
FJ8	4407	FJ21	4845	XA1	4395	XA57	4410	YN386	5232	ZJ770	4200
FJ9	4401	FJ22	4857	XA125	4884	XA72	4845	ZJ062	4407	ZJ0707	4404
FJ10	4407	FJ23	4404	XA127	3876	XA8	4164	ZJ070	4293	ZJ0709	4200
FJ11	4404	FJ25	4200	XA139	4518	XA98	4404	ZJ212	4920	ZJ805	4092
FJ12	4407	FJ26	4839	XA144	4176	YN1101	4401	ZJ371	4200	ZJ916	4611
FJ13	4407	FJ27	4521	XA161	4407	YN1122	5229	ZJ383	4413	ZJ959	5163
FJ14	4404	FJ28	4401	XA163	4407	YN123	4947	ZJ448	4200		

长度多态性区域

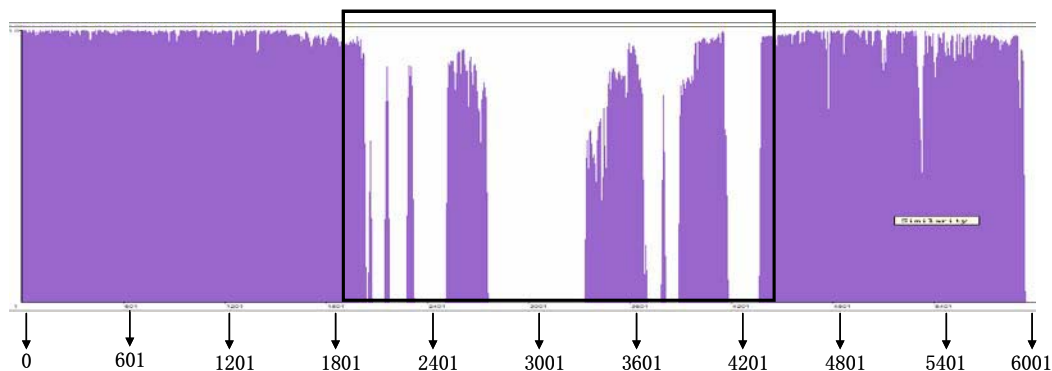


图 1 cagY 基因长度多态性区域

Fig 1. The diagram of the cagY repeat sequences polymorphism region

框内显示为 cagY 的重复片段长度多态性区域

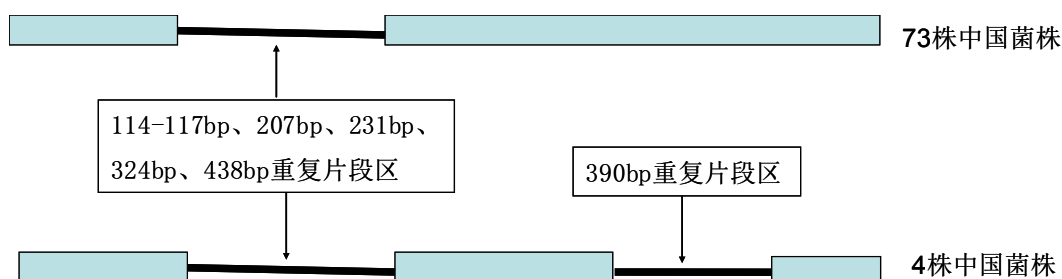


图 2 cagY 基因中重复片段序列区域分布模式图

Fig 2. Diagram of the repeat sequences regional distribution in cagY

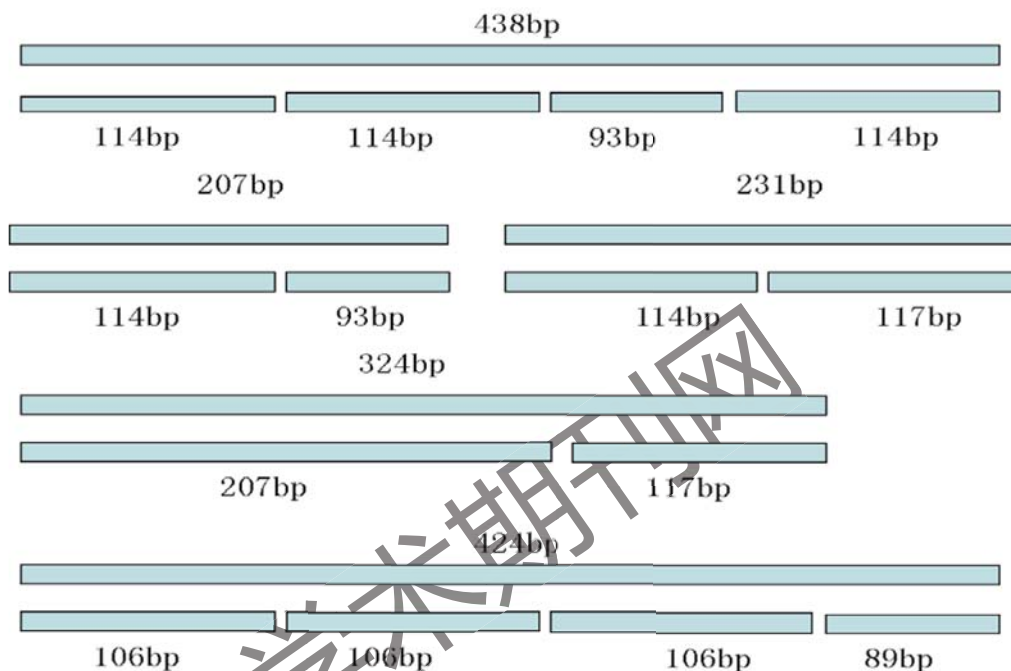


图 3 cagY 各类重复片段结构模式图

Fig 3. Diagram of all kinds of repeat fragments

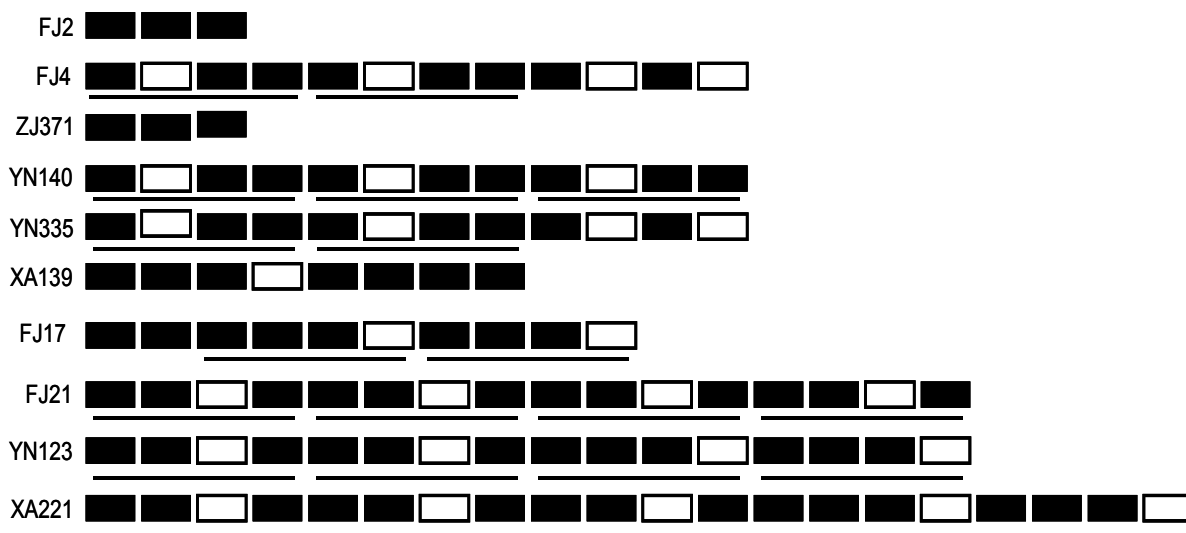


图 4 不同菌株 cagY 重复片段区 114-117bp 和 93bp 核心重复序列的排列方式

Fig 4. Arrangement of the 114-117bp and 93bp core repeat fragments in different strains

从上到下分别代表: FJ2 (3534bp)、FJ4 (4404bp)、ZJ371 (4200bp)、YN140 (4407bp)、YN335 (4407bp)、XA139 (4521bp)、FJ17 (4644bp)、FJ21 (4845bp)、YN123 (4947bp)、XA221 (5394bp)。黑色方块代表 114-117bp 序列, 白色方块代表 93bp 序列, 菌株名称后的括号内为 cagY 长度, 下划线代表一个 438bp 重复片段。