



## 论文

# 复杂的正选择压力驱动不同细胞嗜性 HIV-1 的进化

张驰宇<sup>①\*</sup>, 丁娜<sup>①</sup>, 陈克平<sup>①</sup>, 杨荣阁<sup>②\*</sup><sup>①</sup> 江苏大学生命科学研究院, 镇江 212013;<sup>②</sup> 中国科学院武汉病毒研究所, 病毒学国家重点实验室, 武汉 430071

\* 联系人, E-mail: zhangcy1999@hotmail.com; ryang@wh.iov.cn

收稿日期: 2010-03-17; 接受日期: 2010-08-01

国家自然科学基金(批准号: 30600352)、国家重点基础研究发展计划(批准号: 2006CB504200)、病毒学国家重点实验室开放基金(批准号: 2009008)和江苏大学“拔尖人才培养工程”资助项目

**摘要** 对 HIV-1 细胞嗜性的研究是理解 HIV-1 传播和发病机制的关键. 通过对 HIV-1 B 和 C 亚型的 R5 和 X4 型病毒的全基因组进行适应性进化分析, 发现 R5 和 X4 病毒经历了不同的进化方式, 并且不同的 HIV-1 基因受到不同的正选择压力, 意味着复杂的自然选择压力驱动 HIV-1 的进化. 分析 HIV-1 Gp120 超变区上的正选择位点, 发现相对于其他超变区, 更多的正选择位点发生在 B 亚型 X4 病毒的 V3 区, B 亚型 R5 病毒的 V2 区以及 C 亚型 X4 病毒的 V1 和 V4 区域. 因为这些区域通常影响和决定 HIV-1 的细胞嗜性, 更多的正选择发生在这些特定的超变区意味着作用于 Gp120 上的选择压力与 Gp120 的受体识别和结合功能有关. 值得注意的是, 无论是 B 和 C 亚型还是 R5 和 X4 型病毒, 显著更多的正选择位点发生在 Gp120 的 C3 区域(33.3%~55.6%,  $P < 0.05$ ), 意味着 C3 区对 HIV-1 的生存和适应比先前认识的更为重要. 另外, 在 R5 和 X4 病毒的 *env* 基因中, 约有一半的正选择位点是相同的, 尤其是 Gp41 上的第 96, 113 和 281 位正选择位点均出现在所分析的 4 种病毒类型中. 这些共同的正选择位点不仅意味着对病毒生存和适应的重要性, 也意味着 R5 和 X4 存在交叉免疫源性位点的可能性, 这对于 AIDS 疫苗的发展具有重要意义.

**关键词**HIV-1 细胞嗜性  
适应性进化  
正选择位点  
R5 和 X4  
Gp120  
抗原表位

HIV-1 进入宿主细胞是通过与细胞表面的主要受体 CD4 和辅助受体(co-receptor)CCR5 或 CXCR4 相互作用实现的. 使用不同的辅助受体感染细胞决定了 HIV-1 不同的细胞嗜性和表型<sup>[1]</sup>. R5 和 X4 是 HIV-1 的两个最主要的细胞嗜性类型, 分别利用 CCR5 和 CXCR4 作为辅助受体进入细胞<sup>[2]</sup>. R5 和 X4 具有不同的病毒学特征, R5 病毒是控制 HIV-1 早期

感染的优势病毒株, 并且这种优势可能一直维持到整个 AIDS 发病过程<sup>[1]</sup>. X4 病毒在外周血单个核细胞中具有比 R5 型更快的复制能力和更高的毒力, 通常在感染后期或感染者发病后出现. X4 病毒的出现往往伴随着急剧的 CD4<sup>+</sup> T 淋巴细胞数目减少, 并且导致加快的 AIDS 疾病进程<sup>[3]</sup>. HIV-1 的细胞嗜性是理解 HIV-1 传播和发病机理的关键<sup>[4,5]</sup>. 在大约一半的

HIV-1 B 亚型感染者中, 伴随着病程的快速发展, HIV-1 的细胞嗜性将从 R5 型向 X4 型转换<sup>[6]</sup>. 然而, HIV-1 从 R5 型向 X4 型的转换却很少发生在 C 亚型感染者中<sup>[4]</sup>. 因此, 比较 HIV-1 B 和 C 亚型病毒的进化将有助于理解 HIV-1 细胞嗜性转换机制.

HIV-1 具有很高的突变率, 并受到来自机体免疫系统的强大选择压力<sup>[7]</sup>. 在 HIV-1 的 M 组、O 组, 甚至 HIV-2 和 SIV 的整个基因组(特别是 *env* 基因)中普遍可以检测到达尔文正选择<sup>[8-14]</sup>. 此外有研究显示, 无论是儿童还是成人 HIV-1 感染中, 作用于 HIV-1 基因组的正选择与 AIDS 的疾病发展有关<sup>[15,16]</sup>. 对 R5 病毒 *env* 基因 C2V3C3 区的研究显示, 相对于快速发病的感染者, 在疾病进程较慢的感染者中病毒更容易受到正选择压力<sup>[7,17]</sup>. 这可能是因为慢进展的感染者具有相对更强的免疫系统, 而快速发病的感染者的免疫系统遭到了破坏. 然而, 另外一个基于 B 亚型 V3 区的研究显示, 合胞体诱导型(syncytium-inducing, SI)病毒比非合胞体诱导型(non-syncytium-inducing, NSI)病毒进化得更快, 意味着 SI(类似 X4)病毒受到比 NSI(类似 R5)病毒更强的正选择压力<sup>[18]</sup>. 这些不一致的结果可能是因为这些研究只分析了 *env* 基因的部分片段. 因此, 要揭示 HIV-1 R5 和 X4 病毒受到的选择压力特征需要对 HIV-1 整个 *env* 基因, 甚至整个基因组进行分析. 本研究对 HIV-1 B 和 C 亚型的 R5 和 X4 型病毒的全基因组进行了适应性进化分析. 结果显示 R5 和 X4 病毒具有明显不同的进化方式, 并且 HIV-1 的不同基因受到不同的选择压力. 在 Gp120 的保守区, 显著更多的正选择位点发生在 C3 区, 意味着 C3 区对 HIV-1 的生存和适应比先前认识的更为重要. 另外, 在 R5 和 X4 型病毒的 *env* 基因中, 约有一半的正选择位点是相同的, 意味着它们不仅在病毒生存和适应中起作用, 也对 AIDS 疫苗的研究具有重要意义.

## 1 材料与方法

### 1.1 序列收集和排序

本研究所用序列均从 Los Alamos 国家实验室(Los Alamos National Laboratory, LANL)的 HIV 序列数据库中下载(<http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html>). 因为数据库中含有明确细胞嗜性信

息的 HIV-1 全长序列非常有限, 首先从数据库下载了所有 HIV-1 B 和 C 亚型的全长序列, 其中包括 624 条 B 亚型和 466 条 C 亚型序列. 为了确定这些病毒序列的细胞嗜性, 用 3 个权威在线预测工具对这些序列的细胞嗜性特征进行预测. 这 3 个在线预测工具分别是 WebPSSM(<http://indra.mullins.microbiol.washington.edu/webpssm/>)<sup>[19]</sup>, Geno2pheno(<http://coreceptor.bioinf.mpi-inf.mpg.de/index.php>)<sup>[20]</sup>和 HIV-1 PhenoPred(<http://yjxy.ujs.edu.cn/R5-X4%20pred.rar>)<sup>[21]</sup>. 对于每条序列, 只有 3 种软件都预测到一致的细胞嗜性结果时, 该序列才被选择用于后续分析. 选择的序列被分成 R5 和 X4 数据集. 因为进化距离过近的序列会导致辨识度过低<sup>[22]</sup>, 每个数据集中进化距离过近的序列被删除. 最后, 在 B 亚型中, 共有 37 个 R5 型和 33 个 X4 型序列被保留, 在 C 亚型中, 共有 28 个 R5 型和 13 个 X4 型序列被保留. 这些序列的 GenBank 登录号在附表 1 中列出.

HIV-1 的不同基因在 HIV-1 的生活周期中起着不同的生物学作用. 为了比较这些基因受到的选择压力的差异, HIV-1 的基因组被分割成相应的 5 个基因片段, 即 *gag*, reverse transcriptase(RT), integrase(IN), *gp120* 和 *gp41*. 在 C 亚型 X4 病毒序列中, 因为删除进化距离过近的序列后, *gag*, RT 和 IN 的序列过少, 不适合做进一步的适应性进化分析, 所以, C 亚型 X4 病毒只保留 *gp120* 和 *gp41* 基因(各 13 个序列)序列用于适应性进化分析.

所有序列经 MEGA 4.0 软件<sup>[23]</sup>的 Clustal W 工具比对后进行手工校正. 4 种病毒类型的各个基因的系统进化树均用最大似然法(PHYML 软件 2.4.4)构建<sup>[24]</sup>. 序列的长度和密码子位置均依据 HXB2 标准序列(<http://www.hiv.lanl.gov/content/sequence/LOCATE/locate.html>)确定.

### 1.2 适应性进化分析

正选择一般通过比较非同义替代率(nonsynonymous substitution rate,  $d_N$ )与同义替代率(synonymous substitution rate,  $d_S$ )的比值来确定.  $d_N/d_S$  的比值( $\omega$ ) $>1$  表示非同义突变被固定的速率大于同义突变, 说明该基因受到了正选择(positive selection),  $\omega=1$  为中性选择(neutral selection),  $\omega<1$  代表负选择(negative selection)<sup>[25]</sup>. PAML 是 Yang 等人<sup>[25,26]</sup>基于最大似然法构建的用于检测单一位点(氨基酸或密码子)是否受到正选

择的经典工具. PAML4.0 中的 6 个模型 M0 (one-ratio), M1a(nearly neutral), M2a(positive selection), M3(discrete), M7( $\beta$ )和 M8( $\beta$  &  $\omega$ )被用于分析 HIV-1 R5 和 X4 型病毒序列的适应性进化<sup>[26]</sup>. 这 6 个模型中, M0, M1a 和 M7 为零假设模型(不允许正选择出现), 分别与 3 个正选择模型 M3, M2a 和 M8 形成嵌套模型用于检测正选择<sup>[27]</sup>. 似然率检验(likelihood ratio test, LRT)通过计算 2 倍的正选择模型与零假设模型最大似然率之差( $2\Delta L=2(L_1-L_0$ )), 并在特定自由度下进行卡方检验, 来确定正选择是否存在<sup>[27]</sup>. 如果 LRT 值具有统计学意义( $P<0.05$ ), 表明零假设被拒绝, 而正选择假设成立, 说明存在正选择. 通过正选择模型(M2a, M3 和 M8)可以鉴定出一系列受到正选择的单个氨基酸(密码子)位点. 每个氨基酸位点的后验概率通过贝叶斯经验检验(Bayes empirical Bayes, BEB)获得<sup>[28]</sup>. 当一个位点的  $\omega > 1$ , 并且其后验概率  $>0.95$  时被鉴定为正选择位点(positively selected site).

因为 M3 模型经常会导致对正选择位点的过度

估计<sup>[27,28]</sup>, M3 模型不被用于鉴定 HIV-1 基因的正选择位点. 为了进一步减少和避免假阳性结果, 只有同时被 M2a 和 M8 检测到的正选择位点才被作为最后确认的正选择位点<sup>[28]</sup>. 为了确定和验证 PAML 获得的结果, 用另外一个公认的软件包 DataMonkey package 做相似分析进行验证<sup>[29]</sup>.

## 2 结果

### 2.1 HIV-1 主要基因的正选择分析

HIV-1 的 R5 和 X4 型病毒在细胞嗜性和复制能力上显示出不同的表型特征<sup>[3]</sup>. 病毒的 Gp120 和 RT 直接与 HIV-1 的表型有关, 前者通过对辅助受体的特异性识别决定细胞嗜性, 而后者直接决定病毒的复制. 因此, 对 HIV-1 B 和 C 亚型的 R5 和 X4 病毒的 *gp120*, *gp41*, RT, IN 和 *gag* 基因进行了适应性进化分析. B 亚型 *gp120* 和 *gp41* 基因的结果分别列于表 1 和 2, C 亚型 *gp120* 和 *gp41* 基因的结果分别列于表 3

表 1 HIV-1 B 亚型 R5 和 X4 病毒 *gp120* 基因的似然值和参数估计

细胞嗜性	模型	lnL	参数估计	2Δl	正选择位点 <sup>a)</sup>
R5	M0	-13809.11	$\omega=0.52$	1998.50	无
	M3	-12809.86	$p_0=0.5982, p_1=0.2658 (p_2=0.1360), \omega_0=0.07, \omega_1=0.85, \omega_2=2.75$	$P=0.0000$	未显示 <sup>b)</sup>
	M1a	-12963.07	$p_0=0.6514 (p_1=0.3486)$	302.66	不允许 <sup>c)</sup>
	M2a	-12811.74	$p_0=0.6169, p_1=0.2620 (p_2=0.1211), \omega_2=2.99$	$P=0.0000$	<b>19T 31T 85V 87V 164S 169V 178K 183P 195S 200V 219A 232T 283T 290T 293E 308R 333I 336A 337K 340N 343K 344Q 347S 360I 362K 363Q 440S 442Q 444R</b>
	M7	-12947.00	$p=0.1795, q=0.3087$	306.48	不允许
	M8	-12793.76	$p_0=0.8659 (p_1=0.1342), p=0.2272, q=0.5413, \omega=2.66$	$P=0.0000$	<b>4K 10L 13W 19T 31T 85V 87V 164S 169V 175F 178K 183P 195S 200V 219A 232T 283T 290T 291S 293E 308R 333I 336A 337K 340N 343K 344Q 346A 347S 360I 362K 363Q 389Q 440S 442Q 444R</b>
	X4	M0	-11488.66	$\omega=0.67$	1420.24
M3	-10778.54	$p_0=0.5291, p_1=0.3181 (p_2=0.1527), \omega_0=0.06, \omega_1=0.89, \omega_2=3.11$	$P=0.0000$	未显示	
M1a	-10925.70	$p_0=0.6162 (p_1=0.3838)$	294.34	不允许	
M2a	-10778.53	$p_0=0.5481, p_1=0.3272 (p_2=0.1247), \omega_2=3.43$	$P=0.0000$	<b>10L 12R 19T 33K 49T 85V 87V 161I 169V 170Q 200V 232T 275V 300N 301N 302N 303T 306R 308R 317F 328Q 333I 337K 343K 347S 354G 363Q 389Q 440S 442Q 467I</b>	
M7	-10922.28	$p=0.1662, q=0.2439$	295.74	不允许	
M8	-10774.41	$p_0=0.8487 (p_1=0.1513), p=0.2050, q=0.3892, \omega_2=2.97$	$P=0.0000$	<b>10L 12R 19T 33K 49T 85V 87V 161I 169V 170Q 192K 200V 208V 232T 275V 291S 300N 301N 302N 303T 306R 308R 316A 317F 318V 328Q 333I 335R 337K 340N 343K 347S 354G 363Q 389Q 440S 442Q 467I</b>	

lnL: 最大似然值的对数; 2Δl: 似然率检验, 指 lnL 之差的 2 倍. P 值代表在自由度下的卡方检验. M0 与 M3 比较的自由度为 3, M1a 与 M2a 以及 M7 与 M8 比较的自由度为 2. a) 后验概率大于  $>95\%$  的正选择位点被列出, 其中后验概率  $>99\%$  的正选择位点用黑体显示; b) 因为 M3 模型经常会导致对正选择位点的过度估计, M3 模型不被用于鉴定正选择位点; c) 零假设模型不允许位点的  $\omega$  值  $>1$

表 2 HIV-1 B 亚型 R5 和 X4 病毒 gp41 基因的似然值和参数估计<sup>a)</sup>

细胞嗜性	模型	lnL	参数估计	2Δl	正选择位点
R5	M0	-10580.74	$\omega=0.57$	1518.12	无
	M3	-9821.68	$p_0=0.6478, p_1=0.2360 (p_2=0.1162), \omega_0=0.08,$ $\omega_1=0.86, \omega_2=3.74$	$P=0.0000$	未显示
	M1a	-10014.68	$p_0=0.7306 (p_1=0.2694)$	384.62	不允许
	M2a	-9822.37	$p_0=0.6731, p_1=0.2198 (p_2=0.1072), \omega_2=3.96$	$P=0.0000$	<b>7L 24M 77K 96A 107S 113N 125N 129S</b> <b>130L 133S 137E 189A 209H 213P 235I 236R</b> <b>239N 253C 281A 306A 318V 321V 325A</b> <b>326C 334R 340L 343I</b>
	M7	-10032.47	$p=0.2010, q=0.3902$	418.48	不允许
	M8	-9823.23	$p_0=0.8789 (p_1=0.12113), p=0.3113,$ $q=0.8201,$ $\omega=3.55$	$P=0.0000$	<b>7L 24M 32Q 77K 96A 107S 113N 119E 125N</b> <b>129S 130L 133S 137E 189A 209H 213P 235I</b> <b>236R 239N 245I 253C 281A 304L 306A 318V</b> <b>321V 325A 326C 334R 340L 343I</b>
X4	M0	-8388.95	$\omega=0.63$	1183.34	无
	M3	-7797.28	$p_0=0.6907, p_1=0.2243 (p_2=0.0497), \omega_0=0.11,$ $\omega_1=1.29, \omega_2=5.08$	$P=0.0000$	未显示
	M1a	-7963.59	$p_0=0.7054 (p_1=0.2946)$	327.88	不允许
	M2a	-7799.65	$p_0=0.6497, p_1=0.2545 (p_2=0.0959), \omega_2=4.57$	$P=0.0000$	<b>24M 96A 108L 109E 110Q 113N 129S 130L</b> <b>133S 163N 167N 189A 209H 210L 212T 213P</b> <b>235I 236R 239N 281A 308A 325A 326C 343I</b>
	M7	-7980.27	$p=0.1591, q=0.2925$	352.78	不允许
	M8	-7803.88	$p_0=0.8908 (p_1=0.1092), p=0.2182,$ $q=0.4835,$ $\omega=4.18$	$P=0.0000$	<b>24M 32Q 96A 101A 108L 109E 110Q 113N</b> <b>129S 130L 133S 163N 167N 189A 209H 210L</b> <b>212T 213P 235I 236R 239N 247D 281A 308A</b> <b>322V 325A 326C 343I</b>

a) 表中参数注释参见表 1

表 3 HIV-1 C 亚型 R5 和 X4 病毒 gp120 基因的似然值和参数估计<sup>a)</sup>

细胞嗜性	模型	lnL	参数估计	2Δl	正选择位点
R5	M0	-10763.38	$\omega=0.47$	1581.90	无
	M3	-9972.43	$p_0=0.7142, p_1=0.2445 (p_2=0.0413),$ $\omega_0=0.10,$ $\omega_1=1.29, \omega_2=5.76$	$P=0.0000$	未显示
	M1a	-10127.16	$p_0=0.7172 (p_1=0.2828)$	305.00	不允许
	M2a	-9974.66	$p_0=0.68058, p_1=0.2529 (p_2=0.0666),$ $\omega_2=4.31$	$P=0.0000$	<b>7Y 10L 84V 181I 240T 281A 295N 300N 335R</b> <b>344Q 346A 350R 365A 389Q 404G 405S</b>
	M7	-10133.41	$p=0.1878, q=0.4031$	312.76	不允许
	M8	-9977.03	$p_0=0.9223 (p_1=0.0777), p=0.2293,$ $q=0.5610,$ $\omega=3.72$	$P=0.0000$	<b>7Y 10L 84V 181I 240T 281A 295N 300N 335R</b> <b>344Q 346A 350R 365A 389Q 404G 405S</b>
X4	M0	-8030.39	$\omega=0.54$	993.48	无
	M3	-7533.65	$p_0=0.6346, p_1=0.2910 (p_2=0.0744),$ $\omega_0=0.10,$ $\omega_1=1.30, \omega_2=6.87$	$P=0.0000$	未显示
	M1a	-7647.70	$p_0=0.6544 (p_1=0.3456)$	224.46	不允许
	M2a	-7535.47	$p_0=0.5929, p_1=0.3111 (p_2=0.0959), \omega_2=5.31$	$P=0.0000$	<b>7Y 137D 138T 140T 141N 173Y 186N 295N</b> <b>300N 320I 335R 344Q 346A 360I 362K 389Q</b> <b>404G 405S 406N 408T 410G 429K 461S 500K</b>
	M7	-7658.08	$p=0.1654, q=0.2725$	242.32	不允许
	M8	-7536.92	$p_0=0.8857 (p_1=0.1144), p=0.2070,$ $q=0.3918,$ $\omega=4.49$	$P=0.0000$	<b>7Y 17G 85V 87V 132T 137D 138T 140T 141N</b> <b>169V 173Y 186N 281A 295N 300N 320I 335R</b> <b>343K 344Q 346A 360I 362K 363Q 389Q 404G</b> <b>405S 406N 408T 410G 429K 460N 461S 500K</b>

a) 表中参数注释参见表 1

和 4. 其他基因的结果列于附表 2~7.

3 个正选择模型(M2a, M3 和 M8)结果显示, 除了 B 亚型 R5 病毒的 IN 基因外, 所有 HIV-1 基因都受到了正选择压力. 对不同的基因进行比较发现, *env* (*gp120* 和 *gp41*)和 *gag* 基因受到的正选择压力要强于 RT 和 IN 基因受到的正选择压力. 特别是 *env* 基因, 其受到的正选择压力最强<sup>[8-11]</sup>. 另外, 不同基因上鉴定的正选择位点的数量也支持 *env* 和 *gag* 基因受到更强的正选择压力. 这些结果表明, 不同的 HIV-1 基因在病毒进化中经历不同的正选择模式.

### 2.2 *env(gp120* 和 *gp41)*基因上正选择位点的鉴定

在 B 亚型 R5 和 X4 病毒的 *gp120* 基因上分别有 29 和 31 个正选择位点被鉴定, 在 C 亚型 R5 和 X4 病毒的 *gp120* 基因上分别有 16 和 24 个正选择位点被鉴定(表 5). 无论 B 亚型和 C 亚型, 在 X4 病毒中鉴定的正选择位点数均多于 R5 病毒, 说明 X4 病毒的 *gp120* 受到了比 R5 病毒更强的正选择压力. 然而对于 *gp41* 基因来说, 27 和 17 个正选择位点分别发生在 B 和 C 亚型的 R5 病毒中, 均多于 B(24 个)和 C(11 个)亚型的 X4 病毒中的正选择位点数(表 5). 这说明 R5

病毒的 *gp41* 基因受到比 X4 病毒更强的正选择压力. 虽然整个 R5 和 X4 病毒的 *env* 基因的正选择位点数没有明显不同, 但是 R5 和 X4 病毒的 *gp120* 和 *gp41* 基因却经历了不同的进化方式.

另一方面, 如果从亚型的角度分析, B 亚型 *env* 基因的正选择位点要多于 C 亚型的正选择位点数(表 5). 对于 *gp120* 基因, B 亚型 R5 和 X4 病毒分别有 29 和 31 个正选择位点, 均多于 C 亚型 R5 和 X4 病毒的 16 和 24 个位点. 对于 *gp41* 基因, B 亚型 R5 和 X4 病毒分别有 27 和 24 个正选择位点, 也均多于 C 亚型 R5 和 X4 病毒的 17 和 11 个位点. 这些结果说明 B 亚型 *env* 基因在进化中受到比 C 亚型更强的正选择压力.

### 2.3 *gag* 和 *pol*基因上正选择位点的鉴定

HIV-1 *gag* 基因编码 4 个结构蛋白, 这些蛋白是人体免疫系统识别的重要靶点. 在 B 和 C 亚型的 R5 病毒中, 分别有 9 和 10 个正选择位点发生在 *gag* 基因中, 多于 B 亚型 X4 病毒中鉴定的 6 个正选择位点(附表 2 和 3). 有趣的是, 其中 4 个正选择位点(91, 138, 280 和 374)在 B 亚型的 R5 和 X4 病毒中是相同的. 尤其是第 91 和 138 位位点在 C 亚型的 R5 病毒中也受

表 4 HIV-1 C 亚型 R5 和 X4 病毒 *gp41* 基因的似然值和参数估计<sup>a)</sup>

细胞嗜性	模型	lnL	参数估计	2Δl	正选择位点
R5	M0	-7964.17	$\omega=0.53$	811.42 $P=0.0000$	无
	M3	-7558.46	$p_0=0.6953, p_1=0.2258 (p_2=0.0789),$ $\omega_0=0.1083, \omega_1=1.1385, \omega_2=3.39$		未显示
	M1a	-7629.70	$p_0=0.7019 (p_1=0.2982)$	141.48 $P=0.0000$	不允许
	M2a	-7558.76	$p_0=0.6775, p_1=0.2281 (p_2=0.0944),$ $\omega_2=3.12$		<b>96A 101A 108L 109E 113N 129D 144K</b> <b>156A 160N 163N 212T 239N 256S 266I</b> <b>281A 321V 345L</b>
	M7	-7652.02	$p=0.2013, q=0.3581$	174.02 $P=0.0000$	不允许
	M8	-7565.01	$p_0=0.8674 (p_1=0.1326), p=0.3539,$ $q=0.9639, \omega_2=2.66$		<b>96A 101A 107S 108L 109E 113N 129D</b> <b>137E 140N 144K 154K 156A 160N 163N</b> <b>210L 212T 232D 239N 256S 266I 281A</b> <b>321V 324G 326C 345L</b>
X4	M0	-4927.55	$\omega=0.57$	430.82 $P=0.0000$	无
	M3	-4712.14	$p_0=0.6766, p_1=0.2808 (p_2=0.0427),$ $\omega_0=0.09, \omega_1=1.46, \omega_2=5.67$		未显示
	M1a	-4758.74	$p_0=0.6516 (p_1=0.3484)$	86.78 $P=0.0000$	不允许
	M2a	-4715.35	$p_0=0.6277, p_1=0.2940 (p_2=0.0783),$ $\omega_2=3.97$		<b>96A 108L 109E 113N 133S 163N 210L</b> <b>258H 281A 321V 330R</b>
	M7	-4768.49	$p=0.0970, q=0.1519$	98.00 $P=0.0000$	不允许
	M8	-4719.49	$p_0=0.9026 (p_1=0.0975), p=0.1526,$ $q=0.3012,$ $\omega=3.64$		<b>84I 96A 108L 109E 113N 133S 163N</b> <b>210L 258H 281A 321V 330R</b>

a) 表中参数注释参见表 1

到正选择. 这些在不同细胞嗜性特征病毒的 *gag* 基因上均受到正选择的位点可能对 HIV-1 的适应具有重要作用.

HIV-1 的 *pol* 基因编码逆转录酶(RT)和整合酶(IN), 它们是 HIV 生活周期中的关键酶. 比较 R5 和 X4 病毒中这两个酶受到的选择压力将有助于了解 R5 和 X4 病毒在复制力上的差异. 在 HIV-1 B 亚型中, RT 的第 162 和 376 位位点在 R5 病毒中受到正选择, 而第 211 位位点在 X4 病毒中受到正选择(附表 4 和 5). 对于 IN, 第 118 和 123 位位点在 X4 病毒中受到正选择(表 S4), 但在 R5 病毒中没有检测到正选择位点. 在 C 亚型 R5 病毒中, RT 上第 123, 344 和 377 号位点和 IN 上第 50, 72 和 125 位位点受到正选择(附表 4~7).

## 2.4 HIV-1 *env* 基因上正选择位点的分布

HIV-1 表面糖蛋白 Env 包括 Gp120 和 Gp41. Gp120 不仅决定病毒的细胞嗜性, 也是宿主免疫反应的最重要靶点. Gp120 上包含 5 个不连续的保守区(C1~C5)和 5 个不连续的超变区(V1-V5). 比较正选择位点在 Gp120 上的分布发现, 相对于超变区占整个 Gp120 的比例(28.8%), 明显更多的正选择位点发生在 B(41.9%,  $P=0.041$ )和 C 亚型(62.5%,  $P=0.0002$ )X4 病毒的 Gp120 超变区(表 5). 表明 X4 病毒 Gp120 的超变区遭受更强的正选择压力.

V3 区决定 HIV-1 辅助受体的使用, 是决定病毒细胞嗜性的关键区域, 也是引起中和抗体的关键免疫表位<sup>[30,31]</sup>. 本研究进一步分析了正选择位点在 5 个超变区的分布(图 1). 在 B 亚型 X4 病毒中, 61.5%的

正选择位点出现在 V3 区, 显著高于 V3 占整个超变区的比例 24.5%( $P=0.0011$ ). 该结果表明, V3 区受到更强的正选择压力<sup>[7,18]</sup>. 在 C 亚型 X4 病毒中, 更多的(66.7%)正选择位点出现在 V1 和 V4, 显著高于这两个超变区在整个超变区所占的比例(34%,  $P=0.0316$ )(表 5). 另外, 在 B 亚型 R5 病毒中, 有 83.3%的位点出现在 V2 区, 显著高于 V2 占整个超变区的比例 26.5%( $P=0.0013$ ).

有趣的是, 当分析正选择位点在保守区的分布时, 发现在所有被分析的 4 种病毒类型中, C3 区明显包含了比其他保守区更多的正选择位点(B 亚型 R5 病毒: 43.5%,  $P<0.0001$ ; B 亚型 X4 病毒: 33.3%,  $P=0.0206$ ; C 亚型 R5 病毒: 45.5%,  $P=0.0032$ ; C 亚型 X4 病毒: 55.6%,  $P=0.0004$ )(表 5). 该结果表明, C3 区在 Gp120 进化过程中起到比先前认识的更为重要的作用. 在 Gp41 中, 只有很少的正选择位点(9.1%~22.2%)发生在两个七肽重复区(heptad repeat region, HR), 正选择位点在 HR1 和 HR2 区的分布比例甚至低于 2 个 HR 区在整个 Gp41 中所占的比例(24.3%)(表 5, 图 1).

另一方面, 发现无论是 HIV-1 B 还是 C 亚型, *env* 基因上约一半的正选择位点在 R5 和 X4 病毒中是一致的(图 1). 例如, 在 B 亚型中, 共有 56 个正选择位点发生在 R5 病毒中. 然而, 其中 30 个同时也发生在 X4 病毒中. 而在 C 亚型中, 有 16 个正选择位点既发生在 R5 病毒, 也发生在 X4 病毒中. 另外, 在所有分析的 4 种病毒类型中, Gp41 上的第 96, 113 和 281 位位点均受到正选择(图 1).

表 5 正选择位点在 Gp120 和 Gp41 上的分布比较

亚型	细胞嗜性	氨基酸位点特征	Gp120						Gp41					
			超变区 <sup>a)</sup>					C3 <sup>a)</sup>	C1~C5	合计	HR1+HR2	其他区域	合计	
			V1	V2	V3	V4	V5							Total <sup>b)</sup>
B	R5	正选择位点	0	<b>5</b>	1	0	0	6	<b>10</b>	23	29	6	21	27
		非正选择位点	26	<b>34</b>	35	34	12	141	<b>43</b>	341	482	78	240	318
	X4	正选择位点	0	3	<b>8</b>	1	1	<b>13</b>	<b>6</b>	18	31	3	21	24
		非正选择位点	26	36	<b>28</b>	33	11	<b>134</b>	<b>47</b>	346	480	81	240	321
C	R5	正选择位点	0	1	1	3	0	5	<b>5</b>	11	16	2	15	17
		非正选择位点	26	38	35	31	12	143	<b>48</b>	353	485	82	246	328
	X4	正选择位点	4	2	2	6	1	<b>15</b>	<b>5</b>	9	24	1	10	11
		非正选择位点	22	37	34	28	11	<b>132</b>	<b>48</b>	355	487	83	251	334

a) 黑体数字表示当某一个超变区与剩余其他超变区比较时, 或者是整个超变区(V1~V5)与保守区(C1~C5)比较时具有统计学显著差异( $P<0.05$ , Fisher 精确检验); b) 黑体数字表示 C3 区与剩余其他保守区(C1, C2, C4 和 C5)比较具有统计学显著差异( $P<0.05$ , Fisher 精确检验)

### 3 讨论

通过比较作用于 HIV-1 B 和 C 亚型, 以及 R5 和 X4 病毒几个关键基因上的选择压力, 发现 *gag* 和 *env(gp120* 和 *gp41)* 基因受到更高的自然选择压力. 这表明 HIV-1 的各个基因遭受不同的选择压力<sup>[8]</sup>. 比较发生在 R5 和 X4 病毒 *env* 基因上的正选择位点发现, 发生在 X4 病毒 *gp120* 基因上的正选择位点略多于 R5 病毒(表 5). 这与先前发现 SI 病毒 V3 区比 NSI 病毒 V3 区受到更强的正选择压力一致<sup>[18]</sup>. 然而不同于 *gp120*, 在 *gp41* 基因上, R5 病毒受到的正选择压力略强于 X4 病毒. 这表明 R5 和 X4 病毒经历不同的进化模式.

在未接受治疗的 HIV-1 感染者中, 病毒的进化至少受到来自两方面的选择压力, 即宿主免疫系统和靶细胞资源<sup>[4]</sup>. 宿主的免疫反应主要包括体液免疫和细胞免疫, 通常被认为是 HIV-1 适应性进化的最关键驱动力<sup>[7]</sup>. *gag* 基因编码病毒的结构蛋白, 它们并不涉及病毒的复制和辅助受体识别. 在 3 种 HIV-1 病毒类型的 *gag* 基因上共有 19 个正选择位点被鉴定(附表 2 和 3), 并且这些位点的定位至少与 3 种免疫表位(抗体、细胞毒性 T 淋巴细胞(cytotoxic

T-lymphocyte, CTL)和 T 辅助细胞)中的一种有关. 例如, 83.3% 和 66.7% 的正选择位点分别与 CTL 和 T 辅助细胞表位有关. 这说明 *gag* 基因受到的选择压力主要来自宿主的免疫反应. 特别是在所有 Gag 的正选择位点中, 第 91 和 138 位正选择位点在 3 种 HIV-1 病毒类型中均被发现, 表明这两个位点对 HIV-1 的适应性进化非常重要. 之前的研究证明, Gag 蛋白第 30 位氨基酸的突变能够使 HIV 或 SIV 在宿主中获得一种种属特异性的复制优势<sup>[32]</sup>. 至于 Gag 蛋白第 91 和 138 位氨基酸对 HIV-1 适应性的作用则需要进一步的点突变实验来验证.

逆转录酶(RT)是 HIV-1 中负责复制的关键酶<sup>[4]</sup>. X4 病毒通常比 R5 病毒具有更高的复制力. 在 3 种病毒类型的 RT 中, 共有 6 个正选择位点被发现. 这些位点都位于 RT 的 DNA 聚合酶结构域<sup>[33]</sup>, 并且在 R5 和 X4 病毒中鉴定的正选择位点完全不同, 说明这些正选择位点可能与 R5 和 X4 病毒不同的复制特性有关. 另外, 这些位点都定位于 CTL 免疫表位中, 说明细胞免疫反应也参与了 RT 的进化.

HIV-1 表面糖蛋白(Env)暴露在病毒颗粒的表面, 对 HIV-1 的感染和生存起着关键性作用. 它不仅决定 HIV-1 的细胞嗜性, 并且是体液和细胞免疫反应的主

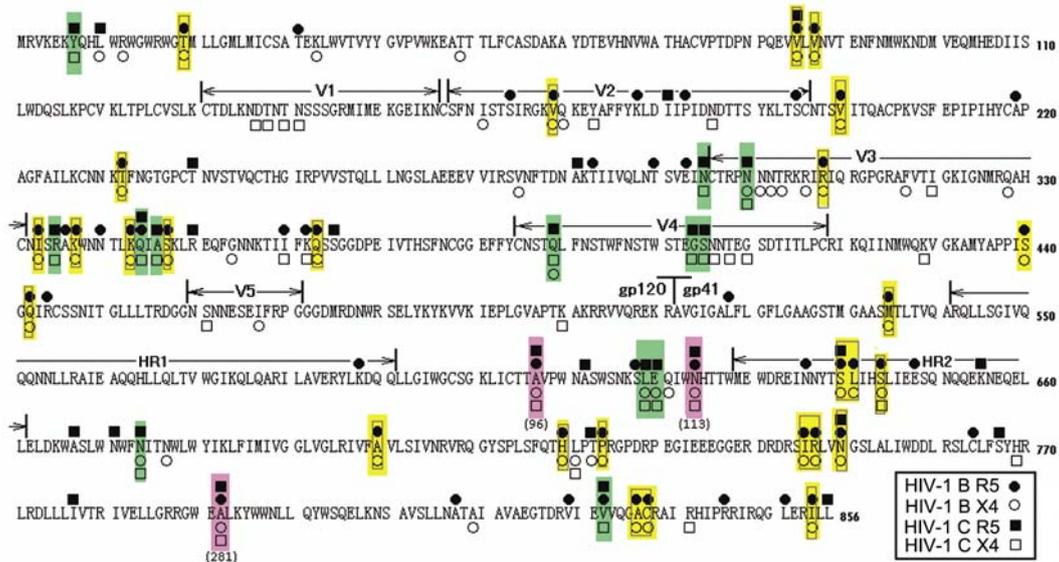


图 1 HIV-1 B 和 C 亚型 R5 和 X4 病毒 Env 上正选择位点的分布

黄色阴影代表 B 亚型 R5 与 X4 型病毒共同的正选择位点; 绿色阴影代表 C 亚型 R5 与 X4 型病毒共同的正选择位点; 粉红色阴影代表 4 个病毒类型中共同存在的正选择位点

要靶点。此外, 在 HIV-1 基因组中, *env* 基因的突变率最高, 这赋予 HIV-1 较高的逃逸宿主免疫反应的潜力<sup>[34,35]</sup>。因此, *env* 基因的适应性进化可能会涉及多种选择因素, 例如细胞资源、宿主免疫反应, 甚至病毒自身进化力等<sup>[4,8]</sup>。

HIV-1 的细胞嗜性由 Gp120 的超变区决定, 特别是 V3 区在 HIV-1 的细胞嗜性决定中扮演关键角色<sup>[30,36,37]</sup>。V3 外的其他超变区, 如 V1, V2 和 V4 区可以影响 HIV-1 辅助受体的使用<sup>[38-44]</sup>。比较正选择位点在 Gp120 上的定位发现, 在 X4 病毒中, 显著更多的正选择位点位于超变区( $P < 0.05$ )。然而, 类似的正选择分布并没有出现在 R5 病毒中。进一步分析显示, 更多的正选择位点发生在 B 亚型 R5 病毒的 V2 区(83.3%,  $P = 0.0013$ )和 X4 病毒的 V3 区(61.5%,  $P = 0.0011$ ), 以及 C 亚型 X4 病毒的 V1V4 区(66.7%,  $P = 0.0316$ )(表 5)。这些结果清楚地表明, 发生在超变区的正选择位点与 Gp120 的辅助受体的识别和结合功能紧密相关。此外, V1V2, V4 和 V5 区已经被证实与 HIV-1 病毒的自身中和反应有关。因此, 体液免疫介导的正选择压力也参与了 Gp120 的进化。另一方面, 相对于其他保守区, 明显更多的正选择位点(33.3%~55.6%,  $P < 0.05$ )定位于 Gp120 的 C3 区, 表明 C3 区在 Gp120 功能上的作用比

先前认识的更重要。在 HIV-1 C 亚型感染早期, Gp120 的 C3 区可以通过与 V4 区结合形成一个特殊的结构基序来诱导自身中和免疫反应的产生<sup>[45]</sup>。因此, 更多的正选择位点位于 C3 区也支持了体液免疫介导的正选择压力在 Gp120 进化中起重要作用的观点。

病毒表面蛋白上的七肽重复区(HR)在介导病毒膜融合中起重要作用<sup>[46]</sup>。与 SARS-CoV 锚定(spike, S)蛋白的 S2 结构域相似, HIV-1 的 Gp41 也包括 2 个 HR。然而, 在 HIV-1 Gp41 中, 只有很小比例(9.1%~22.2%)的正选择位点发生在这 2 个 HR 区(表 5, 图 1), 显然不支持膜融合作为一种主要的选择因素参与 HIV-1 Gp41 的进化<sup>[47]</sup>。然而, Gp41 的第 96, 113 和 281 位在 4 种病毒类型中均被检测为正选择位点, 并且其中的第 96 和 113 位位点还位于 Gp41 的两个 HR 区之间, 可能影响 Gp41 的膜融合功能。这表明膜融合功能可能只是参与 HIV-1 Gp41 进化的一种次要选择因素。此外, 同时发现在 *env* 基因上的正选择位点中, 有接近一半的位点既出现在 R5 病毒, 也出现在 X4 病毒中(图 1)。这些共同的正选择位点不仅说明它们对 R5 和 X4 病毒的生存和适应非常重要, 也意味着免疫反应可能针对 R5 和 X4 病毒的相同区域发挥作用, 这对 AIDS 疫苗的研究具有重要意义。

## 参考文献

- Berger E A, Murphy P M, Farber J M. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu Rev Immunol*, 1999, 17: 657-700
- Berger E A, Doms R W, Fenyo E M, et al. A new classification for HIV-1. *Nature*, 1998, 391: 240
- Bjorndal A, Deng H, Jansson M, et al. Coreceptor usage of primary human immunodeficiency virus type 1 isolates varies according to biological phenotype. *J Virol*, 1997, 71: 7478-7487
- Regoes R R, Bonhoeffer S. The HIV coreceptor switch: a population dynamical perspective. *Trends Microbiol*, 2005, 13: 269-277
- Moore J P, Kitchen S G, Pugach P, et al. The CCR5 and CXCR4 coreceptors—central to understanding the transmission and pathogenesis of human immunodeficiency virus type 1 infection. *AIDS Res Hum Retroviruses*, 2004, 20: 111-126
- Connor R I, Sheridan K E, Ceradini D, et al. Change in coreceptor use correlates with disease progression in HIV-1--infected individuals. *J Exp Med*, 1997, 185: 621-628
- Ross H A, Rodrigo A G. Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. *J Virol*, 2002, 76: 11715-11720
- Choisy M, Woelk C H, Guegan J F, et al. Comparative study of adaptive molecular evolution in different human immunodeficiency virus groups and subtypes. *J Virol*, 2004, 78: 1962-1970
- Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, 1998, 148: 929-936
- Yang W, Bielawski J P, Yang Z. Widespread adaptive evolution in the human immunodeficiency virus type 1 genome. *J Mol Evol*, 2003, 57: 212-221
- Travers S A, O'Connell M J, McCormack G P, et al. Evidence for heterogeneous selective pressures in the evolution of the *env* gene in

- different human immunodeficiency virus type 1 subtypes. *J Virol*, 2005, 79: 1836–1841
- 12 Zanotto P M, Kallas E G, de Souza R F, et al. Genealogical evidence for positive selection in the *nef* gene of HIV-1. *Genetics*, 1999, 153: 1077–1089
- 13 de Oliveira T, Salemi M, Gordon M, et al. Mapping sites of positive selection and amino acid diversification in the HIV genome: an alternative approach to vaccine design? *Genetics*, 2004, 167: 1047–1058
- 14 Soares A E, Soares M A, Schrago C G. Positive selection on HIV accessory proteins and the analysis of molecular adaptation after interspecies transmission. *J Mol Evol*, 2008, 66: 598–604
- 15 Carvajal-Rodriguez A, Posada D, Perez-Losada M, et al. Disease progression and evolution of the HIV-1 *env* gene in 24 infected infants. *Infect Genet Evol*, 2008, 8: 110–120
- 16 Leal E, Janini M, Diaz R S. Selective pressures of human immunodeficiency virus type 1 (HIV-1) during pediatric infection. *Infect Genet Evol*, 2007, 7: 694–707
- 17 Williamson S. Adaptation in the *env* gene of HIV-1 and evolutionary theories of disease progression. *Mol Biol Evol*, 2003, 20: 1318–1325
- 18 Yamaguchi Y, Gojbori T. Evolutionary mechanisms and population dynamics of the third variable envelope region of HIV within single hosts. *Proc Natl Acad Sci USA*, 1997, 94: 1264–1269
- 19 Jensen M A, Li F S, van't Wout A B, et al. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 *env* V3 loop sequences. *J Virol*, 2003, 77: 13376–13388
- 20 Sing T, Low A J, Beerwinkel N, et al. Predicting HIV coreceptor usage on the basis of genetic and clinical covariates. *Antivir Ther*, 2007, 12: 1097–1106
- 21 Xu S, Huang X, Xu H, et al. Improved prediction of coreceptor usage and phenotype of HIV-1 based on combined features of V3 loop sequence using random forest. *J Microbiol*, 2007, 45: 441–446
- 22 Nozawa M, Suzuki Y, Nei M. Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. *Proc Natl Acad Sci USA*, 2009, 106: 6700–6705
- 23 Tamura K, Dudley J, Nei M, et al. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol*, 2007, 24: 1596–1599
- 24 Guindon S, Lethiec F, Duroux P, et al. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res*, 2005, 33: W557–559
- 25 Yang Z, Nielsen R, Goldman N, et al. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, 2000, 155: 431–449
- 26 Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*, 2007, 24: 1586–1591
- 27 Anisimova M, Bielawski J P, Yang Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol*, 2001, 18: 1585–1592
- 28 Anisimova M, Bielawski J P, Yang Z. Accuracy and power of bayes prediction of amino acid sites under positive selection. *Mol Biol Evol*, 2002, 19: 950–958
- 29 Pond S L, Frost S D. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics*, 2005, 21: 2531–2533
- 30 Hwang S S, Boyle T J, Lyerly H K, et al. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science*, 1991, 253: 71–74
- 31 Javaherian K, Langlois A J, LaRosa G J, et al. Broadly neutralizing antibodies elicited by the hypervariable neutralizing determinant of HIV-1. *Science*, 1990, 250: 1590–1593
- 32 Wain L V, Bailes E, Bibollet-Ruche F, et al. Adaptation of HIV-1 to its human host. *Mol Biol Evol*, 2007, 24: 1853–1860
- 33 Sarafianos S G, Das K, Tantillo C, et al. Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *EMBO J*, 2001, 20: 1449–1461
- 34 Parren P W, Moore J P, Burton D R, et al. The neutralizing antibody response to HIV-1: viral evasion and escape from humoral immunity. *AIDS*, 1999, 13: S137–S162
- 35 Klenerman P, Wu Y, Phillips R. HIV: current opinion in escapology. *Curr Opin Microbiol*, 2002, 5: 408–413
- 36 Fouchier R A, Groenink M, Kootstra N A, 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
- 37 Xiao L, Owen S M, Goldman I, et al. CCR5 coreceptor usage of non-syncytium-inducing primary HIV-1 is independent of phylogenetically distinct global HIV-1 isolates: delineation of consensus motif in the V3 domain that predicts CCR-5 usage. *Virology*, 1998, 240: 83–92

- 38 Boyd M T, Simpson G R, Cann A J, et al. A single amino acid substitution in the V1 loop of human immunodeficiency virus type 1 gp120 alters cellular tropism. *J Virol*, 1993, 67: 3649–3652
- 39 Smyth R J, Yi Y, Singh A, et al. Determinants of entry cofactor utilization and tropism in a dualtropic human immunodeficiency virus type 1 primary isolate. *J Virol*, 1998, 72: 4478–4484
- 40 Groenink M, Fouchier R A, Broersen S, et al. Relation of phenotype evolution of HIV-1 to envelope V2 configuration. *Science*, 1993, 260: 1513–1516
- 41 Pastore C, Nedellec R, Ramos A, et al. Human immunodeficiency virus type 1 coreceptor switching: V1/V2 gain-of-fitness mutations compensate for V3 loss-of-fitness mutations. *J Virol*, 2006, 80: 750–758
- 42 Sullivan N, Thali M, Furman C, et al. Effect of amino acid changes in the V1/V2 region of the human immunodeficiency virus type 1 gp120 glycoprotein on subunit association, syncytium formation, and recognition by a neutralizing antibody. *J Virol*, 1993, 67: 3674–3679
- 43 Ghaffari G, Tuttle D L, Briggs D, et al. Complex determinants in human immunodeficiency virus type 1 envelope gp120 mediate CXCR4-dependent infection of macrophages. *J Virol*, 2005, 79: 13250–13261
- 44 Cho M W, Lee M K, Carney M C, et al. Identification of determinants on a dualtropic human immunodeficiency virus type 1 envelope glycoprotein that confer usage of CXCR4. *J Virol*, 1998, 72: 2509–2515
- 45 Bunnik E M, Pisas L, van Nuenen A C, et al. Autologous neutralizing humoral immunity and evolution of the viral envelope in the course of subtype B human immunodeficiency virus type 1 infection. *J Virol*, 2008, 82: 7932–7941
- 46 Chambers P, Pringle C R, Easton A J. Heptad repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins. *J Gen Virol*, 1990, 71: 3075–3080
- 47 Zhang C Y, Wei J F, He S H. Adaptive evolution of the spike gene of SARS coronavirus: changes in positively selected sites in different epidemic groups. *BMC Microbiol*, 2006, 6: 88

附表 1 本文所用 HIV-1 基因组的 GenBank 登录号

亚型	细胞嗜性	序列数目	GenBank 登录号
HIV-1 B	R5	37	AB286956, AB253432, AF003888, AF042101, AF224507, AY173952, AY037282, EU576191, AY586543, AY713412, AY835748, AY713411, EU574998, AY561236, AY970946, AY839827, AY857022, D10112, DQ854714, DQ837381, DQ886031, FJ469746, EF363124, EF637046, EF514699, EF637049, EU786675, FJ460501, FJ469770, FJ495937, FJ469703, FJ469731, M93258, U23487, U63632, FJ496085, FJ496150, K02007
	X4	33	EU281726, K02013, AB287363, AB287365, AF049494, AF086817, AF146728, AY037268, AY736821, AY173956, AY180905, AY560108, AY835767, AY835768, D86068, DQ127534, DQ396398, DQ823363, EF514712, FJ469686, FJ469692, FJ469736, FJ469737, FJ469739, FJ469748, FJ469753, FJ469759, L02317, L31963, M17449, M26727, FJ496166
HIV-1 C	R5	28	AB254141, DQ369991, AY734550, DQ275642, EU786673, AY878054, AF286227, AY945738, FJ496185, U46016, AY713414, AF110978, AF110981, AF286224, AF286231, AY444800, AY463217, AY563170, AF286233, AF286234, AF290027, AY043176, AF391231, AY118165, AY253303, AY228556, AY228557, AY253321
	X4	13	FJ846637, FJ846642, AY878064, DQ093600, AY529666, FJ846647, AY529678, DQ382362, DQ382372, DQ382378, AY529677, AY529673, AF411966

附表2 HIV-1 B亚型 R5 和 X4 病毒 gag 基因的似然值和参数估计<sup>a)</sup>

细胞嗜性	模型	lnL	参数估计	2Δl	正选择位点
R5	M0(one-ratio)	-10957.60	$\omega=0.24$		无
	M3(discrete)	-10293.04	$p_0=0.7101, p_1=0.2228 (p_2=0.0671), \omega_0=0.03, \omega_1=0.56, \omega_2=2.58$	1329.12 $P=0.0000$	<b>12E 30K 67S 84T 91R 138I 146A 215V 223I 252N 280T 374A 375T 389I 425D 441Y 473P 478P 483L 487T</b>
	M1a(Nearly neutral)	-10382.84	$p_0=0.7869 (p_1=0.2131)$	145.4 $P=0.0000$	不允许
	M2a(Positive selection)	-10310.14	$p_0=0.7719, p_1=0.1899 (p_2=0.0381), \omega_2=3.82$		<b>84T 91R 138I 223I 280T 374A 389I 473P 483L</b>
	M7(beta)	-10364.21	$p=0.1217, q=0.4289$	157.08 $P=0.0000$	不允许
	M8(beta& $\omega$ )	-10285.67	$p_0=0.9509 (p_1=0.0491), p=0.1539, q=0.6937, \omega=3.06$		<b>67S 84T 91R 138I 146A 223I 280T 374A 375T 389I 425D 473P 478P 483L</b>
	M0(one-ratio)	-8991.79	$\omega=0.25$		无
	M3(discrete)	-8566.49	$p_0=0.6919, p_1=0.2246 (p_2=0.08350), \omega_0=0.02, \omega_1=0.49, \omega_2=2.06$	850.6 $P=0.0000$	<b>12E 15R 30K 67S 79Y 84T 91R 93E 102D 119D 125S 124N 138I 219H 223I 252N 280T 374A 389I 403R 418K 441Y 473P 478P 473L 477T</b>
X4	M1a(Nearly neutral)	-8607.63	$p_0=0.7827 (p_1=0.2173)$	51.28 $P=0.0000$	不允许
	M2a(Positive selection)	-8581.99	$p_0=0.7777, p_1=0.0448 (p_2=0.1775), \omega_2=2.86$		<b>67S 91R 138I 280T 374A 478P</b>
	M7(beta)	-8603.18	$p=0.1197, q=0.4075$	73.92 $P=0.0000$	不允许
	M8(beta& $\omega$ )	-8566.22	$p_0=0.9279 (p_1=0.0720), p=0.1844, q=0.9994, \omega=2.20$		<b>67S 84T 91R 125S 138I 223I 280T 374A 478P 473L</b>

a) 后验概率大于>95%的正选择位点; 后验概率>99%的正选择位点用黑体显示

附表3 HIV-1 C亚型 R5 病毒 gag 基因的似然值和参数估计<sup>a)</sup>

模型	lnL	参数估计	2Δl	正选择位点
M0 (one-ratio)	-9276.18	$\omega=0.27$	1131.26 $P=0.0000$	无
M3 (discrete)	-8710.55	$p_0=0.7768, p_1=0.1893 (p_2=0.0338), \omega_0=0.05, \omega_1=0.93, \omega_2=3.99$		<b>28K 54S 79Y 90Q 91R 138I 146A 241S 371T 440S 458P</b>
M1a (Nearly neutral)	-8778.63	$p_0=0.7926 (p_1=0.2074)$	135.54 $P=0.0000$	不允许
M2a (Positive selection)	-8710.86	$p_0=0.7825, p_1=0.1852 (p_2=0.0324), \omega_2=4.14$		<b>28K 79Y 90Q 91R 138I 146A 241S 371T 440S 458P</b>
M7 (beta)	-8782.96	$p=0.1194, q=0.4021$	147.9 $P=0.0000$	不允许
M8 (beta& $\omega$ )	-8709.01	$p_0=0.9639 (p_1=0.0360), p=0.1426, q=0.5506, \omega=3.71$		<b>28K 54S 79Y 90Q 91R 138I 146A 223V 241S 371T 440S 458P</b>

a) 后验概率大于>95%的正选择位点; 后验概率>99%的正选择位点用黑体显示

附表4 HIV-1 B亚型 R5 和 X4 病毒 RT 基因的似然值和参数估计<sup>a)</sup>

细胞嗜性	模型	lnL	参数估计	2Δl	正选择位点
R5	M0(one-ratio)	-7803.46	$\omega=0.14$	602.52 $P=0.0000$	无
	M3(discrete)	-7502.20	$p_0=0.7218, p_1=0.2236 (p_2=0.0546), \omega_0=0.02, \omega_1=0.27, \omega_2=1.46$		未显示
	M1a(Nearly neutral)	-7534.98	$p_0=0.8974 (p_1=0.1026)$	22.66 $P=0.0000$	不允许
	M2a(Positive selection)	-7523.65	$p_0=0.8971, p_1=0.0897 (p_2=0.0132), \omega_2=3.05$		<b>162S 376T</b>
	M7(beta)	-7525.67	$p=0.1371, q=0.7393$	49.5 $P=0.0000$	不允许
	M8(beta& $\omega$ )	-7500.92	$p_0=0.9604 (p_1=0.0396), p=0.2362, q=2.1903, \omega=1.75$		<b>162S 211R 245V 297E 332Q 360A 376T 386T</b>
	M0(one-ratio)	-6749.04	$\omega=0.17$	488.7 $P=0.0000$	无
	M3(discrete)	-6504.69	$p_0=0.6640, p_1=0.2606 (p_2=0.0753), \omega_0=0.01, \omega_1=0.23, \omega_2=1.51$		未显示
X4	M1a(Nearly neutral)	-6522.31	$p_0=0.8821 (p_1=0.1179)$	18.96 $P=0.0001$	不允许
	M2a(Positive selection)	-6512.83	$p_0=0.8834, p_1=0.0940 (p_2=0.0226), \omega_2=2.65$		<b>211R</b>
	M7(beta)	-6526.67	$p=0.1286, q=0.6409$	46.92 $P=0.0000$	不允许
	M8(beta& $\omega$ )	-6503.21	$p_0=0.9479 (p_1=0.0521), p=0.2461, q=2.1554, \omega_2=1.86$		<b>207Q 211R 215T 245V 357M 376T</b>

a) 后验概率大于>95%的正选择位点; 后验概率>99%的正选择位点用黑体显示

附表 5 HIV-1 C 亚型 R5 病毒 RT 基因的似然值和参数估计<sup>a)</sup>

模型	lnL	参数估计	2Δl	正选择位点
M0(one-ratio)	-7159.79	$\omega=0.15$	573.44	无
M3(discrete)	-6873.07	$p_0=0.8749, p_1=0.1176 (p_2=0.0075), \omega_0=0.05,$ $\omega_1=0.90, \omega_2=5.24$	$P=0.0000$	<b>123D 334Q 377T</b>
M1a(Nearly neutral)	-6900.15	$p_0=0.8841 (p_1=0.1159)$	53.04	不允许
M2a(Positive selection)	-6873.63	$p_0=0.8817, p_1=0.1109 (p_2=0.0073), \omega_2=5.41$	$P=0.0000$	<b>123D 334Q 377T</b>
M7(beta)	-6909.81	$p=0.1378, q=0.6736$		不允许
M8(beta& $\omega$ )	-6877.18	$p_0=0.9919 (p_1=0.0081), p=0.1647, q=0.9165,$ $\omega=4.91$	$65.26 P=0.0000$	<b>123D 334Q 377T</b>

a) 后验概率大于>95%的正选择位点; 后验概率>99%的正选择位点用黑体显示

附表 6 HIV-1 B 亚型 R5 和 X4 病毒 IN 基因的似然值和参数估计<sup>a)</sup>

细胞嗜性	模型	lnL	参数估计	2Δl	正选择位点
R5	M0(one-ratio)	-4567.46	$\omega=0.12$	304.34	无
	M3(discrete)	-4415.29	$p_0=0.82278, p_1=0.1239 (p_2=0.0533),$ $\omega_0=0.03, \omega_1=0.34, \omega_2=1.62$	$P=0.0000$	<b>10E 16S 27L 38S 44L 71L 100L</b> 111T <b>121T</b> 124T 199I
	M1a(Nearly neutral)	-4427.40	$p_0=0.8994 (p_1=0.1006)$	6.9	不允许
	M2a(Positive selection)	-4423.95	$p_0=0.9011, p_1=0.0816 (p_2=0.0173), \omega_2=2.64$	$P=0.0317$	无
	M7(beta)	-4435.26	$p=0.1511, q=0.8502$	37.46	不允许
	M8(beta& $\omega$ )	-4416.53	$p_0=0.9444 (p_1=0.0556), p=0.3879, q=4.7314,$ $\omega=1.58$	$P=0.0000$	16S 71I 100L 124T
X4	M0(one-ratio)	-3994.45	$\omega=0.14$	198.36	无
	M3(discrete)	-3895.27	$p_0=0.7992, p_1=0.1833 (p_2=0.0175), \omega_0=0.04,$ $\omega_1=0.51, \omega_2=3.19$	$P=0.0000$	100L 118S 123A 124T
	M1a(Nearly neutral)	-3909.48	$p_0=0.8738 (p_1=0.1262)$	12.02	不允许
	M2a(Positive selection)	-3903.47	$p_0=0.8744, p_1=0.1123 (p_2=0.0133), \omega_2=3.95$	$P=0.0025$	118S 123A
	M7(beta)	-3910.05	$p=0.1933, q=0.9867$	26.96	不允许
	M8(beta& $\omega$ )	-3896.57	$p_0=0.9821 (p_1=0.0179), p=0.2779, q=1.8653,$ $\omega=3.13$	$P=0.0000$	100L <b>118S 123A</b> 124T

a) 后验概率大于>95%的正选择位点; 后验概率>99%的正选择位点用黑体显示

附表 7 HIV-1 C 亚型 R5 病毒 IN 基因的似然值和参数估计<sup>a)</sup>

模型	lnL	参数估计	2Δl	正选择位点
M0(one-ratio)	-4143.50	$\omega=0.16$	292.76	无
M3(discrete)	-3997.12	$p_0=0.7776, p_1=0.1910 (p_2=0.0314), \omega_0=0.04,$ $\omega_1=0.44, \omega_2=2.70$	$P=0.0000$	<b>11E 50M 72V 125T</b> 269R
M1a(Nearly neutral)	-4016.89	$p_0=0.8952 (p_1=0.1048)$	26.58	不允许
M2a(Positive selection)	-4003.60	$p_0=0.8940, p_1=0.0851 (p_2=0.0209), \omega_2=3.32$	$P=0.0000$	<b>50M 72V 125T</b>
M7(beta)	-4021.05	$p=0.1546, q=0.7387$		不允许
M8(beta& $\omega$ )	-3996.88	$p_0=0.9696 (p_1=0.0304), p=0.2734, q=1.9860,$ $\omega=2.74$	$48.34 P=0.0000$	<b>11E 50M 72V 125T</b>

a) 后验概率大于>95%的正选择位点; 后验概率>99%的正选择位点用黑体显示