

综述

全基因组测序在食源性疾病监控和暴发调查中的应用

李柏生,柯昌文,张永慧

(广东省疾病预防控制中心,广东 广州 511430)

摘要:全基因组测序(WGS)分型在微生物的遗传与变异特征分析、菌株进化和暴发溯源调查中已经显示出了其极大的应用价值和发展潜力。本文简要概述了WGS的基本原理,介绍了基于全基因组的单核苷酸多态性分型和基于全基因组多位点序列分型两种主要的测序数据解读和分析方式,比较了WGS与脉冲场凝胶电泳技术、多位点序列分型等常见分子分型手段的主要特点,列举了典型的WGS应用于食源性疾病监控和暴发识别的案例,最后对WGS分型在食源性疾病监控和暴发调查领域的发展前景进行了展望。

关键词:全基因组测序;食源性疾病;暴发调查;溯源分析;食源性致病菌

中图分类号:R155 **文献标志码:**A **文章编号:**1004-8456(2016)02-0269-04

DOI:10.13590/j.cjfh.2016.02.027

Application of whole genome sequencing for the foodborne disease surveillance and outbreak investigation

LI Bai-sheng, KE Chang-wen, ZHANG Yong-hui

(Center for Disease Control and Prevention of Guangdong Province, Guangdong Guangzhou 511430, China)

Abstract: In the past few years, the whole genome sequencing (WGS) analysis has shown its great value and potentials in the field of microbial heredity and variation characteristics, evolution and source tracking. This paper reviews the basic principles of WGS and introduces two main methods to interpret and analyze the WGS sequencing data by wgSNP and wgMLST. Then, the main characteristics of WGS such as discriminatory ability and sensitivity were compared with the other common molecular typing methods of PFGE and MLST. Also, it's illustrated that the current state-of-art of the WGS applied to the foodborne disease surveillance, identification of outbreaks and source tracking. Finally, the advantages of the WGS and the prospective of its future application in the foodborne disease surveillance and identification of outbreaks were briefly expounded.

Key words: Whole genome sequencing; foodborne disease; identification of outbreaks; source tracking; foodborne pathogenic bacteria

食源性疾病监控和暴发调查主要通过发现异常疾病的聚集,准确、快速地识别病原体,发现传染源,以达到控制甚至终止疾病暴发的目的。20世纪90年代,美国疾病预防控制中心组建了微生物国家分子分型网络PulseNet,用于准确及时地发现和识别疾病暴发,在大肠杆菌O157和沙门菌暴发调查中发挥了重要的作用^[1-2]。脉冲场凝胶电泳技术(PFGE)是当前PulseNet采用的主流技术,通过限制

性内切酶消化细菌基因组后产生的DNA指纹图谱进行暴发识别和溯源分析。此外,细菌多位点序列分型(MLST)和多位点可变数目串联重复序列分析(MLVA)等分子分型方法,通过互联网数据库方便了不同实验室间的数据交流与共享,有利于快速进行大范围的暴发调查和溯源分析。但是,这些分型方法仅反映了病原菌基因组的局部遗传信息,分辨力有限,对于高度克隆化的菌株或进行大时间跨度的流行病学研究时稍显不足。

全基因组测序(WGS)分型,可在全基因组碱基序列的基础上进行流行病学分析,全面反应了病原菌的遗传与变异特征,提高了菌株进化和溯源分析的分辨力,有助于快速地确认并追踪病原体。随着新一代测序技术的成熟,WGS分型所需时间不断缩

收稿日期:2016-02-17

基金项目:广东省省级科技计划项目(2014A020219004)

作者简介:李柏生 男 副主任技师 研究方向为食源性疾病监测

E-mail:libsn@126.com

通信作者:张永慧 男 主任医师 研究方向为食品安全风险评估

E-mail:zyh@cdep.org.cn

短,成本不断降低,在食源性疾病监控和暴发调查中具有越来越明显的优势。本文将从病原菌全基因组测序分型的基本原理、主要特点和在食源性疾病监控和暴发识别中的应用等方面进行概述。

1 病原菌全基因组测序分型的基本原理

病原菌全基因组测序分型是指对病原菌整个基因组的核苷酸序列进行测序,并通过分析这些核苷酸携带的全部遗传信息而达到分子分型的目的。全基因组测序分型包含了病原菌全部的基因序列信息,可用于发现基因组结构的变异,发现毒力、耐药基因,发现可区分的核酸分子靶标,也可用于细菌分型和微进化研究^[3]。目前,病原菌全基因组测序分型主要有两种方式:基于全基因组的单核苷酸多态性分型(wgSNP)和基于全基因组的多位点序列分型(wgMLST)。

wgSNP分型方法主要有两种方式:基于标准参考序列的SNPs分析和基于K-mer的单核苷酸多态性(SNPs)分析。基于标准参考序列的SNPs分析通过与参考序列比对分析,根据覆盖率和阅读频率选择SNPs位点,从而获得高质量的SNPs,通过比较不同菌株之间的SNPs可以揭示菌株之间的种系发生关系。基于标准参考序列的SNPs分析可以获得菌株间良好的种系发生关系,有助于阐明出现SNPs位点的基因或内含子区域;关键在于如何从大量菌型中去获取合适的密切相关的标准参考序列,此外,还需要在计算机上进行多重比对分析。基于K-mer的SNPs分析,则是通过将测序数据切割成一序列K长度的序列,这些K-mer可以有13~35个碱基的差异,通过K-mer之间的比对筛选分析SNPs位点。该方法不需要将测序结果预先拼接组装,也不需要标准参考序列,运算速度比较快,但是没有考虑序列的质量,也无法对SNPs位点在基因组中的确切位置进行准确定位,无法鉴定相邻的两个SNPs位点。wgMLST分型方法以核心基因组为基础,纳入了更多的等位基因(一般可达1500~2500个),沿用经典MLST分析方法,以基因比对的方式在核心基因组中搜寻等位基因差异,可以获得比经典MLST更高的分辨率^[4]。wgSNP分析无法确定SNPs的产生是由于点突变还是发生重组,而wgMLST以等位基因变异为基础,基因重组或多个位点的插入或缺失都将被归为单个的进化事件,这种分析方式比仅仅考虑单个碱基的突变更适合于生物进化过程的分析。wgMLST对生物信息分析的要求较低,但方法的建立需要大量的菌株做基础数据进行评估,菌株数量的多少影响最终筛选出核心基因的数

量。病原菌种类、菌株来源和菌型的不同,也将影响到wgMLST方法的适用范围。此外,进行wgMLST分析对成千上百个等位基因的筛选,也需要合适的自动化筛选工具。

2 病原菌全基因组测序分型的主要特点

全基因组测序分型不同于PFGE、MLST、扩增片段长度多态性(AFLP)和多位点数目可变串联重复序列分析(MLVA)等常见的分子分型技术,这些技术针对的是病原菌基因组中的某一小部分变异信息,仅反映了病原菌基因组的少数遗传元件的变异和菌株的局部遗传信息,分辨力有限,尤其是对高度克隆化的菌株。如肠炎沙门菌,菌群本身具有高度克隆化,常规的分子分型方法对其缺乏足够的分辨力,85%的肠炎沙门菌可以归为5个PFGE型别^[5]。因此,及时有效地监控和识别沙门菌暴发,需要一个更为可靠的分型工具。2009年7月至2010年4月,一起由蒙得维沙门菌污染了制作意大利腊肠的原料胡椒而引起的食源性暴发波及美国44个州,导致272人发病。Bakker等^[6]采用病例对照研究的方法,选取20株暴发菌株和27株暴发前分离到的病例、食品、动物和环境菌株,这些菌株经*Xba*I内切酶消化后,具有完全相同的PFGE图谱,无法识别出暴发菌株,而通过wgSNP分析,不仅准确地识别出本次暴发相关的菌株,还识别出之前未发现的一些小型的暴发。这充分说明wgSNP分子分型在病原菌的暴发识别和溯源分析上比PFGE分型具有更高的分辨力。

与经典的MLST方法只对7个管家基因进行测序和分析比对不同,wgMLST对病原菌核心基因组上的成百上千个基因位点的序列差异进行区分和分型,获得比经典MLST更高的分型能力。Leopold等^[7]对18株流行病学背景清晰的耐甲氧西林金黄色葡萄球菌(MRSA)进行wgMLST分析,不仅能把具有直接流行病学关联的菌株聚集成簇,还能够将未发现直接流行病学关联的菌株区分开,揭示出经典的微生物分型方法不能揭示菌株间的传播关系。Ruppitsch等^[8]对单增李斯特菌17个血清型代表菌株和2起暴发菌株进行wgMLST分析方法应用评价,结果显示wgMLST分析方法对单增李斯特菌具有很好的分型力,可以很好地区分不同暴发的菌株以及与暴发相关或不相关的菌株。同时,wgMLST可以区分MLST型别和PFGE带型均一致但流行病学不相关的菌株。与wgSNP分析不同,wgMLST分型对生物信息分析的要求较低,其较好的分型能力在病原菌的分子分型和分子流行病学研究中具有

更广阔的应用前景。

与 PFGE 和 MLST 等分子分型方法只针对病原菌基因组中的某一部分遗传元件进行分析不同, wgSNP 可在整个基因组基础上对单个核苷酸的变异进行分析,因此,具有较高的分辨力和灵敏度。2010 年 10 月,海地首次暴发霍乱大流行。关于疫情的源头,曾引起各方诸多争论, Hendriksen 等^[9]对分离自尼泊尔、海地的霍乱弧菌进行了全基因组测序,并与全球其他地区分离的霍乱弧菌进行 SNPs 比对分析,结果发现,此次海地暴发菌株和尼泊尔菌株聚集成簇,两者之间只存在 1~2 个 SNP 的差异,提示海地霍乱暴发的菌株来源于可能尼泊尔。可见, wgSNP 分析可在全基因组的水平上对不同分离菌株进行精确甄别,达到精准确认暴发源头、明确传播途径的目的;甚至能够精确区分疾病大流行中不同暴发事件的先后顺序,这些是 PFGE 等常见的分子分型方法难以实现的,这些特点在疾病大流行控制中的作用至关重要。

3 病原菌全基因组测序在食源性疾病监控和暴发中的应用

在进行食源性疾病监控中,对病原体进行快速准确地分型鉴定是早期发现和预警食源性疾病暴发的关键因素。经典微生物分离鉴定和分型需要多种不同的技术手段以及大量的人力、物力和财力,并且分辨力不高,无法有效识别在时间、空间和型别上有交叉的暴发,更无法进行准确溯源。通过全基因组测序,比较基因组学和群体遗传学构建系统发育树,结合流行病学调查,可准确推测病原体之间的进化关系,为暴发识别和溯源分析提供依据。

在全基因组测序的基础上,找出编码和表达某一特定血清型的分子标记,并建立相应的检测和分型方法,可以辅助甚至替代传统的血清分型方法。Joensen 等^[10-11]将流行病学和病原体的实时聚类分析联系在一起,通过 WGS 分析建立了一个新的产 *vero* 毒素大肠杆菌 (VTEC) 分型和监测体系,能够及时准确地识别暴发。同时,他们还通过基因组流行病学中心 (CGE) 建立了一个基于 web (在 Http 协议基础之上,利用浏览器进行访问的网站) 的大肠杆菌血清分型服务器,可针对特定的 O 抗原和 H 抗原基因分析处理。通过该公共数据库,可获得 682 个大肠杆菌的全基因组数据,并完成血清分子分型。Dallman 等^[12]在产志贺毒素大肠杆菌 O157 的检测工作中,通过 WGS 分析快速发现了病例的关联和聚集性,并指引流行病学调查,还发现并确认了一些之前未发现的规模较小但分布广泛的聚类和

暴发,这说明与传统的分子分型方法相比,WGS 在食源性疾病监控工作中具有更高的分辨力、敏感性和实时性。

由于全基因组测序包含了病原菌全部的基因和信息,可用于充分的发现基因组结构的变异,包括插入缺失,以及完整的毒力和耐药基因的变化,对于病原体变异或重组产生的新型病原体有较高的实用价值。2011 年 5 月至 2011 年 7 月,德国大肠杆菌 O104:H4 暴发并迅速蔓延至整个德国和欧洲部分国家,导致近 4 000 人发病,近千人出现溶血性尿毒素综合征,超过 50 人死亡。流行病学调查将暴发来源指向埃及进口的豆芽,经典微生物学显示暴发菌株具有相同的亚型和表型特征。大肠杆菌 O104:H4 引起如此大规模暴发引发了微生物学家极大的兴趣,WGS 分析发现,引起此次暴发的大肠杆菌 O104:H4 携带有典型的肠聚集性大肠杆菌 (EAEC) 的毒力因子和分泌蛋白,菌株进化分析也显示出与典型的 EAEC 较近,而与肠出血性大肠杆菌 (EHEC) 较远的遗传关系^[13]。虽然此次 O104:H4 暴发菌株也产志贺毒素,常见于肠出血性大肠杆菌,但是 Rohde^[14]和 Rasko 等^[15]通过 WGS 分析,明确其为肠聚集性大肠杆菌,进一步的种系发生进化分析显示该菌株与非洲国家的不产志贺毒素的肠聚集性大肠杆菌不同,形成单独的进化分支。这表明,暴发菌株可能是通过水平转移的方式获取编码志贺毒素的噬菌体以及其他致病因子,从而引起这起严重的疾病暴发。

4 问题与展望

在食源性疾病监控和暴发调查中,WGS 正逐渐成为一个更加有效的鉴定工具,它有助于弄清菌株抗生素耐药和毒力产生的遗传基础、精确地推测耐药性、描述菌群结构、评估某种致病菌群的出现和传播机制,为病例的有效管理提供依据。WGS 还可用于新的抗生素以及疫苗研发,宏基因组的研究也使得研究者可以进一步了解包括人体肠道微生态在内的各种微生物生态系统^[16-17]。综上所述,WGS 将更加广泛应用于食源性致病菌的精确鉴定、特征性描述和进化分析。

WGS 分型的关键不在于如何得到基因组序列,而是如何对这些庞大的基因组数据进行快速的计算和解释,其结果不但可以直接与经典微生物分型方法(如 PFGE、MLST)进行比对,同时还可通过公共数据库方便全球共享和分析,通过比较基因组间的单核苷酸多态性,精确识别和区分食源性疾病暴发机制。未来还需进行更多的研究,

比如基于 SNPs 进化分析需要对相关微生物的变异进行更加详细的分析论证。尤其是在食源性疾病暴发调查中,若想通过简单的划定一个 SNPs 阈值来区分是否暴发存在一定的困难,需要通过进一步的调查和比较来对测序数据进行深入有效的解释^[3]。同时,WGS 的推广和使用还需要成熟易用的生物信息学分析流程和专家解释系统,并需要一线的临床医生、感染控制专家、实验室专家对 WGS 分析的接受和认同。

参考文献

- [1] Marder E P, Garman K N, Ingram L A, et al. Multistate outbreak of *Escherichiacoli* O157; H7 associated with bagged salad [J]. Foodborne Pathog Dis, 2014, 11(8) : 593-595.
- [2] Angelo K M, CHU A, Anand M, et al. Outbreak of *Salmonella* Newport infections linked to cucumbers-United States, 2014 [J]. MMWR Morb Mortal Wkly Rep, 2015, 64(6) : 144-147.
- [3] Kwong J C, McCallum N, Sintchenko V, et al. Whole genome sequencing in clinical and public health microbiology [J]. Pathology, 2015, 47(3) : 199-210.
- [4] Maiden M C, Jansen R M, Bray J E, et al. MLST revisited: the gene-by-gene approach to bacterial genomics [J]. Nat Rev Microbiol, 2013, 11(10) : 728-736.
- [5] DenBakker H C, Allard M W, Bopp D, et al. Rapid whole-genome sequencing for surveillance of *Salmonella entericaserovarenteritidis* [J]. Emerg Infect Dis, 2014, 20(8) : 1306-1314.
- [6] Bakker H C, Switt A I, Cummings C A, et al. A whole-genome single nucleotide polymorphism-based approach to trace and identify outbreaks linked to a common *Salmonella enterica* subsp. *enterica* serovar Montevideo pulsed-field gel electrophoresis type [J]. Appl Environ Microbiol, 2011, 77(24) : 8648-8655.
- [7] Leopold S R, Goering R V, Witten A, et al. Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes [J]. J Clin Microbiol, 2014, 52(7) : 2365-2370.
- [8] Ruppitsch W, Pietzka A, Prior K, et al. Defining and evaluating a core genome multilocus sequence typing scheme for whole-genome sequence-based typing of *Listeria monocytogenes* [J]. J Clin Microbiol, 2015, 53(9) : 2869-2876.
- [9] Hendriksen R S, Price L B, Schupp J M, et al. Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian outbreak [J]. MBio, 2011, 2(4) : e00157-11.
- [10] Joensen K G, Scheutz F, Lund O, et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli* [J]. J Clin Microbiol, 2014, 52(5) : 1501-1510.
- [11] Joensen K G, Tetzschner A M, Iguchi A, et al. Rapid and easy in silico serotyping of *Escherichia coli* using whole genome sequencing (WGS) data [J]. J Clin Microbiol, 2015, 53(8) : 2410-2426.
- [12] Dallman T J, Byrne L, Ashton P M, et al. Whole-genome sequencing for national surveillance of Shiga toxin-producing *Escherichia coli* O157 [J]. Clin Infect Dis, 2015, 61(3) : 305-312.
- [13] Cheung M K, Li L, Nong W, et al. 2011 German *Escherichia coli* O104: H4 outbreak: whole-genome phylogeny without alignment [J]. BMC Res Notes, 2011, 4(1) : 533.
- [14] Rohde H, Qin J, Cui Y, et al. Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104: H4 [J]. N Engl J Med, 2011, 365(8) : 718-724.
- [15] Rasko D A, Webster D R, Sahl J W, et al. Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany [J]. N Engl J Med, 2011, 365(8) : 709-717.
- [16] Gordon N C, Price J R, Cole K, et al. Prediction of *Staphylococcus aureus* antimicrobial resistance by whole-genome sequencing [J]. J Clin Microbiol, 2014, 52(4) : 1182-1191.
- [17] Behar A, Fookes M C, Goren S, et al. Whole genome analysis to detect potential vaccine-induced changes on *Shigella sonnei* genome [J]. Vaccine, 2015, 33(26) : 2978-2983.

欢迎投稿《中国食品卫生杂志》网址: www.zgspws.com