维普资讯 http://www.cqvip.c

第19卷5期
2004年10月

中国病毒学
VIROLOGICA SINICA

19（5）：444-448
October 2004

我国中部HIV-1 B′亚型分离株基因组结构特征及系统发育分析

冯福民，李敬云**，鲍作义，庄道民，刘思杨，李 林

（军事医学科学院微生物流行病研究所 北京 100071）

Strutural Characterization and Phylogenetic Analysis of HIV-1 Genome of a Prevalent Subtype B′ Strain in Central China

FENG Fu-min，LI Jing-yun，BAO Zuo-yi，ZHUANG Dao-min，LIU Si-yang，LI Lin

(Institute of Microbiology and Epidemiology, Academy of Military Medical Science, Beijing, 100071, China)

Abstract: In order to analyze the gene characterization of *Human immunodeficiency virus-1*(HIV-1) prevalence in Henan Province, a HIV-1 strain, named CNHN24, was isolated from one HIV-1 infected blood donor. The proviral DNA was extracted from productively infected PBMCs (peripheral blood mononuclear cells) and full-length HIV-1 genomes were amplified. The sequences were determined by Walking sequencing. All of the reading frames were compared with HIV-1 RL42 and HIV-1 HXB2. The secondary structure prediction of the envelope proteins of HIV-1 CNHN24 and RL42 were performed by Anthewin software. Multiple alignments of the whole genome, *env* and *V3* loop gene with domestic HIV-1 strains and standard representatives of HIV-1 were performed by Clustal software package, and phylogenetic tree were constructed. The results showed that the HIV-1 CNHN24 was B′ subtype (according to *V3* nucleotide sequence). There were more purines in the HIV-1 CNHN24 genome, it was 60.24%. Compared with HIV-1 RL42 and HXB2, the homegeneity of *gag* and *pol* reading frames was higher, *env* and non- structural genes was lower. The envelope proteins secondary structure of CNHN24 and RL4 were homoplastic, but the amino acid variability was higher in conservative *C4*. Phylogenetic trees showed the closest relationship between CNHN24 and RL42, the next was between HIV-1 Lai and HIV-1 HXB2. HIV-1 *V3* sequence phylogenetic tree also indicated the closest relationship of CNHN24 to CR206 and RL42, which were isolated in Yunnan Province. We inferred that the HIV-1 CNHN24 may derive from Yunnan Province.

Key word: *Human immunodeficiency virus-1*(HIV-1); Genome; Phylogenetic analysis

摘要：为了分析河南省 *Human immunodeficiency virus-1*(HIV-1)流行株的基因特征和传染来源，我们从一位 HIV-1 感染的有偿献血员体内分离 HIV-1 毒株。提取感染细胞的全基因组 DNA，扩增病毒全基因组，Walking 法测得全长序列。与 HIV-1 RL42 和 HIV-1 HXB2 株比较，分析基因特征。用 Anthewin 软件比较分析 CNHN24 株和 RL42 株 *env* 蛋白的二级结构。用 Clustal 软件对国内 HIV-1 分离株和国际参考株的全基因组、*env* 基因和 *V3* 环基因分别进行了系统发育分析。结果发现，CNHN24 株为 HIV-1 B′ 亚型，富含嘌呤，含量达到 60.26%。与 RL42 株及 HXB2 株比较，*pol* 基因和 *gag* 基因变异度较小，*env* 基因及非结构基因变异度较大。与 RL42 株相比，*env* 蛋白二级结构变异不大，但 *env* 保守区 *C4* 的氨基酸呈高度变异。系统发育分析显示，CNHN24 株与 RL42 株遗传距离最近，HIV-1 Lai 株和 HXB2 株次之，与国内的 HIV-1 B/C 及 AE/BC 重组毒株的遗传距离较远；从 *V3* 序列的比较可以发现，CNHN24 株与来自云南省的毒株 HIV-1 CR206 及 RL42 遗传距离最近。认为 HIV-1 CNHN24 株可能由云南传入。

投稿日期：2004-04-05，修回日期：2004-05-31

作者简介：冯福民(1963-)，男，河北唐山籍，华北煤炭医学院任职，副教授，博士研究生，研究方向为传染病防治。

** 通讯作者：李敬云(1963-)，女，山东籍，博士，研究员，研究方向为艾滋病防治。

Corresponding auhtor. Tel: 010-66948566, Email: lijy@nic.bmi.ac.cn

维普资讯 http://www.cqvip.c

关键词：I 型艾滋病病毒；基因组；系统发育分析
中图分类号：R511 文献标识码：A 文章编号：1003-5125(2004)05-0444-05

艾滋病的流行已有 20 年的历史，1983 年法国巴斯德研究所的Montagnier和美国癌症研究中心的Gallo 分别从男性同性恋者血液中分离出艾滋病病毒(Human immunodeficiency virus, HIV)。我国 1985 年报告首例 HIV 感染者[1]，从此，艾滋病开始在国内流行，其间经历了传入期、播散期并进入快速增长期。特别是进入九十年代，HIV-1 B′ 亚型在我国部分地区的有偿献血员中传播[2]，这种经非法单采血浆而形成的中国特有的传播途径造成短时间内大量人群感染，带来严重的医疗问题和社会问题。由于 HIV 的变异率较高，各地流行的 HIV 具有不同的基因特征，表现为不同的进化过程。本研究于2002年初在我国中部分离一株HIV-1 B′ 亚型流行株，对其基因组进行了详细的全序列分析及进化途径分析，其结果不仅有助于阐明这一流行株的进化特征，而且还将为有关病毒流行特征等方面的研究提供一个完整、清楚的分子生物学背景。

1 材料与方法

1.1 血液样品

2002年2月在河南省农村地区进行的HIV-1传播途径调查中，无菌采集一名女性有偿献血员（HIV-1 感染者）静脉血 10mL 于 EDTA 抗凝管中，同时准备健康人抗凝血。研究对象有多次献血史，发现 HIV-1 感染<1 年，无其他危险性暴露。

1.2 病毒分离与基因提取

按照参考文献[3]方法分离健康人和 HIV-1 感染者单个核细胞，将健康人淋巴细胞作为滋养细胞短期增殖培养 HIV-1。按照 1:1 的比例加入细胞，于 37℃培养箱中 7d 后收集细胞，成功分离 HIV-1 病毒命名为 CNHN24。用博大泰克公司试剂盒提取细胞基因组 DNA。

1.3 基因分型

根据 Bachmann MH[4]的分型研究方法，利用 HIV-1 *env* 区的两对引物和 *gag* 区的两对引物分别从 HIV-1 感染者基因组 DNA 中套式扩增出 686bp（HXB2 株的 6681-8007 和 7185-7871）和 413bp（HXB2 株的 816-2061 和 836-1249）的片段，送交测序公司进行 PCR 产物测序，用 NCBI 的 BLAST 程序在 GenBank 中进行序列一致性比较，根据一致性高低确定被测样品的 HIV-1 亚型。

1.4 全长基因扩增与测序

根据 3′ RACE 确定的序列合成 3′ 端引物 LTR3（gaa gca ctc aag gca agc ttt att gag g），同时根据文献[5]提供的 HIV-1 mRNA 5′ 端保守区序列合成 5′ 端引物 LTR5（aat ctc tag cag tgg cgc ccg aac ag），利用 TaKaRa 公司的 LA Tag 在基因组 DNA 中扩增 HIV-1 全长基因。扩增条件为：94℃ 3min，(热启动)94℃ 30s, 68℃ 10min, 3cycles；94℃ 15s, 68℃ 7min and 30s (每个循环递增 5s), 20cycles;72℃ 10min。全长 9kb 的PCR产物送测序公司采用 walking 方法分段测序，每个测序反应的长度为 400-500bp，按顺序拼接成全长基因。

1.5 二级结构与系统发育分析

采用 Anthewin 软件分析 HIV-1 *env* 蛋白的二级结构；用 CLUSTAL X 1.8 软件进行系统发育分析，绘制系统树。

2 结果

2.1 CNHN24 株病毒基因组结构特征

在基因组 DNA 中扩增得到全长 9010nt 的 HIV-1 B 亚型基因，测序后发现在 9010nt 基因组中，四种碱基分别为：3240a、1583c、2189g、1998t。在其基因组成上，嘌呤碱含量相对较高（5429/9010），占 60.26%，符合逆转录病毒属基因组碱基组成特征。该基因编码区含有 *env*、*gag* 和 *pol* 三个结构基因以及 *tat*、*nef*、*rev*、*vpr*、*vpu* 和 *vif* 六个非结构基因，共编码 3154 个氨基酸。非编码区为 5' 端和 3'端重复序列，分别编码 56 个和 217 个氨基酸，其中 5'端为非完整序列。

2.2 核苷酸和氨基酸变异度分析

将本实验室分离株（CNHN24 株）、国内 HIV-1 B 亚型克隆株(RL42株)及流行于欧美地区的HIV-1 B 亚型株（HXB2 株）的全序列输入计算机，用 DNAsis 2.5 软件进行比较分析。结果表明，CNHN24 株与 RL42 株的基因变异度小于 HXB2 株。其中结构基因 *pol* 的变异度最小，*env* 变异度最大。而在非结构基因中，*vif* 和 *vpr* 变异度较小，*enf* 变异度较大。在氨基酸水平上也表现出明显的差异。表 1 显示了 CNHN24 株与 RL42 株及 HXB2 株比较时不同基因的核苷酸及氨基酸差异。

从表 1 也可以看出，在核苷酸水平和氨基酸水平上，CNHN24 株各基因的系统发育程度不同。相对保守的结构基因 *gag* 和 *pol*，氨基酸变异率虽然

维普资讯 http://www.cqvip

也高于核苷酸的变异率，但差别不大，而 *env* 基因及非结构基因的差别就比较大。

表1 CNHN24株核苷酸及氨基酸与RL42株及HXB2株的比较

Table 1 The variety of nucleic acid and amino acid of HIV-1 CNHN24 compare with HIV-1 RL42 and HIV-1 HXB2

Gene	RL42 strain		HXB2 strain	
	Difference of nucleic acid(%)	Difference of amino acid(%)	Difference of nucleic acid(%)	Difference of amino acid(%)
env	6.29	13.47	10.38	20.80
gag	3.29	6.13	4.92	7.06
pol	2.92	4.08	4.91	6.88
tat	6.43	13.56	12.55	23.17
nef	8.06	12.95	14.15	21.82
rev	6.57	12.46	12.04	21.64
vpr	4.58	8.06	14.72	27.64
vpu	5.69	13.19	17.39	30.32
vif	4.32	10.47	8.12	18.76

HIV-1 *gp120* 基因由五个保守区 *C1*～*C5* 和五个高变区 *V1*～*V5* 组成。因为氨基酸序列表示一种基因的稳定遗传特征优于核苷酸序列，因此，我们比较了进化关系最近的三株 HIV-1 B′ 亚型，CNHN24 株、RL42 株及 HIV-1 97TH6-107 株的 *gp120* 氨基酸序列，任意两株的变异率都超过16%，而且，三株病毒的 *gp120* 氨基酸总变异率达到了35%。各区的变异率差别较大，如果按变异率高低排序，则三株病毒 *V1*、*V2*、*C4*、*V5*、*V3* 和 *V4* 的总变异率依次 65.62%、46.34%、44.23%、41.67%、40%和40%；CNHN24 株与 RL42 株相比，*V1*、*V4*、*V3*、*V5*、*C4* 和 *V2* 的变异率依次 46.87%、33.33%、25.71%、25%、23.08%和21.95%。后者的变异率明显低于前者，提示 CNHN24 株与 RL42 株的遗传关系较密切。值得注意的是，*C4* 区作为保守区，氨基酸变异率却与高变区相当，而且 *C2* 区与 *V3* 区相邻的25个氨基酸的变异率却超过了 *V3* 区。

2.3 包膜蛋白二级结构分析

CNHN24 株包膜蛋白由846个氨基酸组成，编码 gp120 和 gp41 两种糖蛋白。在 HIV 复制及传播过程中，基因处于动态变化中。与分离自我国云南的 HIV-1 B′ 亚型 RL42 株相比，传播时间仅相隔5年，但基因变异非常明显。有114个氨基酸发生突变，其中，无义突变41个，缺失和插入14个，27个氨基酸的亲疏水性发生改变，32个氨基酸的发生电荷改变。而与流行于欧美地区的 HXB2 株相比，包膜蛋白区氨基酸差异达到176个。氨基酸变异导致蛋白质二级结构的改变，进而影响到蛋白质的功能。但图1的 CNHN24 株与 RL42 株的包膜蛋白的抗原性和亲\疏水性并未见明显差别。

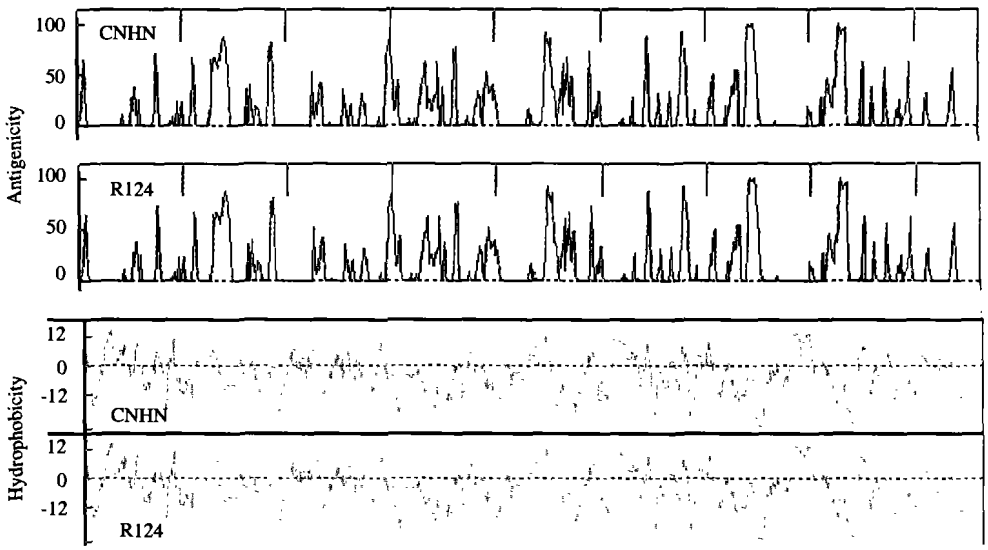

图1 HIV-1 CNHN24 株与 RL42 株包膜蛋白二级结构预测

Fig. 1 The secondary structure prediction of envelope proteins of HIV-1 CNHN24 and HIV-1 RL42

2.4 HIV-1 CNHN24 株病毒基因组系统发育分析

利用 Los Alamos HIV 数据库中检索国内 HIV-1 分离株及 HIV-1 B 亚型和 C 亚型国际参考株全基因组序列，进行系统发育分析，结果表明，CNHN24 株与 RL42 株遗传距离最近，其次为国外 HIV-1 B 亚型参考株 HXB2 株和 Lai 株。进化树如图2所示。

在全基因组分析的基础上，我们对上述 HIV-1 分离株的 *env* 区也进行了系统发育分析（图3）。比较图

维普资讯 http://www.cqvip.c

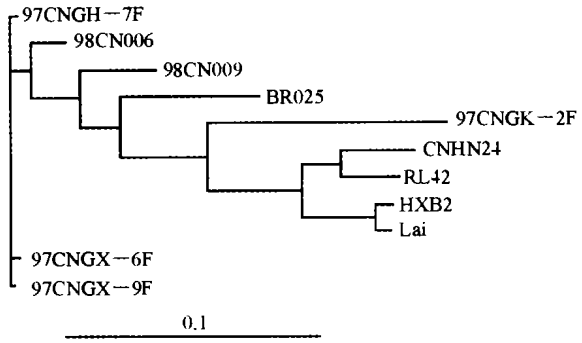

图 2 不同 HIV-1 分离株全基因组系统发育树

Fig. 2 The phylogenetic tree of some HIV-1 strains complete genome

2 和图 3，未发现两者存在明显的差别。但进行系统发育分析需要考虑的因素很多，如编码区和非编码区、高变区和保守区、转换和颠换、插入和缺失、无义突变和有义突变、突变速率的差异等。在 *env* 区内包含 5 个保守区和 5 个高变区，混合在一起进行进化分析有可能得到错误的结论。为此，我们将 *V3* 环序列单独提取出来进行系统发育分析，以比较其中的差别，同时也加入国内另外几株 HIV-1 B′ 亚型的 *V3* 环序列，结果见图 4。从这三个图的结果，我们不难看出，与 CNHN24 株同源关系最近的依次为分离自云南的 CR206 株和 RL42 株，其次为云南的 CR7 株和山东的 SD8 株，然后才是国外参考株 Lai 株与 HXB2 株。值的注意的是，来自广西的 96CHI31 株 HIV-1 *V3* 环序列与其它地区的序列差异较大，系统发育树也显示，96CHI31 株与其它 HIV-1 B 亚型株的遗传距离较远，而与同样来自广西的一株 AE/BC 重组型 HIV-1 遗传距离较近。

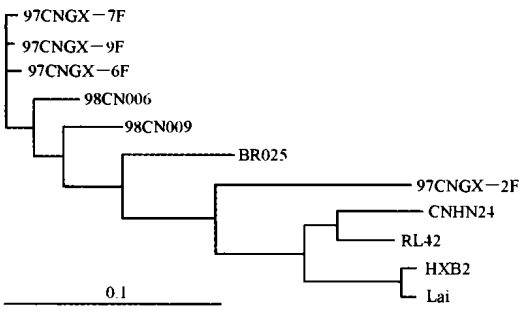

图 3 不同 HIV-1 分离株的 env 区系统发育树

Fig. 3 The phylogenetic tree of *env* gene of some HIV-1 strains

3 讨论

HIV 分为 HIV-1 和 HIV-2 两型，而 HIV-1 型包括至少 10 个亚型。根据亚型不同，HIV-1 基因组为大约 9100~9800 个核苷酸的单股正链 RNA。嘌呤碱

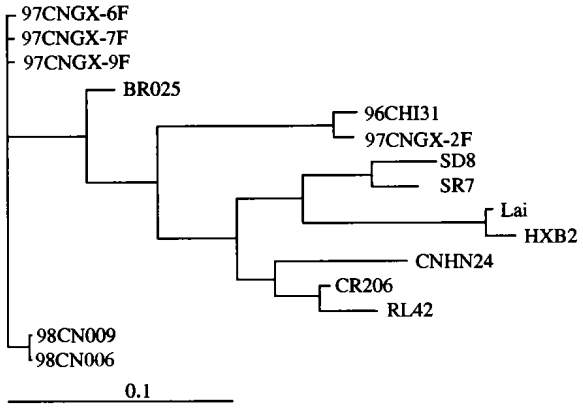

图 4 不同株 HIV-1 基因的 V3 环系统发育树

Fig. 4 The phylogenetic tree of *V3* loop of some HIV-1 strains

含量丰富，本文分离的 CNHN24 株病毒基因组中腺嘌呤占 35.96%，鸟嘌呤占 24.30%。从基因组的 5' 末端至 3'末端依次排列着 *5'LTR*、*gag*、*pol*、*vif*、*vpr*、*tat*、*rev*、*vpu*、*env*、*nef*、*3'LTR* 及多聚腺苷酸尾。其中，*gag*、*pol* 和 *env* 为结构基因，其余为调控基因。

env 基因编码 HIV-1 的包膜蛋白，是病毒进入宿主细胞的特异结合部位，分为表面糖蛋白 *gp120* 和跨膜糖蛋白 *gp41*。目前对 HIV-1 基因组变异的研究多局限于 *env* 区，一方面是 *env* 区具有重要作用，另一方面也是因为缺乏不同地区优势流行株的基因组数据，如国内仅有几株 HIV-1 B′ 亚型和 B/C 亚型基因组序列。HIV-1 B′ 亚型只有 1998 年分离自云南省的 RL42 株和 2002 年分离自河南省的 CNHN24 株。这两株病毒是于不同流行年代从不同流行地区分离到的，在病毒传入该地区的时间上，两株病毒至少相差 5 年，因此，比较这两株病毒的基因组序列，可以分析 HIV-1 B′ 亚型在国内传播过程中基因组变化特征。通过 *env* 区和 *V3* 环基因的系统发育分析认为，CNHN24 株来自云南，虽然传播时间只相差几年，但其核苷酸的总一致率只有 95.64%，证实 HIV-1 的高变异性，与 Zhong 等人[6] 的研究结果相符合；但是，在一个国家或地区的 HIV-1 流行株的同源性仍然高于其他国家或地区的毒株[7]。HIV-1 的逆转录酶没有校正功能，是目前已知的变异率最高的病毒[8]，在每个复制周期中至少一个核苷酸突变[9]。这种现象对 HIV-1 的变异、进化、传播及致病带来很大的变数或不确定的影响因素，大大增加了新变异株的起源和发生的可能性，大大增强和丰富了病毒抗原变异的能力，也大大增加了人体甚至整个医学界低抗 HIV-1 感染和致病的成本与难度，使得 HIV-1 不断突破人体的免疫屏障，造成长期感染和人体免疫缺陷，为 HIV 的防

维普资讯 http://www.cqvip.c

治研究提出了更多的难题。

与RL42株比较，CNHN24株的*env*区虽然有13.5%的氨基酸发生突变，但两者的二级结构却相差不大，表明这些氨基酸的突变不是无义突变，就是突变点不是关键位点。即使将*env*区二级结构示意图放大，也只能看到两株病毒包膜蛋白的个别位点存在差别，提示我们虽然两株病毒包膜蛋白氨基酸差异比较明显，但在蛋白质结构上却表现出相当高的稳定性。在HIV-1 *gp120*基因的五个保守区*C1*～*C5*和五个高变区*V1*～*V5*中，对进化关系最近的三株HIV-1 B′亚型，CNHN24株、RL42株及97TH6-107株的*gp120*氨基酸序列进行的比较，各区的变异率差别较大，与以往报道的排序不符。一些保守区的氨基酸变异率却超过了高变区。这样的结果说明，随着病毒在不同地区的传播，同型病毒也会产生较大的变异。所谓保守区也只是相对的，我们分离的CNHN24株*C4*区和*C2*区的高变异率即说明了这一点。

在进行系统发育分析时，进化速度相差较大的基因不能作为一个整体进行分析，如HIV-1 *env*区的*C2*区为保守区，而*V3*区为高变区，将这两段基因合并进行分析则可能得到有偏的甚至是错误的结论，又如同一序列高变区和保守区的变化速度不同，若只从整个序列变化的平均数去推算遗传距离则会造成低估，如HIV-1型*env*基因*V4-C4-V5*区是两个快速进化区，位于一个保守区两侧，这样把它们分开来分析较为适宜[10]。虽然如此，仍有些研究者未能区分保守区和高变区而将两者合并进行系统发育分析[11,12]，使其结果的可信度遭到质疑。对HIV-1 *gp120 V3*高变区的研究进一步说明了区分不同变异区域进行系统发育分析的合理性[13]。

从系统发育分析结果可以看出：虽然国内HIV-1分离株与国外HIV-1 B亚型参考株全基因组、*env*区和*V3*环的进化方向是相似的，但也不能说明将多段进化程度不同的基因合并在一起分析是合理的。在大多数情况下，使用局部比对是较为合理的，这种比对方法可能会揭示一些匹配的序列段，而本来这些序列段是夹杂在一些完全不相关联的碱基对中的，因此，不恰当地使用全程比对很可能会掩盖一些局部的相似性。一组相关病毒的进化史称为种系发生，通常以分叉的进化树表示，两个枝端的毒株在其起点处有着共同的祖先。毒株的亲缘关系的远近以树枝的长短表示。国内的HIV-1B′亚型和HIV-1 B/C及AE/BC重组型分离株的遗传距离较远，即使同是B′亚型，遗传距离差别亦较大，表明国内艾滋病感染的传染源不同，体现出不同的传入途径和在国内的进化过程。

参考文献

[1] Zeng Y, Fan J, Zhang O, *et al*. Detection of antibody to LAV/HTLV- III in sera from hemophiliacs in China[J]. AIDS Res, 1986, 2 (Suppl 1): S147-S149.

[2] Su B, Liu L, Wang F, *et al*. HIV-1 subtype B' dictates the AIDS epidemic among paid blood donors in the Henan and Hubei provinces of China[J]. AIDS, 2003, 21, 17(17):2515-2520.

[3] Salminen M O, Koch C, Eric S B, *et al*. Recovery of virtually full-length HIV-1 provirus of diverse subtypes from primary virus cultures using the polymerase Chain reaction[J]. Virology, 1996, 213: 80-86.

[4] Bachmann M H, Delwart E L, Shpaer E G, *et al*. Rapid genetic characterization of HIV type 1 strains from four World Health Organization-sponsored vaccine evaluation sites using a heteroduplex mobility assay[J]. AIDS Res Hum Retroviruses, 1994, 10(11): 1345- 1353.

[5] Graf M, Shao Y, Zhao Q, *et al*. Cloning and characterization of a virtually full-length HIV type 1 genome from a subtype B'-Thai strain representing the most prevalent B-Clade isolate in China[J]. AIDS Research and human retroviruses, 1998, 14(3):285-288.

[6] Zhong P, BUrda S, Konings F, *et al*. Genetic and biological properties of HIV type 1 isolates prevalent in villagers of the Cameroon equatorial rain forests and grass fields: further evidence of broad HIV type 1 genetic diversity[J]. AIDS Res Hum Retroviruses, 2003, 19(12): 1167-1178.

[7] Harris M E, Maayan S, Kim B, *et al*. A cluster of HIV type 1 subtype C sequences from Ethiopia, observed in full genome analysis, is not sustained in subgenomic regions[J]. AIDS Res Hum Retroviruses, 2003, 19(12):1125-1133.

[8] McCutchan F E, Salminen M O, Carr J K, *et al*. HIV-1 genetic diversity[J]. AIDS. 1996; 10 Suppl 3:S13-20.

[9] Hammer S M, Yeni P. Antiretroviral therapy: where are we[J]? AIDS, 1998, 12 Suppl A: S181-188.

[10] Balfe P. Concurrent evolution of human immunodeficiency virus type 1 in patients infected from the same source: rate of sequence change and low frequency of inactivating mutations[J]. J Virol, 1990, 64: 4221-4428.

[11] Rodenburg C M, Li Y Y, Trask S A, *et al*. Near full-length clones and reference seqences for subtype C isolates of HIV type 1 from three different continents[J]. Aids research and human retroviruses, 2001, 17(2):161-168.

[12] Su L, Graf M, Zhang Y ZH, *et al*. Characterization of a virtually full-length Human Immunodeficiency Virus type 1 genome of a prevalent intersubtype(C/B') recombinant strain in China[J]. J Virology, 2000, 74(23): 11367-11376.

[13] Holmes E C, Zhang L Q, Simmonds P, *et al*. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 with a single infected patient [J]. Proc Natl Acad Sci USA, 1992, 1:89-96.