

· 艾滋病预防与控制 ·

中国艾滋病病毒 1 型 AE 循环重组型毒株
env 基因的变异和进化分析

梁浩 邢辉 魏民 陈钊 关琪 黄海龙 全宇 陈健平 洪坤学 施侣元 邵一鸣

【摘要】 目的 研究中国艾滋病病毒 1 型(HIV-1)AE 循环重组型(CRF01-AE)毒株 env 不同基因区序列变异的特点及其与进化压力的关系。方法 应用巢式聚合酶链反应(nested-PCR)对从全国部分省收集来的 HIV-1 感染者血液样本中的 HIV-1 外膜蛋白(env)基因进行扩增和亚型鉴定后,选择 34 份 CRF01-AE 重组型 HIV-1 毒株 env 基因 V3~V4 区及其邻近区域的序列进行系统进化树和氨基酸变异分析,并计算和分析氨基酸同义替换(K_s)值和非同义替换(K_a)值及 K_s/K_a 比值。结果 基因系统树显示中国的 34 份样本 CRF01-AE 重组型毒株与我国代表株 AE.97CNGX2F 和泰国代表株 AE.CM240、AE.93TH253 聚集在一起。氨基酸替换主要发生在 C3 和 V4 区,而 V3 区和 V3 上游区氨基酸序列相对保守,糖基化位点也比较保守。V3 环顶端四肽以 GPGQ 为主(87.50%)。大部分毒株的第 306 和 320 位点上没有出现带正电荷的氨基酸。整个 V3~V4 区的 K_s 值显著高于 K_a 值($P<0.001$),且 K_s/K_a 比值显著高于 1($P<0.001$),只有 V4 区 K_s/K_a 比值显著低于 1($P<0.01$)。结论 目前多数流行于中国的 CRF01-AE 重组型 HIV-1 毒株具有较高的同源性,在进化上关系密切。氨基酸替换主要发生在 C3 和 V4 区,而不是 V3 区。由于大部分毒株的第 306 和 320 位点上没有出现带正电荷的氨基酸,推测这些毒株为不引起细胞融合的 NSI 型,其 V3~V4 区序列的变异主要受到功能性的限制而与正向选择压力无密切关系,但其中 V4 区的变异与正向选择压力有关。

【关键词】 人类免疫缺陷病毒;基因;变异;进化;分子

Study on the evolutionary pressure on the env gene of the human immunodeficiency virus type 1 CRF01-AE strains circulating in China LIANG Hao*, XING Hui, WEI Min, CHEN Zhao, GUAN Qi, HUANG Hai-long, QUAN Yu, CHEN Jian-ping, HONG Kun-xue, SHI Lv-yuan, SHAO Yi-ming.
*National Center for Sexually Transmitted Disease and Acquired Immune Deficiency Syndrome Prevention and Control, Chinese Center for Disease Prevention and Control, Beijing 100050, China
Corresponding author: SHAO Yi-ming

【Abstract】 Objective To identify variations in the env gene of human immunodeficiency virus type 1(HIV-1) subtype CRF01-AE strains circulating in China and to elucidate the potential relationship between genetic variation and evolutionary pressure. **Methods** Fragments of the HIV-1 env gene were amplified by nested-polymerase chain reactor(n-PCR) from the whole blood of HIV-1 infected individuals from four provinces in Southeast China (Guangdong, Hunan, Jiangsu and Jiangxi). The PCR products were then directly sequenced by ABI 377 DNA sequencers. The sequences covering the env V3-V4 region of 34 HIV-1 subtype CRF01-AE strains were selected to analyse phylogenetic trees and amino acid mutations. The accumulation of synonymous(K_s) and antonymous(K_a) substitutions as well as K_s/K_a ratios were calculated using DIVERGE. **Results** Phylogenetic trees showed that the 34 HIV-1 subtype CRF01-AE strains from China clustered with the Chinese AE reference strain(AE.97CNGX2F), as well as with the reference strains from Thailand(AE.CM240 and AE.93TH253). The amino acid sequences of the env V4 and C3 regions in the samples were highly variable, compared with those of V3 and V3-downstream regions. The V3 loop central motif in the majority(87.5%) of the strains was GPGQ. The majority of strains did not contain positively charged amino acids at positions 306 and 320 in V3 loop. The N-linked glycosylation sites in the V3-V4 region and flanking regions in these strains were relatively conserved. Analysis of the entire region showed that the mean K_s values were significantly higher than that of the K_a values($P<0.001$), with the K_s/K_a significantly higher than 1.0($P<0.001$). In contrast,

基金项目:2000 年度国家科技部基础性工作专项资金资助项目(120);国家重点“973”基础研究资助项目(G1999054107)

作者单位:100050 北京 中国疾病预防控制中心性病艾滋病预防控制中心(梁浩、邢辉、魏民、陈钊、关琪、黄海龙、全宇、陈健平、洪坤学、邵一鸣);华中科技大学同济医学院公共卫生学院(施侣元)

通讯作者:邵一鸣

the Ks/Ka ratio in the V4 region was significantly lower than 1.0 ($P < 0.01$). **Conclusions** Our study indicated that the majority of HIV-1 subtype CRF01-AE strains circulating in China were highly homogeneous. The amino acid sequences of the V4 and C3 regions were significantly more variable than those of the V3 loop. Our analysis also suggested that the phenotype of nearly all strains was likely to be non-syncytium inducing (NSI). Finally, the variation found in the V3-V4 sequence was significantly influenced by functional constraints as opposed to positive selective pressure, while the variability of the lone V4 region was strongly related to positive selective pressure.

【Key words】 Human immunodeficiency virus; Gene; Variation; Evolution, molecular

艾滋病病毒 1 型 (HIV-1) 在体内的独特复制方式 (无校正功能) 和基因重组以及快速的更新换代致使其基因组具有高度的多样性。目前这方面的机理已有许多研究报道。但宿主的选择压力与 HIV 基因的变异关系知之较少。此外,不同个体的 HIV-1 型的进化率是否反映了不同宿主的选择压力,目前尚未明了^[1]。最近,我们曾对我国 B 和 B/C 亚型 HIV-1 毒株的 env 基因 gp120 区序列的变异与选择压力的关系进行了研究,结果表明 B 和 B/C 亚型毒株具有不同的进化特点^[2]。同时,我们通过分子流行病学调查发现,我国 CRF01-AE 重组型 HIV-1 型在西南边境和东南沿海地区呈逐年上升趋势^[3]。为了探讨目前我国 HIV 感染人群中的 CRF01-AE 重组型 HIV-1 流行株的基因变异是由于正向选择压力所造成、还是病毒本身自然突变的结果,我们对随机收集的 CRF01-AE 重组型 HIV-1 流行株的 env 基因 V3~V4 区及其临近区域序列进行测序分析,结果报道如下。

材料与方 法

1. HIV-1 血液样本来源:本项研究所分析的 34 份 CRF01-AE 重组型 HIV-1 毒株为 2001~2002 年在我国广东、江西、江苏和湖南 4 省 HIV 感染者中采集鉴定的血清样本。感染者在知情同意后进行流行病学调查和采集 5~8 ml 静脉血,其中男性 29 例,女性 5 例,年龄 22~45 岁,感染时间 1~5 年,均为无症状 HIV 携带者,经吸毒 (27 例占 79.41%) 和性乱 (7 例占 20.59%) 途径感染。

2. 核酸提取:用德国 QiaGen 公司的 QIAamp Blood Mini Kit 试剂盒从每位感染者的抗凝全血中提取细胞 DNA,核酸最终浓度大约为 20~40 ng/ μ l,并冻存于 -80℃ 冰柜备用。

3. HIV-1 env 基因的扩增:用巢式聚合酶链反应 (nested-PCR) 对 HIV-1 env 基因 V3~V4 及邻近基因区 686 bp 进行扩增,外侧引物为 ENV-D5:5'-ATG GGA TCA AAG CCT AAA GCC ATG TG-3'

和 ENV-D12:5'-AGT GCT TCC TGC TGC TCC CA-3',内侧引物为 ENV-7:5'-CTG TTA AAT GGC AGT CTA GC-3'和 ENV-8:5'-CAC TTC TCC AAT TGT CCC TCA-3'。第一轮反应体系:总体积 30 μ l,10 \times buffer 3 μ l,dNTP 浓度 200 μ mol/L,Taq 酶 1.5 U,引物浓度 0.4 μ mol/L,核酸样本 5 μ l 及相应的 ddH₂O。反应条件:94℃ 1 min,55℃ 1 min,72℃ 2 min,3 个循环;94℃ 15 s,55℃ 45 s,72℃ 90 s,32 个循环;72℃ 10 min。第二轮反应体系:总体积 50 μ l,dNTP 浓度、引物浓度、Taq 酶用量及反应条件同第一轮,模板为 5 μ l 第一轮扩增产物。

4. 扩增片段的回收和纯化:PCR 产物经 1% 琼脂糖凝胶电泳,与 Marker 对照判断无误后,切下特异性扩增带,用德国 QiaGen 公司的 Qiaex 试剂,按说明书提纯扩增的 env 基因 V3~V4 区片段;回收得到的 DNA 溶于 pH 8.7 的 10 mmol/L Tris-HCl 缓冲液中。

5. 序列测定:以 D3 为测序引物,提纯的 PCR 产物为模板,采用美国 ABI 公司 Dye Terminator 循环测定反应试剂盒,在美国 PE 公司的 9700 型 PCR 仪上进行测序反应,反应条件:95℃ 10 s,50℃ 5 s,60℃ 4 min,25 个循环;反应产物经提纯后,在美国 ABI377 型 DNA 全自动测序仪上进行序列测定和分析。

6. 序列分析与数据处理:所测得的序列用美国 ABI 公司的 SeqED 软件进行编辑校正后,使用 Wisconsin Package GCG (Version 10.0) 软件进行分析。以 Pileup 程序对样本及国际标准序列进行排列和比较,其国际标准参考序列来自美国 Los Alamos HIV 基因数据库;用 Distances 程序计算序列间的基因离散率;用 Translation 和 Pretty 程序计算氨基酸序列的同源性序列;用 Meg 软件的 Neighbor-joining 法绘制系统树;用 Diverge 程序计算蛋白质的每个氨基酸位点的同义替换 (synonymous substitution, Ks) 与非同义替换 (antonymous substitution, Ka) 的比值^[4],采用 SPSS

软件包对数据进行统计学分析。

结 果

1. env 基因 V3~V4 区序列的进化关系的分析 将所研究的 34 份样本 CRF01-AE 重组型 HIV-1 毒株和来自中国、泰国、非洲国家的 CRF01-AE 重组型代表毒株以及国际各亚型标准毒株的 env 基因 V3~V4 区序列(471 bp)进行系统进化树分析。系统树显示我国的 34 份样本 CRF01-AE 重组型毒株与我国代表株 AE. 97CNGX2F 和泰国代表株 AE. CM240、AE. 93TH253 聚集在一起,而远离非洲 AE 代表株及其他亚型代表株。其中 31 份毒株来源于同一个分支(空心小圆圈),进化关系上非常密切, Bootstrap 值为 99%,而且进化树水平距离较短。其余 3 份毒株 hun0203、hun0201、js0127 分散在 CRF01-AE 重组型构成的大分支树中(实心小圆圈),进化树水平距离较长(图 1)。

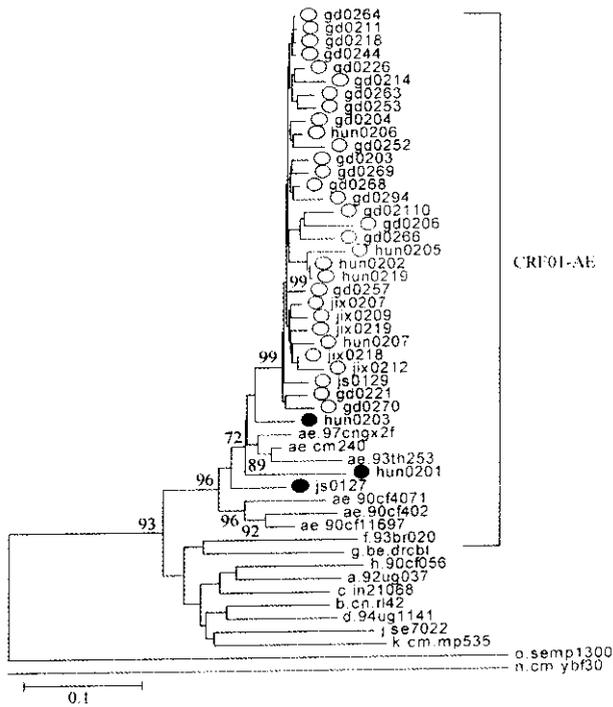


图1 根据 34 份 CRF01-AE 重组型毒株 V3~V4 区序列及各亚型国际代表株序列绘制的系统树 (系统树节点只显示高于 70% 的 Bootstrap 值)

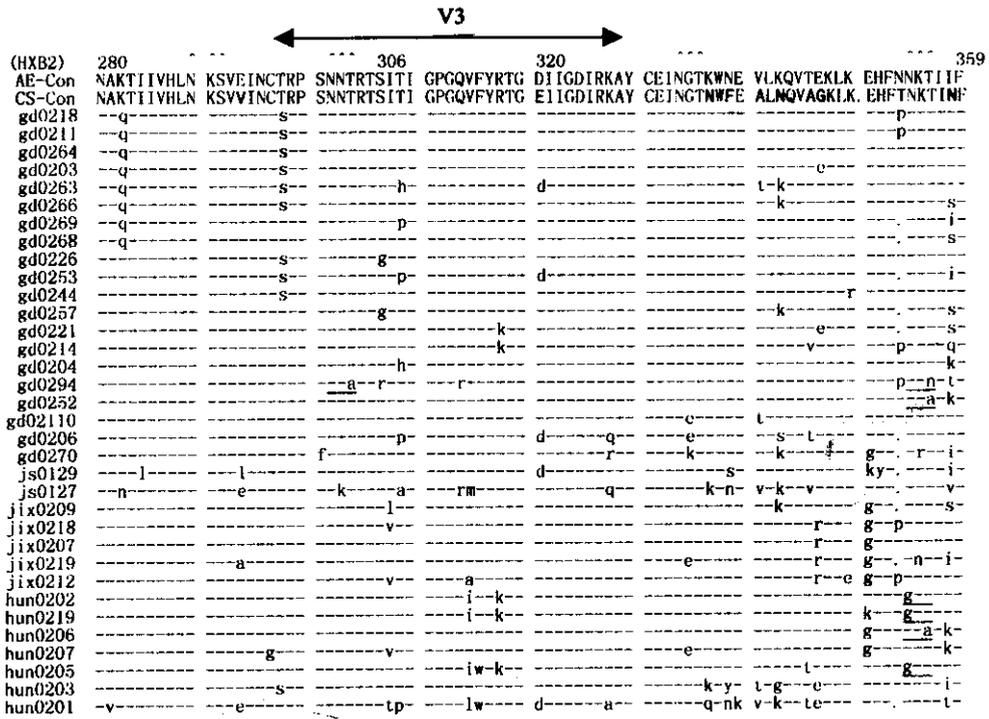
2. env 基因 V3~V4 区氨基酸序列变异的分析和核苷酸基因距离的计算 将 34 份样本 CRF01-AE 重组型 HIV-1 毒株 env 基因 V3~V4 区及其邻近区域氨基酸序列和共享序列与国际标准 CRF01-AE 重组型氨基酸共享序列进行比较,结果发现我国

CRF01-AE 重组型的氨基酸替换主要发生在 V3 下游区(C3)和 V4 区,而 V3 区和 V3 上游区氨基酸序列发生较少的替换。V3 环顶端四肽以 GPGQ 为主,占分析毒株的 87.50% (21/24),只有 2 个毒株(gd0294 和 js0127)的 V3 环顶端四肽为 GPGR。绝大部分毒株在与生物表型密切相关的 306 和 320 位点上没有出现带正电荷的氨基酸,即精氨酸(R)或赖氨酸(K),只有 js0127 在 306 位点上为带正电荷的精氨酸。同时,我们还发现大部分毒株的糖基化位点(N-X-T 和 N-X-S)比较保守,只有少数毒株的某些糖基化位点发生丢失,而且主要发生在 C3 和 V4 区(图 2)。通过对 V3~V4 区及其邻近区域核苷酸基因距离的计算,结果显示 V4 区平均基因距离及离散范围最大,其次是 C3 区、V3 区,最小是 V3 上游区(表 1);与上述各区氨基酸序列的变异具有一致性。

3. env 基因 V3~V4 区 Ks 值、Ka 值及 Ks/Ka 比值的计算 为了分析 V3~V4 区的变异是否与选择压力有关,我们分别计算了整个 V3~V4 区、V3 上游区、V3 区、C3 区和 V4 区的 Ks 值、Ka 值及 Ks/Ka 比值(表 1)。从表 1 可见整个 V3~V4 区、V3 上游区和 C3 区的 Ks 平均值分别显著高于各自基因区的 Ka ($P < 0.001$),而且各区的 Ks/Ka 比值均显著高于 1 ($P < 0.001$)。而 V3 区 Ka 平均值虽然高于 Ks,但差异无显著性 ($P > 0.05$),其 Ks/Ka 比值与 1 比较差异也无显著性 ($P > 0.05$);V4 区 Ks 和 Ka 平均值差异无显著性 ($P > 0.05$),但其 Ks/Ka 比值显著低于 1 ($P < 0.01$)。

讨 论

最近的研究表明,正向选择压力在 HIV-1 的进化中起着重要的作用^[5]。正向选择压力主要作用于 HIV-1 env 基因高度变异区,致使其氨基酸序列多态性的增高^[6]。HIV-1 env 基因 V3 区具有很高的免疫原性,能被中和抗体识别。因此,V3 区氨基酸通过有义突变逃避抗体的识别而赋予 HIV-1 选择优势,从而使 Ka 超过 Ks^[6]。但是,另一方面,V3 区是 HIV-1 与辅助受体结合的关键部位。所以,V3 区的变异除了受到正向选择压力的作用外,同时还受到功能性的限制。由于非融合诱导(NSI)毒株主要受到功能性限制,因此其 V3 环氨基酸序列的多样性和阳性电荷通常低于融合诱导(SI)毒株^[7]。



AE-Con 国际 CRF01-AE 毒株的共享序列 CS-Con 34 份样本的共享序列。三个漏字符号代表糖基化位点，下画线表明糖基化位点丢失

图2 34 份样本 env 基因 V3~V4 的氨基酸序列及其共享序列和国际 CRF01-AE 重组型毒株共享序列

表1 env 不同基因区平均基因距离、Ks 值、Ka 值及 Ks/Ka 比值

基因区	比较序列数	基因距离($\bar{x} \pm s$)	Kd($\bar{x} \pm s$)	Ka($\bar{x} \pm s$)	P1 值	Ks/Ka($\bar{x} \pm s$)	P2 值
V3-upstream~V4	34	6.13 ± 3.44	0.303 ± 0.011	0.254 ± 0.008	<0.001	1.07 ± 0.02	<0.001
V3-upstream	34	2.76 ± 2.53	0.042 ± 0.002	0.021 ± 0.001	<0.001	1.49 ± 0.12	<0.001
V3	34	4.36 ± 2.99	0.037 ± 0.002	0.041 ± 0.001	>0.05	1.09 ± 0.06	>0.050
C3	34	5.95 ± 5.26	0.343 ± 0.014	0.262 ± 0.009	<0.001	1.14 ± 0.04	<0.001
V4	34	12.48 ± 10.69	0.247 ± 0.018	0.213 ± 0.009	>0.050	0.87 ± 0.04	<0.010

我们通过应用系统进化树对 env 基因 V3~V4 区序列分析表明,目前流行于我国的 CRF01-AE 重组型 HIV-1 毒株主要通过吸毒和性途径由泰国传入我国,而且这些毒株大多数具有较高的同源性,在进化上关系密切。另外 3 个流行株(hun0203、hun0201、js0127)在进化树上的水平距离较大,与这 3 个毒株的感染者感染时间较长(5 年左右)以及其中 1 名感染者早期在缅甸居住期间(3 年)有过性乱史有关。我国 CRF01-AE 重组型 env 基因 V3~V4 区氨基酸序列分析发现 C3 和 V4 区的氨基酸变异显著高于 V3 区。核苷酸基因距离的分析亦证实了这一结果。有报道指出,流行于泰国的 CRF01-AE 重组型 NSI 毒株 V3 环顶端四肽以 GPGQ 为主^[8]。我国大部分 CRF01-AE 重组型 HIV-1 毒株 V3 环顶端四肽为 GPGQ(87.50%),这可能是奠基者效应(founder effect)所造成的结果:即从泰国传入我国的 CRF01-AE 重组型毒株,还没有经过长时间的流行,各毒株间在基因水平上的差异还很小,而形成这种奠基者效应^[9]。有报道 V3 区某些位点氨基酸的改变与细胞嗜性有关,其中两个重要的位点第 306 位和 320 位为正电荷时,病毒为 SI 型^[10]。我们的结果显示,绝大部分毒株的第 306 和 320 位点上没有出现带正电荷的氨基酸,推测这些毒株可能为 NSI 型。同时我们研究发现,大部分 CRF01-AE 重组型毒株 V3~V4 区的糖基化位点比较保守,这与 Ichimura 等^[8]对泰国 CRF01-AE 重组型 NSI 毒株报道结果相似。

那么,上述 CRF01-AE 重组型毒株 V3~V4 区的变异是由正向选择压力所造成,还是病毒本身自然突变的结果?为了回答这个问题,我们分别计算了上述各个基因区的 K_s 值、 K_a 值及 K_s/K_a 比值。 K_s 或 K_a 值越高,序列之间的基因离散率越高; K_s/K_a 比值越低,作用于氨基酸变异的选择压力越高,如果 K_s/K_a 比值低于 1,则表明基因变异是在正向选择压力作用下发生的^[11]。由于我国 CRF01-AE 重组型毒株整个 V3~V4 区的 K_s 值显著高于 K_a 值,且 K_s/K_a 比值显著高于 1,因此,总的来说正向选择压力在我国 CRF01-AE 重组型毒株整个 V3~V4 区变异中不起主要的作用,而是病毒本身自然突变的结果。由于目前我国 CRF01-AE 重组型毒株大部分可能为 NSI 型,病毒的变异主要受到功能性的限制,因此,病毒自然突变尽管可以使得核苷酸的多样性(K_a 值和 K_s 值)增高,但不能导致 K_a 值高

于 K_s 值^[12]。值得注意的是,虽然 V3 区 K_a 值高于 K_s 值,但差异无显著性,V4 区 K_s/K_a 比值显著低于 1,提示 V3 区的变异可能与正向选择压力无关,而 V4 区主要受到正向选择压力的作用而发生变异。这与 Shiino 等^[13]报道,NSI 病毒 V3 区序列与其邻近基因区相比受到功能性限制的作用更强相一致。同时,结果还提示分别对单个高度变异区的分析可以更好地揭示他们各自的变异与选择压力的关系。

(本研究得到广东、湖南、江苏、江西等省疾病预防控制中心万卓越、陈曦、羊海涛、易志强等同仁的大力支持与帮助,深表谢意)

参 考 文 献

- Bagnarelli P, Mazzola F, Menzo S, et al. Host-specific modulation of the selective constraints driving human immunodeficiency virus type 1 env gene evolution. *J Virol*, 1999, 73: 3764-3777.
- 梁浩, 魏民, 陈钊, 等. 我国人类免疫缺陷病毒 1 型 B、C 亚型主要流行株外膜蛋白基因 V3~V4 区序列的变异分析. *中华实验和临床病毒学杂志*, 2003, 17: 153-158.
- 邢辉, 潘品良, 苏玲, 等. 1996~1998 年中国流行的 E 亚型 HIV-1 毒株的分子流行病学研究. *中国性病艾滋病防治*, 2002, 8: 200-203.
- Li WH. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J Mol Evol*, 1993, 36: 96-99.
- Ganeshan S, Dickover RE, Korber BT, et al. Human immunodeficiency virus type 1 genetic evolution in children with different rates of development of disease. *J Virol*, 1997, 71: 663-677.
- Bonhoeffer S, Holmes EC, Nowak MA. Causes of HIV diversity. *Nature*, 1995, 376: 125.
- Chesebro B, Wehrly BK, Nishio J, et al. Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell-tropic isolates: definition of critical amino acids involved in cell tropism. *J Virol*, 1992, 66: 6547-6554.
- Ichimura H, Kliks S, Visrutaratna S, et al. Biological, serological, and genetic characterization of HIV-1 subtype E isolates from North Thailand. *AIDS Res Hum Retroviruses*, 1994, 10: 263-269.
- Myers G. HIV: Between past and future. *AIDS Res Hum Retroviruses*, 1994, 10: 1317-1324.
- Fouchier RA, Groenink M, Kootstra NA, et al. Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J Virol*, 1992, 66: 3183-3187.
- Zhang L, Diaz RS, Ho DD, et al. Host-specific driving force in human immunodeficiency virus type 1 evolution in vivo. *J Virol*, 1997, 71: 2555-2561.
- Sato H, Shiino T, Kodaka N, et al. Evolution and biological characterization of human immunodeficiency virus type 1 subtype E gp120 V3 sequences following horizontal and vertical virus transmission in a single family. *J Virol*, 1999, 73: 3551-3559.
- Shiino T, Sato H, Kodaka N, et al. A group of V3 sequences from human immunodeficiency virus type 1 subtype E non-syncytium-inducing, CCR5-using variants are resistant to positive selection pressure. *J Virol*, 2000, 74: 1069-1078.

(收稿日期: 2003-01-28)

(本文编辑: 尹廉)