

病毒宏基因组学的研究概况及应用

李 晶,杨洪一,刘 崇

(东北林业大学 生命科学学院,黑龙江 哈尔滨 150040)

摘要:病毒宏基因组学(Viral metagenomics)是新兴起的研究环境中病毒多样性及其分子生物学信息的技术。此技术已在人类健康、动植物病毒检测、环境中病毒多样性分析方面得到广泛应用,对病毒的系统研究、病原体分析及预防具有重要意义。为拓展病毒宏基因组学的应用范围和现实作用,综述了病毒宏基因组学在动植物健康、环境问题以及病毒研究领域研究概况及其应用展望。

关键词:病毒宏基因组学;高通量测序;随机引物;PCR

中图分类号:S852.65 **文献标识码:**A **文章编号:**1002-2767(2015)12-0182-04 **DOI:**10.11942/j.issn1002-2767.2015.12.0182

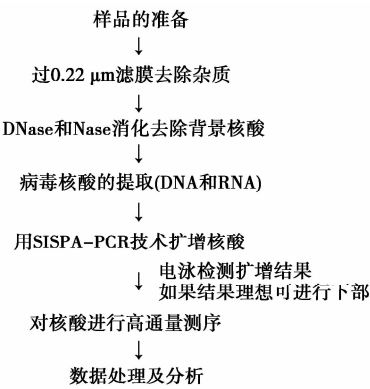
病毒个体微小,多数病毒直径在100 nm(20~200 nm),较大的病毒直径 300~450 nm,较小仅为 18~22 nm,结构简单,不能独立复制需要依赖于宿主细胞复制繁殖,被许多生物学家认为是处于生命和非生命交叉区域的存在物。据估计目前对病毒的发掘还不到 1%^[1],对病毒的研究具有广阔的前景和现实意义。病毒独特的结构和特性给病毒的研究和鉴别带来许多困难,主要体现在两个方面:第一,病毒没有专门的宿主细胞系,60%以上的病毒无法成功的进行离体培养^[2]或在培养中不能表达致病性;第二,病毒基因本身变异率高,通过与宿主间的相互作用进化,增加核酸多样性,产生新病毒,导致宿主范围扩大、跨物种传播^[3]。对细菌的研究可以通过保守的 16sRNA 的分析来定位分类信息、进化关系和种群多样性等。对于真菌有 18sRNA 及 ITS 序列。然而病毒不像细菌真菌,没有固定保守的进化标记基因。所以一些传统研究方法的应用受到限制,不能完全满足病毒研究的需要。如电镜观察病毒的灵敏性不高,细胞培养病毒可能观察不到细胞病变,血清学反应中不但难以获得高价抗体而且容易出现交叉反应导致错误结果,传统 PCR 方法对未知序列及高变异的病毒研究难以发挥作用。加之近年来病毒流行病的频繁发生及其可怕的传染性,对人类及动植物的健康产生严重威胁,如 HIV 病毒、SARS 病毒、禽流感病毒和在西非等地肆虐的埃博拉病毒^[4]等,给人们造成了巨大的恐慌和经济损失。因此,对病毒基因组的研究、致病源的探

索、病毒在生物体和环境如何存在及传播、病毒病防治的研究已迫在眉睫。

随着时代的发展和生物科学技术的进步,新兴的病毒宏基因组学为解决这些问题提供了契机,宏基因组学(Metagenomics)的概念是 1998 年由 Handelsman^[5]首次提出,对特定环境中基因组的总和进行研究,包括培养的和未培养的微生物。病毒宏基因组学(Viral metagenomics)就是宏基因组学在病毒领域的应用,即从环境或生物组织中浓缩病毒粒子的遗传物质进行生物学信息分析的技术。它的应用需要一些交叉学科的创新技术的支持,随机引物 PCR 和新一代测序技术——高通量测序的应用大大提高了研究的效率和获取信息的丰度,克服大环境中病毒浓度低、易受干扰的不足,拓展了病毒宏基因组学的应用范围和现实作用,为探索未知病毒提供广阔的前景和应用空间。在人类预防疾病、开发疫苗方面具有重大贡献。

1 病毒宏基因组学的研究过程

对于未知病毒的研究过程如下:



收稿日期:2015-09-18

第一作者简介:李晶(1989-),女,黑龙江省青冈县人,在读硕士,从事微生物研究。E-mail:1401787417@qq.com。

SISPA 方法是 1991 年 Gregory 和 Jung 在随

机引物 PCR 方法的基础上创造的^[6], SISPA-PCR 使用含有已知片段的随机引物进行逆转录,这个已知片段在接下来的 PCR 反应中将作为引物^[7],此方法先后经 Breitbart^[8]和 Djikeng^[9]等人的改进,在 SISPA 的基础上建立了 RNase、DNase-SISPA 联用的方法,增加了样品过滤和 DNA 酶 RNA 酶消化等步骤,去除外源污染,发现新病毒。高通量测序(High-throughput sequencing)也称第二代测序技术或深度测序^[10],可以对数百万个 DNA 分子进行同时测序,一次可同时输出大量数据,打破样品类型局限,最大限度的富集病毒核酸信息。

样品通过离心、过滤、核酸酶处理等步骤去除病毒粒子以外的物质和背景核酸的干扰^[11],纯化和富集病毒粒子,降低错配率,减少数据量,提高下游分析速率。所以此步骤的富集效果对后来的高通量测序结果的序列数量有决定性影响。将样品中的病毒粒子富集到 $106 \text{ 个} \cdot \text{mL}^{-1}$ 以上为理想状态^[12]。然后进行核酸 DNA 和 RNA 的提取。病毒基因组比真核和原核宿主短,在核酸制备过程中要尽量去除污染,一个小的污染可能导致长基因优先测序^[13],掩盖病毒基因,可能导致宏基因组分类组成成分的偏差。选取几种随机单引物对提取的核酸进行反转录,用 Klenow 酶合成 cDNA 的第二条链,然后用相应的引物进行 PCR 扩增,增加样品中低丰度病毒的检出率。不同的随机扩增方法对不同的核酸类型(线型、环型、ss/dsRNA/DNA)有不同的效率^[14],应用多种扩增方法和引物可以优化检测大环境中可能的病毒核酸的灵敏度,增加覆盖率。产物进行凝胶电泳检测,电泳条带呈现弥散状,送生物公司进行高通量测序。测序结果产生大量数据,数据的筛选处理和分析是整个过程的又一关键步骤,如果数据分析不好,整个实验就不能得到理想的效果,那么海量的测序数据也就失去了意义。

病毒宏基因组的分析流程与一般宏基因组数据分析流程相似,包括原始数据的预处理、基因组装、功能分类和基因预测等。但病毒序列存在高变异性。测序序列中也可能存在原病毒的序列,原病毒以溶原态存在于宿主染色体内,这将会影响病毒基因组与宿主的分离^[15]。

对于发现新病毒可以将单个序列和 contigs 在 blastn 的 nr 数据库中比对,如果比对结果无法判断出是何种序列,则在 blastx 的 nr 数据库中比

对,比对的 E 值如果大于 e^{-5} 则被认为是目前无法确定的生物的基因序列。如果比对的结果小于 e^{-5} 又与其最近的那一条的相似性较低,则可能是我们需要关注的目标^[3]。

2 病毒宏基因组学方法的优势和局限

2.1 优势

传统方法只能以已知病毒为目标,难以发现新病毒,而病毒宏基因组学方法结合新一代测序技术和随机 PCR 技术能够挖掘环境中海量未知的病毒;研究的病毒直接从环境中获取,不需要分离培养;可以对环境中过于分散、丰度低的病毒进行系统分析和鉴定。

2.2 局限

应用病毒宏基因组学方法进行研究,依然存在一些问题需要进一步探索,如从大量环境中提取样本方式,随机扩增方法和引物的选择能否做到无偏差覆盖所有病毒,海量测序数据的处理,依赖样品复杂度的生物信息分析等。

3 病毒宏基因组学的应用

病毒宏基因组学以其显著的优势在病毒研究的各个方面得到广泛应用,如人类、动植物健康、水、环境问题、病毒领域的研究等等。改变着人们对病毒的认识,给涉及病毒的许多问题提供新的解决思路。

3.1 临床应用

2008 年^[16]病毒宏基因组学在人体临床检查中第一次应用,结合深度测序检测器官移植相关的致死疾病的病原体。利用病毒宏基因组学对人体的唾液、呼吸道、血液、排泄物的检测分析,可以发现未知和潜在的致病病毒。Law 等^[17]对一些肝病患者的血浆进行研究,发现病毒宏基因组学方法快速准确,还能将检测范围扩大到其它体液。单同领^[18]根据病毒宏基因组学的方法分析了儿童和猪肠道病毒群落,并且确定了致病原的信息。Zhang 等^[19]应用宏基因组学在人的粪便中发现了大量的植物病毒。Willner 等^[20]应用宏基因组学分析了健康和呼吸道感染的人的呼吸道分泌物中的 DNA 病毒群落。2010 年^[21]在没有先验信息的条件下,应用病毒宏基因组学获得了流感病毒的全部基因。研究者还通过监测特定区域内已知传染人、动植物病毒的昆虫媒介携带的病毒^[22],来检测相关病毒的流行和预防,为突发病原体的鉴定提供方向。蝙蝠作为很多人兽共患

病(例如埃博拉病毒、严重急性呼吸系统综合征(SARS)病毒和尼帕病毒)的自然传播宿主,故蝙蝠体内的病毒多样性成为学者研究的热点。2010年,Li等^[23]和Donaldson等^[24]应用病毒宏基因组学的方法分别研究了北美蝙蝠粪便中的病毒群落。发现了大量新的哺乳动物病毒、昆虫病毒和植物病毒,Donaldson还发现了一株新的冠状病毒。2013年,杨凡力等^[25]应用病毒宏基因组学的方法研究吉林、云南、湖南采集的蝙蝠组织,发现蝙蝠体内有细小病毒、博卡病毒、腺病毒、小双节RNA病毒、圆环病毒等新病毒。2000年用病毒宏基因组学方法对埃博拉病毒进行了鉴定^[26]。

3.2 环境问题

病毒能够抵抗很多常规水处理系统在水中顽固存留,利用病毒宏基因组学快速检测水环境中的病原体,得到的宏基因组数据可以作为排泄物污染的指标,以此检测水质。Rosario等^[27]应用宏基因组学在废水处理厂的污水中发现了大量的来自动植物和人的致病性病毒,从而认为污水可能是病毒传播的一种途径。

3.3 病毒研究

目前科学家正在发展功能病毒宏基因组学,来发掘新病毒的功能酶^[28],用于诊断和生物科学研究。利用病毒宏基因组学可以开发不同环境中的各种病毒如海洋、温泉、岩石层、地下和高温环境等,为研究病毒学的分类和进化提供方法。

通过对环境病毒宏基因组学的研究,可以了解环境中病毒的有机构成、进化关系,进而掌握病毒的分布、变化动态和进化情况,追踪一些病原体的原始自然宿主,以便及时防控和预测一些常发和新发病毒病,还可以通过改变病变环境中的微生物群落来对抗病理性群落,从而达到改善病情甚至是治疗的效果^[29]。

4 病毒宏基因组学的研究意义

病毒宏基因组学的应用为病毒生态学在研究遗传潜力、病毒群落组成和结构以及环境病地理学等问题提供广阔的前景和方向。病毒宏基因组学还涉及了病毒和宿主之间的相互作用,病毒一些裸露在外的基因可能以某种未知的方式操控宿主,这对一些疾病的致病机理的分析、制定有效的治疗方法具有重大意义。病毒宏基因组学发现病毒的潜力能够推动许多相关学科的发展,如进化生物学、病原体的检测和生物技术。

病毒宏基因组学对一定环境内的全部病毒的研究起到了里程碑式的作用,此环境可以是海洋、土壤等无机环境,也可以是各种生物体的组织等有机环境。在病毒研究领域具有划时代意义

5 展望

随着生物科学技术的发展,一些交叉学科创新技术的发现,能够更加完善病毒宏基因组学的研究体系,例如搭建更为丰富的微生物基因组数据库,开发病毒宏基因组学检测领域的分析工具,将宏基因组学与宏转录组学、宏蛋白组学等技术结合起来在组学研究的多个层面上探索微生物的多样性等,使此项技术能够更广泛更准确的应用于病毒相关领域。通过对人体内微生物种群多样性的研究来探索其与某些疾病、耐药情况^[30]的可能关系。此技术还可以用来发现新的酶物质、抗生素以及一些有用的次生代谢产物。地球上海洋面积约占总面积的72%,富含丰富的资源,利用宏基因组学对海洋微生物的开发和利用具有重要意义。

期待未来的某一天能够利用此技术攻克如HIV病毒、SARS病毒、禽流感病毒和埃博拉病毒等高传染性、高致死率、严重威胁全人类生命健康的病毒病。

参考文献:

- [1] Mokili J L,Rohwer F,Dutilh B E. Metagenomics and future perspectives in virus discovery[J]. Virology,2012,2(1):63-77.
- [2] 何彪,涂长春. 病毒宏基因组学的研究现状及应用[J]. 畜牧兽医学报,2012(12):1865-1870.
- [3] 李金松. 应用非序列依赖单引物 PCR(SISPA-PCR)技术筛查腹泻相关的病毒[D]. 北京:中国疾病预防控制中心,2010.
- [4] 许黎黎,秦川. 埃博拉出血热动物模型研究进展[J]. 中国比较医学志,2010,20(9):67-71.
- [5] Handelsman J,Rondon MR,Brady SF,et al. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products[J]. Chem Biol,1998,5(10):R245-R249.
- [6] Reyes G R,Kim J P. Sequence-independent, single-primer amplification (SISPA) of complex DNA populations[J]. Mol Cell Probes,1991,5(6):473-481.
- [7] Froussard P. A random-PCR method (r PCR) to construct whole c DNA library from low amounts of RNA[J]. Nucleic Acids Res, 1992,20(11):2900.
- [8] Breitbart M,Rohwer F. Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing[J]. Biotechniques,2005,39(5):729-736.

- [9] Djikeng A, Halpin R, Kuzmickas R, et al. Viral genome sequencing by random priming methods[J]. BMC Genomics, 2008, 9(5): 1-9.
- [10] 王兴春, 杨致荣, 王敏, 等. 高通量测序技术及其应用[J]. 中国生物工程志, 2012(1): 109-114.
- [11] Thurber R V, Haynes M, Breitbart M, et al. Laboratory procedures to generate viral metagenomes[J]. Nat Protoc, 2009, 4(4): 470-483.
- [12] Djikeng A, Halpin R, Kuzmickas R, et al. Viral genome sequencing by random priming methods[J]. GMC Genomics, 2008, 9(1): 5.
- [13] Fancello L, Raoult D, Desnues C. Computational tools for viral metagenomics and their application in clinical research[J]. Virology, 2012, 434: 162-174.
- [14] Li L, Deng X, Mee E T, et al. Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent[J]. J Virol Methods, 2015, 213: 139-46.
- [15] 浦绍艳, 张鑫磊, 蒋太交, 等. 宏基因组学研究方法及其在病原体检测中的应用[J]. 生物物理学报, 2014(1): 3-14.
- [16] Palacios G, Druce J, Du L, et al. A new arenavirus in a cluster of fatal transplant-associated diseases[J]. N Engl J Med, 2008, 358(10): 991-998.
- [17] Law J, Jovel J, Patterson J, et al. Identification of hepatotropic viruses from plasma using deep sequencing: a next generation diagnostic tool [J]. PLoS One, 2013, 8(4): e60595.
- [18] 单同领. 病毒宏基因组学分析儿童和猪肠道病毒群落及 23 株病毒的初步研究[D]. 上海: 上海交通大学, 2011.
- [19] Zhang T, Breitbart M, Lee WH, et al. RNA viral community in human feces: prevalence of plant pathogenic viruses[J]. PLoS Biol, 2006, 4(1): e3.
- [20] Willner D, Furlan M, Haynes M, et al. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals[J]. PLoS One, 2009, 4(10): e7370.
- [21] Greninger A L, Chen E C, Sittler T, et al. A metagenomic analysis of pandemic influenza A (2009H1N1) Infection in patients from North America [J]. PLoS One, 2010, 185(10): e13381.
- [22] Ng TFF, Duffy S, Polston J E, et al. Exploring the diversity of plant DNA viruses and their satellites using vector-enabled metagenomics on whiteflies[J]. PLoS One, 2011, 6: e19050.
- [23] Li L, Victoria J G, Wang C, et al. Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses [J]. J Virol, 2010, 84 (14): 6955-6965.
- [24] Donaldson E F, Haskew A N, Gates J E, et al. Metagenomic analysis of the viromes of three North American bat species: viral diversity among different bat species that share a common habitat[J]. J Virol, 2010, 84(24): 13004-13018.
- [25] 杨凡力, 王意银, 郑文成, 等. 中国部分地区蝙蝠携带病毒的宏基因组学分析[J]. 生物工程学报, 2013, 29(5): 586-600.
- [26] Sullivan N J, Sanchez A, Rollin P E, et al. Development of a preventive vaccine for Ebola virus infection in primates [J]. Nature, 2000, 408(6812): 605-609.
- [27] Rosario K, Nilsson C, Lim Y W, et al. Metagenomic analysis of viruses in reclaimed water[J]. Environ Microbiol, 2009, 11(11): 2806-2820.
- [28] Schoenfeld T, Liles M, Wommack K E, et al. Functional viral metagenomics and the next generation of molecular tools[J]. Trends Microbiol, 2010, 18: 20-29.
- [29] Willner D, Furlan M, Haynes M, et al. Metagenomic Analysis of Respiratory Tract DNA Viral Communities in Cystic Fibrosis and Non-Cystic Fibrosis Individuals[J]. PLoS One, 2009, 4(10): e7370.
- [30] Maurice C F, Haiser H J, Turnbaugh P J. Xenobiotics Shape the Physiology and Gene Expression of the Active Human Gut Microbiome[J]. Cell, 2013, 152(1): 39-50.

Research and Application of Viral Metagenomics

LI Jing, YANG Hong-yi, LIU Chong

(College of Life Science, Northeast Forestry University, Harbin, Heilongjiang 150040)

Abstract: Viral metagenomics is the rise of new technique, which research viruses' diversity in the environment and their molecular biology information. Viral metagenomics was widely used on human health, animal and plant virus detection, environment diversity analysis, the systematic study of the virus, the pathogen analysis and prevention had great significance. In order to extend the application scope and reality function, the research and application prospects of viral metagenomics on animal and plant health, environmental issues, as well as virus general situation were reviewed.

Keywords: viral metagenomics; high throughput sequencing; random primers; PCR