

副溶血弧菌菌株 229 基因组测序及其致病因子的注释

林毅英¹, 潘力¹, 吴清平², 王斌¹

(1. 华南理工大学生物科学与工程学院, 广东广州 510006)

(2. 广东省微生物研究所华南应用微生物国家重点实验室, 广东广州 510070)

摘要: 本文对从冰冻海鱼中分离到的副溶血弧菌菌株 229 进行基因组测序, 并对注释的基因进行功能聚类分析, 以分析其致病因子, 为研究其致病机制提供基因组全局水平的数据支持。完成了副溶血弧菌菌株 229 的基因组测序及组装, 所得基因组 4.97 Mb, GC 含量为 43.90%。基因注释获得 4940 个蛋白质编码基因 (CDS), 其中 1304 个 CDS 具有 GO 功能注释。WEGO 注释显示大量 CDS 对于副溶血弧菌入侵宿主及随后的代谢过程发挥重要作用。获得了与副溶血弧菌致病性相关的致病因子, 包括溶血素 (hemolysin)、III 型因子 (Type III secretion system, TTSS), 这些因子位于不同的染色体位点, 形成独立的致病因子簇。副溶血弧菌菌株 229 基因组组装和功能注释为研究副溶血弧菌致病性的遗传学基础提供了有价值的信息。

关键词: 副溶血弧菌; 基因组测序; 致病因子; 功能注释

文章编号: 1673-9078(2018)05-76-80

DOI: 10.13982/j.mfst.1673-9078.2018.05.011

Genome Sequencing of *Vibrio parahaemolyticus* Strain 229 and Annotation of Its Pathogenic Factors

LIN Yi-ying¹, PAN Li¹, WU Qing-ping², WANG Bin¹

(1.School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, China)

(2.State Key Laboratory of Applied Microbiology Southern China, Guangdong Institute of Microbiology, Guangzhou 510070, China)

Abstract: In this study, *vibrio parahaemolyticus* isolate 229 isolated from frozen sea fish was carried out by genome sequencing, and gene annotation was performed by gene functional analysis to elucidate its pathogenic factors, which could provide clues and genome-scale data for investigating its pathogenic mechanism. We have accomplished the genome sequencing of *V. parahaemolyticus* 229 and assembled its genome using the obtained paired-end sequencing reads generated on the Illumina Hiseq 2000 Genome Analyzer. The assembled genome was 4.97 Mb in size with GC content of 43.90%. 4,940 protein coding sequences (CDSs) were predicted by gene prediction, of which 1,304 CDSs were annotated with GO function. WEGO illustration revealed that a series of CDSs might play important roles in the invasion of the host of *V. parahaemolyticus* and subsequent metabolic processes. And an array of pathogenic virulence factor genes were annotated including hemolysin and Type III secretion system, which were gathered in different contigs and might act as independent pathogenic factors. The assembled genome and functional annotation of *V. parahaemolyticus* 229 could provide valuable information for further investigating the genetic basis of its pathogenicity.

Key words: *vibrio parahaemolyticus*; genome sequencing; pathogenic factor; functional annotation

副溶血性弧菌 (*Vibrio Parahaemolyticus*), 革兰氏阴性菌, 是人类常见的食源性肠道条件致病菌, 常见于冷冻海鲜类食品高消耗的国家。副溶血弧菌通常存

收稿日期: 2017-11-24

基金项目: 广东省科技计划项目 (2016A050503016、2016A010105004); 华南理工大学中央高校基本科研业务费资助项目 (2015ZP032); 广东省自然科学基金项目 (2017A030313097、S2012030006235)

作者简介: 林毅英 (1992-), 男, 硕士, 研究方向: 生物信息学

通讯作者: 王斌 (1981-), 男, 博士, 副教授, 研究方向: 基因工程及生物信息数据处理

在于海洋水域, 附着于鱼、虾、蟹及其他软件动物的体表生长^[1,2], 人类被其感染后会导致急性肠胃炎^[3-5], 有可能出现腹泻、脓血和脱水等症状, 甚至死亡。对于沿海地区经常食用海鲜制品的人们来说, 副溶血弧菌导致的食物中毒比率非常高^[6], 这类疾病的高发国家和地区包括美国及许多亚洲国家 (如日本、泰国、马来西亚和中国等)^[7,8]。副溶血弧菌在世界范围内的广泛分布, 使得研究其致病因子及致病机制变得非常重要, 以便更好地预防和治疗由其导致的食物中毒、肠胃炎等疾病, 也对检测食品中副溶血弧菌污染及食

品保鲜具有重要的实际应用价值。

虽然已有研究对副溶血弧菌的致病因子进行报道,例如Makino等研究表明副溶血弧菌的致病因子包括溶血素、脲酶、III型分泌因子TTSS、VI型分泌因子T6SS等^[3,9],但是副溶血弧菌引起肠胃炎的具体致病机制仍然不清楚。而近年来,随着抗生素滥用问题的日益加重导致了多重耐药菌株的出现^[8],这更加剧了副溶血弧菌引起肠胃炎等疾病的风险。因此持续监控海鲜制品捕捞、运输及货架期过程中的副溶血弧菌菌株污染情况,揭示新发现菌株的致病因子就显得十分重要。副溶血弧菌基因组测序能够获得全基因组水平的基因注释及功能分析数据,能为寻找其致病因子并分析其致病机制提供可靠的遗传学支持数据。

本研究所用的副溶血弧菌菌株 229,分离于冷冻海洋鱼类制品,其临床感染表现为腹痛、呕吐、腹泻及水样便等症状。本研究利用 Illumina Hiseq 2000 测序平台,制备 400 bp 测序文库,采用双末端 90 bp 测序技术,对其基因组进行测序。所得数据经质量过滤后,进行基因组组装、功能基因注释及基因 WEGO 功能聚类分析,目的在于获得与副溶血弧菌导致肠胃炎等疾病相关的致病因子,为研究副溶血弧菌致病性的遗传学机制提供有价值的数据库支持。

1 材料与方法

1.1 材料

1.1.1 菌株

副溶血弧菌菌株 229,用于基因组测序和功能基因注释,保藏于广东省微生物所菌株保藏中心,编号为 GM1xtf-229。

1.1.2 培养基

高盐 LB 培养基:胰蛋白胨 1%、酵母提取物 0.5%、氯化钠 3%、pH 7.0。培养温度 37 °C。

1.1.3 化学试剂

Tris-HCl、EDTA、SDS、巯基乙醇、RNaseA 等试剂购自北京普博欣生物科技有限公司;琼脂糖、核酸染料购自 TaKaRa 公司;细菌基因组 DNA 提取试剂盒 DNAzol™ Reagent,购自 ThermoFisher Scientific 公司;培养基及组分购自广东环凯微生物科技有限公司。

1.2 试验方法

1.2.1 文库构建及测序

副溶血弧菌菌株 229 在高盐 LB 培养基上,37 °C 培养 12 h 后,用细菌基因组 DNA 提取试剂盒 DNAzol™ Reagent 提取其基因组 DNA,溶于 100 μL

去离子水,并加入 20 μg/mL RNaseA 以去除残留的 RNA 成分。利用 Nano-drop 分析仪检测所得 DNA 样品的浓度和质量,达到 $OD_{260/280}=1.8\sim 2.0$ 的样品,用于基因组测序文库的构建:首先采用超声法 Covaris 将大片段 DNA 随机打断,经片段长度选择(400±25 bp)、平末端修复、末端加 A、接头连接等,获得带有测序接头的 DNA 片段,利用测序接头引物进行 PCR 扩增富集 DNA 片段,电泳回收 400±25 bp 的 PCR 扩增产物片段。利用 Illumina Hiseq 2000 测序平台进行 cluster 文库制备,采用双末端 90 bp 测序策略,在 Illumina Hiseq 2000 测序平台上进行副溶血弧菌菌株 229 的基因组测序。

1.2.2 基因组组装

原始的测序数据会存在一定量的低质量数据,为了使得后续的分析结果更加准确可靠,在进行基因组组装之前对原始的测序数据进行数据过滤处理。依次采用下列步骤对原始数据进行质量控制:(1)去除带有接头污染的reads(默认接头adapter序列与read序列有 15 bp的overlap,允许错配数为 1);(2)去除含N 碱基数目总和达到一定比例的reads(默认 10%,设置为 10 bp);(3)去除质量值连续≤4 的碱基数达到一定程度的reads(默认 40%,设置为 36 个)。经过滤后的 reads 数据,利用 Velvet 1.2.10 软件进行基因组从头组装,软件所用参数采取默认的参数,并根据测序数据进行优化组装结果的优化^[10]。

1.2.3 基因注释及功能聚类分析

蛋白质编码基因(coding sequence, CDS),利用软件Glimmer 3.02 进行预测分析^[11],并通过NCBI非冗余蛋白数据库进行验证。基因GO功能注释通过InterProScan软件完成,GO功能聚类分析通过WEGO 在线程序(<http://wego.genomics.org.cn/cgi-bin/wego/index.pl>)完成^[12]。tRNA的预测由tRNAscan-SE version 1.3.1 软件完成^[13]。菌株 229 基因组与副溶血弧菌模式菌株 10329 基因组的同源比对分析由circos软件完成。进化树构建,利用软件MEGA6 完成,采用Neighbor Joining方法建树,bootstrap参数设置为 1000(即构建 1000 棵进化树,在此基础上整合得到一个进化树)。

1.2.4 致病因子分析

根据目前文献报道的副溶血弧菌的主要致病因子,对菌株 229 组装的基因组及注释的功能基因进行比对分析,预测副溶血弧菌菌株 229 具有的致病因子,并分析这些致病因子的功能。

2 结果与讨论

2.1 基因组组装及统计分析

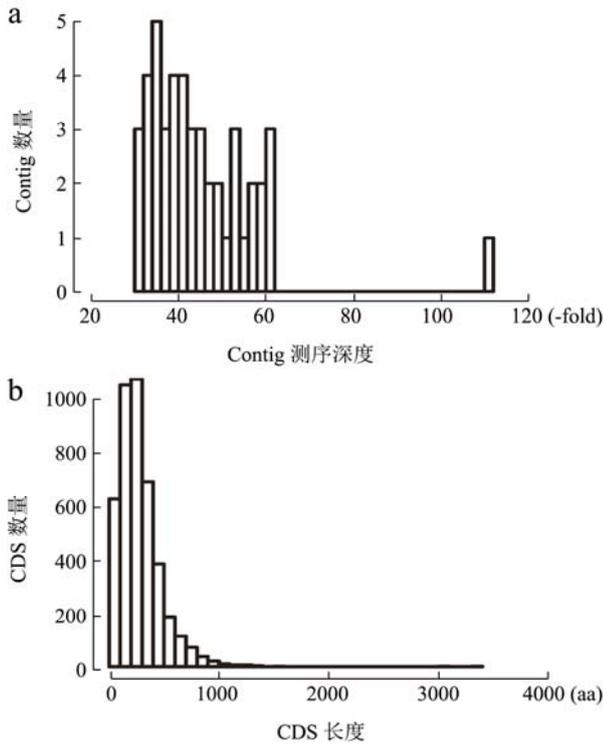


图1 副溶血弧菌 229 基因组 contig 覆盖度及蛋白编码基因 CDS 长度的统计分析

Fig.1 The statistics analysis of contig coverage and gene length (the amount of amino acids) in the genome of *V. parahaemolyticus* strain 229

经质量控制去除低质量及接头污染的 reads, 基因组测序共获得 14,003,024 个 90 bp reads, 用于基因组组装的关键参数 k-mer 数值设置为 71 bp. 组装所得基因组的序列文件提交至 NCBI 数据库, 获得序列号为 MRBW00000000. 原始测序 raw read 数据提交至 NCBI SRA 数据库, 获得序列号为 PRJNA356138.

基因组测序和组装统计信息如表 1, 共获得 63 个 contigs (>1000 bp, N50 为 45,033 bp), 最大的 contig 长度为 716,466 bp, 所得基因组长度为 4.97 Mb, GC 含量为 43.90%, 测序深度达到 254x, 大多数 contig 的覆盖度达到 30~60x (图 1), 所得测序数据能够满足基因组组装、功能基因注释和聚类分析的数据量要求。利用 Glimmer3.02 软件对组装的基因组序列进行功能基因注释, 共获得 4,940 个功能基因, 平均长度达到 933 bp, 多数基因的氨基酸数目少于 1000 个 (图 1), 具有功能注释的基因达到 1270 个, 为分析该菌株导致肠胃炎等疾病的致病因子提供了可靠数据。根据功能基因的平均长度以及预测的基因数目, 可以计算得到副溶血弧菌菌株 229 的功能基因的序列总长约为 4.61 Mb, 占其基因组序列总长 4.97 Mb 的 92.76%, 说明副溶血弧菌的基因密度很高, 基因组上非编码序列十分稀少, 也说明副溶血弧菌基因组具有高度的经

济性, 符合微生物生存及进化的基本原则: 减少基因组冗余序列, 降低代谢能耗。基因组包含 53 个 tRNA 基因, 这是副溶血弧菌蛋白质基因表达的基础。

表 1 副溶血弧菌菌株 229 的测序信息

Table 1 The sequencing information of *V. parahaemolyticus* strain 229

参数	数值
基因组大小/bp	4,965,081
基因组测序深度/fold	254
Contig 数目	63
contig 最大长度/bp	716,466
N50/bp	45,033
GC 含量/%	43.90
蛋白质编码基因 (CDS) 数量	4,940
功能基因数量	1,270
CDS 平均长度/bp	933
tRNA 基因数量	53

2.2 基因功能注释

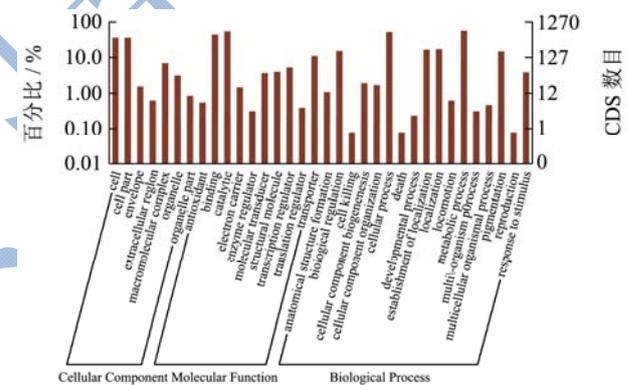


图 2 副溶血弧菌 229 注释基因的 GO 功能分析

Fig.2 GO functional analysis of annotated genes in the genome of *V. parahaemolyticus* strain 229

注: “Cellular component”, “molecular function”, “biological process”: GO 分类单元。

副溶血弧菌基因组中共有 1,270 个基因获得 GO 功能注释, 利用 WEGO 软件对具有 GO 注释的基因进行功能聚类分析。结果显示, 以下 GO 类群均富集了 10% 以上的功能基因: “binding”、“catalytic”、“biological regulation”、“cellular process”、“establishment of localization”、“metabolic process”、“pigmentation”等。尤其是以下两个 GO 类群富集了大量功能基因 (图 2): Biological process categories “metabolic process” (722 CDSs, 55.4%) and “localization” (226 CDSs, 17.3%), 这些基因可能有利于副溶血弧菌菌株对寄主的侵入及其随后在宿主上的生长、繁殖及代谢过程, 这些数据

为研究副溶血弧菌菌株的致病性的机制提供数据支持。

2.3 副溶血弧菌致病因子分析

根据文献报道，副溶血弧菌导致肠胃炎等疾病的致病因子包括溶血素(hemolysin)、III型因子(Type III secretion system, TTSS)和VI型因子(Type VI secretion system)。其中，热稳定的溶血素因子TDH、TRH是主要的致病因子，研究表明所有的临床副溶血弧菌菌株至少含有TDH、TRH二者中的一种致病因子^[14,15]。

基于基因的功能注释结果，副溶血弧菌 229 基因组包括一个溶血素基因 (NODE_11_orf00126) 以及另外三个与溶血素相关的基因 (NODE_24_orf00078、

NODE_2_orf00564、NODE_20_orf00001，表 2)，这是副溶血弧菌致病性的基础因子。III型因子是副溶血弧菌、志贺氏菌、沙门氏菌的重要致病因子之一，属于 Type III secretion system (TTSS) 系统。研究表明 TTSS 系统与副溶血弧菌侵入人类肠道上皮细胞的过程密切相关。对于副溶血弧菌菌株 229，本研究的基因组注释显示其 TTSS 系统包括四个功能基因 (NODE_8_orf00014、NODE_32_orf00049、NODE_2_orf00456、NODE_2_orf00457，表 2)。从这些基因在基因组上的分布情况看，这些致病因子位于不同的染色体位点，因此能够形成独立的致病因子簇。副溶血弧菌菌株 229 的基因组组装和注释为研究副溶血弧菌致病性的遗传学基础提供了有价值的数

表 2 副溶血弧菌 229 致病因子注释及功能分析

Table 2 Annotation and functional analysis of virulence factors in *V. parahaemolyticus* strain 229

致病因子	<i>V. parahaemolyticus</i> 229 同源基因	功能注释
溶血素 hemolysin		
溶血素	NODE_11_orf00126	Leukocidin/Hemolysin toxin family
溶血素相关基因	NODE_24_orf00078	AdipoR/Haemolysin-III-related
	NODE_2_orf00564	Haemolysin-type calcium-binding repeat
	NODE_20_orf00001	Hemolysin-type calcium-binding conserved site
III型因子 (TTSS)	NODE_8_orf00014	TTSS lipoprotein chaperone (YscW)
	NODE_32_orf00049	TTSS flagellar brake protein YcgR
	NODE_2_orf00456	Type III secretion system HrpE
	NODE_2_orf00457	ATPase, type III secretion system, Flii/YscN

2.4 副溶血弧菌菌株 229 基因组同源性分析

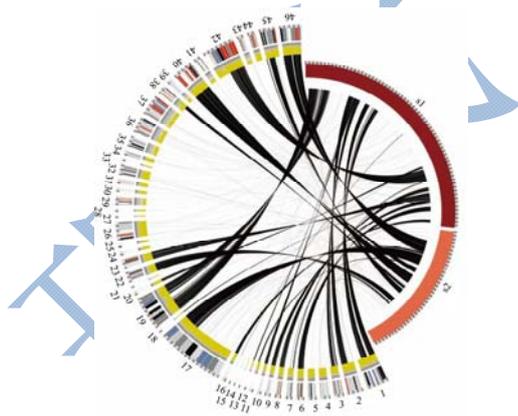


图 3 副溶血弧菌菌株 229 与 10329 的同源性比较分析

Fig.3 The homologous comparative analysis of *V. parahaemolyticus* strain 229 and 10329

注：s1、s2：副溶血弧菌菌株 10329 的 contig；1-46：副溶血弧菌菌株 229 的 contig；黄色圆环：G+C 含量；黑色连接线：两个菌株基因组中的同源性区域。

副溶血弧菌菌株 229 与模式菌株 10329 的同源性比较 (图 3) 表明，二者之间存在大量的同源序列，

说明二者同源关系较近，这有助于利用模式菌株 10329 的基因组信息和基因功能注释信息，分析副溶血弧菌菌株 229 的致病因子基因。同时，两株菌的同源序列存在大量的交错 (图 3)，也就是大量的同源序列在基因组上的位置是不一致的，导致这种差别的原因是多样的，其中之一是基因水平转移机制，这需要深入的遗传学分析加以确定。

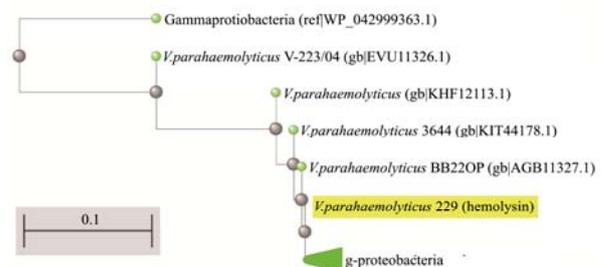


图 4 基于溶血素基因 (hemolysin) 构建副溶血弧菌菌株系统进化树

Fig.4 Phylogenetic distance tree for *V. parahaemolyticus* strains based on the hemolysin gene

基于溶血素基因 (NODE_11_orf00126) 的系统发育分析结果 (图 4)，表明副溶血弧菌菌株 229 与拥有

O4:K8 血清型的菌株BB22OP遗传距离最近^[16]。由于菌株BB22OP与临床菌株RIMD2210633 同源关系十分接近^[3,16], 推测菌株 229 有可能与菌株RIMD2210633 一样具有临床的致病性。

3 结论

3.1 副溶血弧菌菌株 229 的基因组组装及功能

注释

本研究完成了副溶血弧菌菌株 229 的基因组测序、组装和功能基因注释。组装所得基因组序列已提交至 NCBI 数据库, 序列号为 MRBW00000000。基因注释获得 4940 个基因, 其中 1270 个基因具有 GO 注释。WEGO 注释显示大量 CDS 对于副溶血弧菌入侵宿主及随后的生长、繁殖和代谢等过程发挥重要作用。

3.2 副溶血弧菌菌株 229 的致病因子分析

通过基因注释, 获得了与副溶血弧菌致病性相关的致病因子, 包括溶血素 hemolysin 和 III 型因子, 这些因子位于不同的染色体位点, 故而形成独立的致病因子簇。基于溶血素基因 (NODE_11_orf00126) 的系统发育分析结果表明, 副溶血弧菌菌株 229 与拥有 O4:K8 血清型的菌株 BB22OP 遗传距离最近, 推测菌株 229 可能具有临床的致病性。

参考文献

- [1] Letchumanan V, Chan K G, Lee L H. *Vibrio parahaemolyticus*: a review on the pathogenesis, prevalence, and advance molecular identification techniques [J]. *Frontiers in Microbiology*, 2014, 5: 705
- [2] Malcolm TTH, Cheah YK, Radzi CWJWM, et al. Detection and quantification of pathogenic *Vibrio parahaemolyticus* in shellfish by using multiplex PCR and loop-mediated isothermal amplification assay [J]. *Food Control*, 2015, 47: 664-671
- [3] Makino K, Oshima K, Kurokawa K, et al. Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae* [J]. *Lancet*, 2003, 361(9359): 743-749
- [4] Nair G B, Ramamurthy T, Bhattacharya S K, et al. Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants [J]. *Clinical Microbiology Reviews*, 2007, 20(1): 39-48
- [5] Newton A, Kendall M, Vugia D J, et al. Increasing rates of *vibriosis* in the United States, 1996-2010: review of surveillance data from 2 systems [J]. *Clinical Infectious Diseases*, 2012, 54(Suppl 5): S391-395
- [6] 刘秀梅,陈艳,王晓英,等.1992~2001 年食源性疾病暴发资料分析-国家食源性疾病监测网[J].*卫生研究*,2004, 33(6): 725-727
- [7] LIU Xiu-mei, CHEN Yan, WANG Xiao-ying, et al. Foodborne disease outbreaks in China from 1992 to 2001 national foodborne disease surveillance system [J]. *Journal of Hygiene Research*, 2004, 33(6): 725-727
- [7] Chen Y, Stine O C, Badger J H, et al. Comparative genomic analysis of *Vibrio parahaemolyticus*: serotype conversion and virulence [J]. *BMC Genomics*, 2011, 12: 294
- [8] Letchumanan V, Ser H L, Tan W S, et al. Genome sequence of *Vibrio parahaemolyticus* VP152 strain isolated from *Penaeus indicus* in Malaysia [J]. *Frontiers in Microbiology*, 2016, 7: 1410
- [9] Wang R, Zhong Y, Gu X, et al. The pathogenesis, detection, and prevention of *Vibrio parahaemolyticus* [J]. *Frontiers in Microbiology*, 2015, 6(32): 144
- [10] Zerbino D R, Birney E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs [J]. *Genome Research*, 2008, 18(5): 821-829
- [11] Delcher A L, Bratke K A, Powers E C, et al. Identifying bacterial genes and endosymbiont DNA with Glimmer [J]. *Bioinformatics*, 2007, 23(6): 673-679
- [12] Ye J, Fang L, Zheng H, et al. WEGO: a web tool for plotting GO annotations [J]. *Nucleic Acids Research*, 2006, 34(Web Server issue): W293-297
- [13] Lowe T M, Eddy S R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence [J]. *Nucleic Acids Research*, 1997, 25(5): 955-964
- [14] Shirai H, Ito H, Hirayama T, et al. Molecular epidemiologic evidence for association of thermostable direct hemolysin (TDH) and TDH-related hemolysin of *Vibrio parahaemolyticus* with gastroenteritis [J]. *Infection and Immunity*, 1990, 58(11): 3568-3573
- [15] Yamaichi Y, Iida T, Park K S, et al. Physical and genetic map of the genome of *Vibrio parahaemolyticus*: presence of two chromosomes in *Vibrio* species [J]. *Molecular Microbiology*, 1999, 31(5): 1513-1521
- [16] Jensen R V, Depasquale S M, Harbolick E A, et al. Complete genome sequence of pre-pandemic *Vibrio parahaemolyticus* BB22OP [J]. *Genome Announcements*, 2013, 1(1)