

# 多位点测序分型技术在病原微生物分型鉴定中的应用

刘金华<sup>1,2</sup> 贺丹<sup>1</sup> 杨艳秋<sup>3</sup> 张波<sup>3</sup> 王丽<sup>1\*</sup>

(吉林大学基础医学院病原生物学教研室 长春 130021) (吉林出入境检验检疫局技术中心 长春 130062)

(吉林大学第一附属医院 长春 130021)

**摘要** 多位点测序分型(Multilocus sequence typing, MLST)技术是一种以核苷酸序列为基础的病原菌分型方法,它是高通量测序技术与成熟的群体遗传学相结合的产物。该方法简单易行,重复性强,可以通过国际互联网对某一致病菌株在全球范围内的传播分布情况进行追踪监控。目前,MLST技术已被广泛应用于原核病原菌及一些真核病原菌(如真菌)的分型鉴定中。主要对MLST技术的原理及其在一些常见病原菌分型鉴定中的应用进行了简要的阐述。

**关键词** 多位点测序分型,病原菌,分型鉴定

中图分类号:Q93 文献标识码:A 文章编号:0253-2654(2007)06-1188-04

## Application of Multilocus Sequence Typing method on Pathogenic Microorganisms Typing and Identification

LIU Jin-Hua<sup>1,2</sup> HE Dan<sup>1</sup> YANG Yan-Qiu<sup>3</sup> ZHANG Bo<sup>3</sup> WANG Li<sup>1\*</sup>

(Pathogenic Microorganism Department, Basic Medical College of Jilin University, Changchun 130021)

(Jilin Entry-Exit Inspection and Quarantine Bureau, Changchun 130062)

(Jilin University No.1 Clinical Hospital, Changchun 130021)

**Abstract** It is important for identifying, tracking and preventing pathogenic microorganism's outbreak in the epidemiology of infectious diseases. Establish a useful simple method for typing and identifying pathogenic microorganisms is necessary. In 1998, Multilocus sequence typing was proposed which is a nucleotide sequence based approach. It combined developments in high-throughput sequencing techniques with established population genetics. This typing method is a portable and reproducible typing system which can track and investigate pathogenic microbes distributed condition over the global by the internet. Now, MLST schemes have been developed for a variety of prokaryotic and some eukaryotic pathogens (like some fungus). In this paper, the theory of MLST and its application on some pathogenic microbes typing and identifying were introduced.

**Key words** Multilocus sequence typing, Pathogenic microorganism, Typing and identification

病原微生物分型技术的研究对于传染性疾病的流行病学调查、跟踪和阻断其传播途径具有重要意义。如确定已分离的致病菌株引起的感染是否能通过健康人传播?环境中的细菌或从寄主身上分离到的细菌是否变得具有感染性?反复感染或持续感染是否是由同一个菌株所引起的,还是被另一个新的菌株所代替,或是该菌株发生了基因突变产生了新的基因型?一种新的治疗方法是否将一种条件致病菌变成了致病的流行株?要解决这些问题,都需要对病原菌进行分型鉴定<sup>[1]</sup>。

病原微生物分型技术的研究由来已久,从最早

的表型分型技术到现在的核酸指纹分型技术经历了近百年的历史。其中最早应用的表型分型技术主要是通过观察病原微生物的外部形态特征及一些生化特性来对其进行分型鉴定,这种方法虽然具有简单直观的优点,但表型特征数量有限且易受环境因素的影响而具有较强的主观性。目前应用比较广泛的核酸指纹分型技术主要是基于电泳图谱分析的一种分子分型方法,如限制性片段长度多态性(RFLP)、扩增片段长度多态性(AFLP)、可变数目串联重复(VNTR)多态性分析,此外还有核酸探针杂交分型技术、脉冲场凝胶电泳(PFGE)、SNP分型技术等。这

\* 通讯作者 Tel: 0431-85619486, E-mail: wli99@jlu.edu.cn

收稿日期: 2007-03-03, 修回日期: 2007-06-04

些技术在研究病原菌的流行病学特征、了解病原菌的传播途径、追溯传染源及在临床传染性疾病的诊断中都发挥了重大作用<sup>[2]</sup>。

尽管如此,这些基于电泳图谱分析的分型方法都有一些不足,主要表现在不同实验室间的重复性较差,对变异菌株的鉴别能力较差,对两个菌株之间的遗传关系不能有效衡量。其中最主要的缺点是不同的实验室获得相同的结果比较困难。由于这些技术都是通过比较电泳条带的大小和有无来进行判断分析的,不排除某些情况下存在差异较大的两条带大小却相近,此外这些方法操作繁琐,技术复杂,实验结果影响因素较多;不同的分型方法被用于不同的实验室同一病原菌的分型研究上,即使不同的实验室用同一种分型方法所获得的数据也通常难以一致,从而使得进化分析、分类学及群体遗传学等方面的研究而显得不合适。

理论上,鉴定分析两个菌株亲缘关系最精确的方法是他们的基因组序列,只有DNA序列能得出两个菌株是否完全相同的结论,而其他的方法只能得出两者最具有可能的亲缘关系的程度,但是通过比较基因组序列的方法在现实中是不可行的,即使将来可行,对基因组序列海量数据的分析及解释也是一项具有挑战性的工作<sup>[1]</sup>。但序列分析方法对于用PCR技术在基因组某些区域进行扩增所产生的片段进行同源性分析却是可行的。1998年,多位点测序分型(MLST)技术作为一种以核苷酸序列分析为基础的病原菌分型方法被首次提出<sup>[3]</sup>,他是高通量测序技术和成熟的群体遗传学相结合的产物。该方法简单易行,重复性强,能够反映病原菌的群体进化生物学,在流行病学调查和病原菌的分型鉴定中发挥了巨大作用,目前该技术已应用于多种原核及真核病原微生物的分子分型<sup>[4]</sup>、检测鉴定、来源追踪等领域。

## 1 MLST技术的原理

MLST技术是多位点酶电泳(multilocus enzyme electrophoresis, MLEE)技术的延伸,传统的MLEE方法是一种以酶蛋白为基础的分型方法,通过生物个体间多位点酶的差异来反映其遗传基础的差异。如同一种酶在不同的位置出现2条带,则反映存在不同结构的2种酶蛋白,即存在同一基因的两个等位基因,通过对多个酶基因位点的综合分析,就可获得

病原菌的基因型<sup>[5]</sup>。而MLST技术通过测序技术来揭示保守基因的等位基因突变来对病原菌进行分型和鉴定。细菌或真菌的同一属或种中,持家基因(housekeeping gene)几乎都存在保守性,但不同的种或菌株间又存在差异即变异性,一个等位基因可能存在单个或者多个核苷酸的变异。MLST方法就是根据待分型病原菌的基因组序列注释信息,针对数个持家基因设计引物并选择数十甚至上百株菌的染色体DNA进行PCR扩增和产物测序来进行分析的。由于MLST是对每个等位基因进行测序,能获悉每个基因座位中所有的变异情况,从而能发现比MLEE电泳图谱分析更多的等位基因。

在MLST方法中,对某一菌株每个持家基因的所有等位基因都按发现的先后顺序分配一个等位基因序号(allele number),把该菌株所有持家基因的每个等位基因合并在一起组成一个等位基因谱(allelic profile),并给这个等位基因谱分配一个唯一的编号作为该分离株的核酸型或序列型(sequence type, ST)<sup>[6]</sup>。

因为等位基因间的微小差异将会导致菌株间的序列型不同,通过对每个菌株ST的比较就可以揭示各个分离株之间的亲缘关系:亲缘关系较近的菌株具有相近的序列型,反之,序列型不相近的分离株具有较远的亲缘关系。当细菌发生变异时,等位基因多个位点的差异具有比单个位点差异具有较远的亲缘关系。将分离到的未知病原菌,与某一菌种的MLST数据库比较就可以确定它的序列型。

## 2 MLST设计及流程

针对某个菌种建立一个新的MLST时须注意以下三个点:菌种的选择、基因位点的选择和特异引物的设计<sup>[2]</sup>。

首先菌种的选择和菌株数量的确定。根据菌种的流行病学和分型的研究背景,选择存在种群多样性特征的病原菌,所挑选的菌株要具有代表性,避免仅选择可能代表某一亚群的菌株。其次持家基因位点的选择,越来越多的病原菌全基因组序列测定的完成,为MLST筛选候选持家基因提供了宝贵的数据。这些持家基因应即相对保守,又可允许局部碱基发生点突变。持家基因最好不是毒力相关基因或膜相关蛋白基因,因为这种基因在病原菌进化过程受到的选择压力大,其核苷酸顺序可能变化很小或

根本没发生改变,只有其突变位点的数量和种类足够多,才能充分反映病原菌的系统发育关系。此外对于持家基因的数目,一般7个持家基因片段基本能反映出病原菌群体的变异情况。这7个持家基因要相距一定的距离并应尽量覆盖整个染色体。在设计特异引物时要注意扩增片段的长度,一般扩增片段在400bp~500bp足以反映其持家基因的变异情况,并且在进行测序反应时,每个等位基因可通过一个反应就可完成测序。

### 3 MLST在病原菌分型鉴定中的应用

目前,MLST分型技术已被应用于20多种病原细菌及3种病原真菌的分型鉴定中<sup>[7]</sup>,有许多病原菌的MLST数据库已经建立,其中最主要的两个是伦敦帝国学院<sup>[8]</sup>(<http://www.mlst.net>)和牛津大学(<http://www.pubmlst.org>)的MLST数据库。前者连接许多病原菌的MLST数据库,并提供对病原菌等位基因序列及序列型进行分析的工具,还有相应的网页不断更新世界各地研究人员提交的某一菌株的最新流行病学信息。后者主要为那些有兴趣建立自己的MLST的研究人员提供研究工具,诸如用非冗余数据库(non-redundant database, NRDB)确立等位基因,进行不平衡连锁及进化树分析,提供对某一菌株与数据库中的菌株进行亲缘关系比较查询等服务<sup>[9]</sup>。

脑膜炎奈瑟菌(*Neisseria meningitidis*)是最早用于评价MLST的病原菌,研究发现MLST比MLEE分析生成的遗传进化树更为合理,而且能够清晰覆盖全球大多数侵袭性脑膜炎奈瑟菌,MLST技术现在已被许多实验室用于脑膜炎奈瑟菌的分型鉴定<sup>[10]</sup>。国内张怡滨、宋诗铎等对天津地区4株金黄色葡萄球菌(SA)进行多位点测序分型,研究金黄色葡萄球菌临床株的来源和遗传背景。认为MLST作为SA分型方法,分辨率高,结果准确,易于标准化,适于进行分子进化研究<sup>[11]</sup>。

台湾的Chen Kuo-Wei等运用MLST技术对从台湾12家医院分离到的51株白色念珠菌(*Candida albicans*)进行遗传特征分析,对7个持家基因的序列分析共发现83个多态性核苷酸位点,45个二倍体序列类型(diploid sequence types, DSTs),其中36.3%是同义突变,63.9%为非同义突变。认为MLST具有比脉冲场凝胶电泳更高的分辨率,它能够分辨某

一病人所感染的白色念珠菌随时间的推移而发生的微小的种内进化<sup>[12]</sup>。Andrew R. Dodgson等应用MLST方法对不同地理区域收集的107株临床光滑念珠菌菌株及2个参考菌株进行指纹图谱分析,对11个持家基因编码区进行扩增并测序分析,其中有6个基因的扩增片段具有多态性位点,81个核苷酸位点发现有变异。在109株菌中鉴定出了30个序列型(STs),在所有分离株中有5个主要的分支,其中3个进化分枝显示明显的地理差异,表明在不同的地理区域,光滑念珠菌具有明显的遗传分化<sup>[13]</sup>。

MLST如结合real-time PCR探针技术可以精确、快速的对大量分离株进行检测。根据每一个等位基因间碱基的差异,设计等位基因特异引物及探针,从而可以利用real-time PCR的高分辨率及高通量的特性对某一特异菌株进行检验即allele-specific probing检测技术。

### 4 结论

MLST技术通过直接对持家基因的每个基因座位进行测序而不是比较其所编码酶的电泳迁移率。这种改进具有明显的优势,他能获悉每个基因座位中的所有变异情况,因而能在核苷酸水平上识别菌株的差异,这样可以检测到更多的变异,能发现比MLEE电泳图谱分析更多的等位基因。它可用于构建系统发育图,进而推断菌株间的系统发育关系,可以直接鉴定病原菌的亚型。

MLST使用的序列测定技术,较其他的分型方法更明确即无歧义,使得不同实验室间的结果具有高度的可重复性及可比性。正是基于这一点,可将某个菌种的“持家基因”的序列数据提交到MLST网站上,形成该菌种详尽的数据库,实验室间无需交换菌种,即可共享某个菌种相关的序列数据,实现全球范围的数据交流;MLST技术实现了序列数据可以真正交换及共享,这种优势是其他任何方法无法比拟的。他允许世界各地的其他实验室通过互联网访问被存放在数据库中的每个菌株的序列信息,使得不同国家或实验室在本地分离的相关菌株能够被世界各地的实验室利用,从而使得进行全球范围内的流行病学调查成为可能<sup>[14]</sup>。

与其他以PCR为基础的技术一样,MLST具有当菌株不能有效分离时,如某些急性进展性疾病,可以无需病原菌的培养直接从临床样本中提取DNA,可

扩增出“持家基因”进而通过测序获得等位基因谱。此外,它不需要对所有的菌株同时进行分析,可以根据所得到的菌株样本先后进行分析,因为序列数据是不依赖于实验方法的,重复性是绝对的。MLST技术虽然有以上众多优点,但其局限性也不容忽视,首先持家基因的变异情况决定着MLST的分辨力,如某些菌株的持家基因变异程度低或无变异从而导致MLST分型方法的失效,其次MLST需要核酸测序技术,测序成本过高也是限制MLST方法应用的主要原因之一。目前,已有几十种病原菌的MLST数据库已经被建立,随着测序技术的发展和测序成本的不断降低,必然会有更多病原生物体的MLST数据库被建立,从而使得MLST技术得到广泛的应用,这将为病原微生物的分子分型、遗传进化分析及系统发育研究提供有力的工具。

### 参考文献

- [ 1 ] Gil-Lamaignere C , Roilides E , Hacker J , *et al.* Clin Microbiol Infect , 2003 , **9** ( 3 ) : 172 ~ 185 .
- [ 2 ] 杨瑞馥 , 宋亚军 . 微生物法医学 : 理论与技术 . 北京 : 化学工业出版社 , 2005 , pp. 30 ~ 31 .
- [ 3 ] Maiden M C J , Ygraves J A B , Feil E , *et al.* Proc. Natl. Acad. Sci. U. S. A , 1998 , **95** : 3140 ~ 3145 .
- [ 4 ] Tavanti A , Gow A R N , Senesi S , *et al.* Journal of Clinical Microbiology , 2003 , **41** ( 8 ) : 3765 ~ 3776 .
- [ 5 ] Taylor J W , Geiser D M , Burt A , Koufopanou V . Clin Microbiol Rev , 1999 , **12** ( 1 ) : 126 ~ 46 .
- [ 6 ] Urwin R and Maiden M C J . TRENDS in Microbiology 2003 , **11** ( 10 ) : 479 ~ 487 .
- [ 7 ] 吴利先 , 黄文祥 . 中华检验医学杂志 , 2006 , **29** ( 13 ) : 278 ~ 281 .
- [ 8 ] Jolley K A , Feil E J , Chan M S , *et al.* Bioinformatics Applications Note 2001 , **17** ( 12 ) : 1230 ~ 1231 .
- [ 9 ] David M A and Brian G S . Nucleic Acids Research ( Web Server issue ) , 2005 , **33** : W728 ~ W733 .
- [ 10 ] Krizova P , Kalmusova J , Musilek M , *et al.* Epidermal Microbiol Immunol 2004 , **53** : 25 ~ 36 .
- [ 11 ] 张怡滨 , 宋诗铎 , 祁伟 , 等 . 天津医药 , 2004 , **32** ( 10 ) : 627 ~ 629 .
- [ 12 ] Chen K W , Chen Y C , Lo H J , *et al.* Journal of Clinical Microbiology , 2006 , **44** ( 6 ) : 2172 ~ 2178 .
- [ 13 ] Dodgson A R , Pujol C , Denning DW , *et al.* Journal of Clinical Microbiology , 2003 , **41** ( 12 ) : 5709 ~ 5717 .
- [ 14 ] Tavanti A , Davidson A D , Johnson E M , *et al.* Journal of Clinical Microbiology 2005 , **43** ( 11 ) : 5593 ~ 5600 .