Sequencing Primers for HIV-1

Eric Sanders-Buell, Mika O. Salminen, and Francine E. McCutchan

The Henry M. Jackson Foundation Research Laboratory and Division of Retrovirology, Walter Reed Army Institute of Research, Rockville Md. 20850

Introduction

The highly variable genome of the Human Immunodeficiency Virus Type-1 (HIV-1) has been extensively studied for more than a decade. Over several years, multiple, distinct genotypes of HIV-1 were recognized [1]. This finding sparked a multi-national effort to document the full genetic diversity of the virus and to unlock the biological and immunological significance of its genotypic complexity; references 2–7 provide examples of an extensive literature. Significant new discoveries, each incrementally expanding the range of genotypic variation ascribed to HIV-1, have occurred even in the last two years. In addition to at least eight distinct genotypes in the M, or "main" HIV-1 group, a number of rare "O" or outlier forms, and intergenotypic recombinants, combining genetic information from two or more of the major genotypes, have come to light [8–12]. Investigation of the HIV-1 genome is intensifying.

An early trend towards limited sampling of the genetic information contained in the 10 KB genome of HIV-1, fostered in part by the widespread use of the polymerase chain reaction (PCR) to recover HIV-1 genome from clinical samples and virus cultures, resulted in the accumulation of a wealth of short sequences from selected genomic regions and from those genotypes that were first recognized and most accessible for study. Complete genomic sequence information remains scarce or unavailable for many of the genotypes that comprise the prevalent "M" group viruses and, also, for most of the group O and recombinant viruses. Counterbalancing efforts have begun. Increasingly, investigators are focusing on more complete genomic information and on the relatively less characterized genotypes. Using enriched sources or nested PCR, full-length *gag* or envelope genes are accessible by conventional PCR amplification. With the new awareness that recombinant forms occur at a significant frequency [13], the impetus for more complete genetic analysis is strengthened; the technical capability to extend PCR amplification to encompass the full HIV-1 genome has appeared concomitantly [14]. Longer sequence segments, if not complete genomes, may soon contribute substantially to the HIV-1 genotyping effort.

Sequencing and Viral Variation

The genetic variation of HIV-1 poses a challenge to the DNA sequencing laboratory by diminishing the success rate of the primer walking strategy that is frequently employed for long sequence segments. A series of oligonucleotide primers, spaced at intervals along both strands of the DNA template, is used to initiate dideoxy nucleotide sequencing reactions. Genetic variation in the sequences complementary to the primers can interfere, leading to foreshortened sequence segments or failed sequencing reactions.

It is unlikely that universally applicable primer sets can be developed for the sequencing of any but the most conserved regions of the HIV-1 genome. However, the combined experience of many laboratories, if compiled and periodically updated, could facilitate the gathering of sequence data. Here we report our cumulative experience regarding the performance of more than 150 oligonucleotide primers in the *gag* and *env* genes of international HIV-1 isolates and we provide the nucleotide sequences of recommended primer panels. Further, we propose the establishment of a primer database for HIV-1 genomic sequencing that combines the primer walking strategies employed by individual laboratories into an internet accessible, global resource.

Evaluation of Primer Performance

Over the course of a five year HIV-1 sequencing effort, we have evaluated, at some level, 126 primers in the envelope gene and 36 primers in the *gag* gene. All sequencing reactions were performed

on cloned DNA templates using fluorescent dye-labeled dideoxy terminators (Applied Biosystems, Foster City, CA) and cycle sequencing as described by the manufacturer. Data was collected using an Applied Biosystems 373A DNA sequencer. The isolates that were included in this project are listed in Table 1. They include representatives of genotypes A,B,C,D,F, and G for the *gag* gene and genotypes A,B,C,D,E, and F for the envelope gene, respectively. A retrospective compilation of the relative utility of sequencing primers on each of these genotypes is presented here.

Primers were not systematically evaluated. Rather, those that initiated sequencing reactions most often and provided the longest readable sequence segments came into common use, while those that performed sporadically were usually dropped. We have evaluated primer performance in two stages. First, primers that performed poorly and/or were infrequently employed were removed from the recommended set. The remaining were then evaluated for their rate of priming on each genotype and for the length of sequence obtained.

The 1.5 KB, relatively conserved, *gag* gene proved far less problematic than the more highly variable, 2.5 KB envelope gene. Of 36 *gag* primers, 22 came into common use; six of these performed optimally on six genotypes, and ten others were useful for at least four of six genotypes. Only 49 of 126 primers for the envelope gene became part of the core reagent set. Almost half of these were only applicable to 1, 2, or 3 of the six genotypes under study. Clearly, envelope primers, and, to a lesser extent, *gag* primers, require selection for the genotypes under study whenever possible.

Primers for *gag* and Envelope Sequencing

Primers for the collection of sequence data from the coding and non-coding strands of the DNA template were assigned odd or even numbers, respectively, preceded by "G" for *gag* or "E" for envelope. Numbering was sequential from 5′ to 3′ according to the location of priming.

Recommended primers for sequencing of the HIV-1 *gag* gene are listed in Table 2, and their observed performance on various genotypes is illustrated in Figure 1. Primers that provided the highest performance are shown by thick-lined arrows, while those that performed adequately are thin-lined arrows. Primers that did not meet our performance standard or that were not sufficiently evaluated are omitted from pertinent sections of the diagram. Table 3 and Figure 2 illustrate the performance of sequencing primers in the envelope gene region similarly. The unavailability of multiple isolates of some genotypes precluded full evaluation of primer performance at this time.

A Sequencing Primer Database

Ideally, the nucleotide sequences of recommended sequencing primers should be widely accessible and organized by HIV-1 genotype, by genomic region, by the strand of DNA from which sequencing is initiated, and by their nucleotide sequence. Accompanying annotation could include the HIV-1 isolates on which primers were employed, genbank accession numbers of the sequence data that was collected, the specific methodology used for DNA sequencing, the source laboratory, and a bibliography of pertinent publications. Lastly, such information should be available through electronic media.

Towards this goal, we have organized the data available from our laboratory as reported here. It should be re-emphasized that we make recommendations based only on our own experience and not on a complete or systematic study of primer performance, which may vary under different circumstances. Work in progress should provide a similar compilation of primers that were successful for sequencing of additional regions of the HIV-1 genome, including the *pol*, *nef*, *tat*, *rev*, *vif*, *vpu*, *vpr*, and LTR of multiple genotypes.

This compendium, used extensively by participants in the global HIV-1 genotyping effort and having already established itself as a site accessible through the World Wide Web (WWW), would seem optimally positioned to serve as a future repository for the nucleotide sequences of HIV-1 sequencing primers.

Table 1. International HIV-1 Isolates

	GAG[1]							ENV[2]					
Genotype	A		B	C	D	F	G	A	B	C	D	E	F
	VI310	VI57	WR27	VI313	G109	BZ162	LBV21-7	DJ258	WR27	DJ285	SE365	CM235	BZ126
	LBV2-3	VI32	NL4-3	SM145	UG270	VI174		DJ263	NL4-3	DJ373	UG269	CM239	BZ163
	LBV23-10	VI415	BK132	UG268	K31	VI69		DJ264	US1	SE364	UG274	CM240	
Isolates	DJ258	K29	UG280	DJ259	UG274			DJ273	US2	SM145		CM242	
	CI51	K112			VI203				US3	UG268		CM243	
	CI59	K89			SG365				US4	ZAM18		KH03	
	CI20	K7										KH08	
	CI4	VI59											

1. Gag gene sequences have been reported [2] and are available through genbank under accession #s L03696, L03702, L03705, L03707, and L11751-L11803, inclusive. Isolates NL4-3 and WR27 are available under genbank accession #s M19921 and U26546, respectively. Sequences of isolates CM240 and CM245 are unpublished data from our laboratory.

2. Envelope glycoprotein gene sequences have been reported [4] and are available through genbank under accession #s L22939 to L22957, inclusive, and L23064, and L23065. Footnote 1 provides accession numbers for isolates NL4-3 and WR27. Sequences of isolates DJ273, DJ258, KH03, and KH08 are unpublished data from our laboratory.

Table 2. Sequencing Primers for the HIV-1 Gag Gene

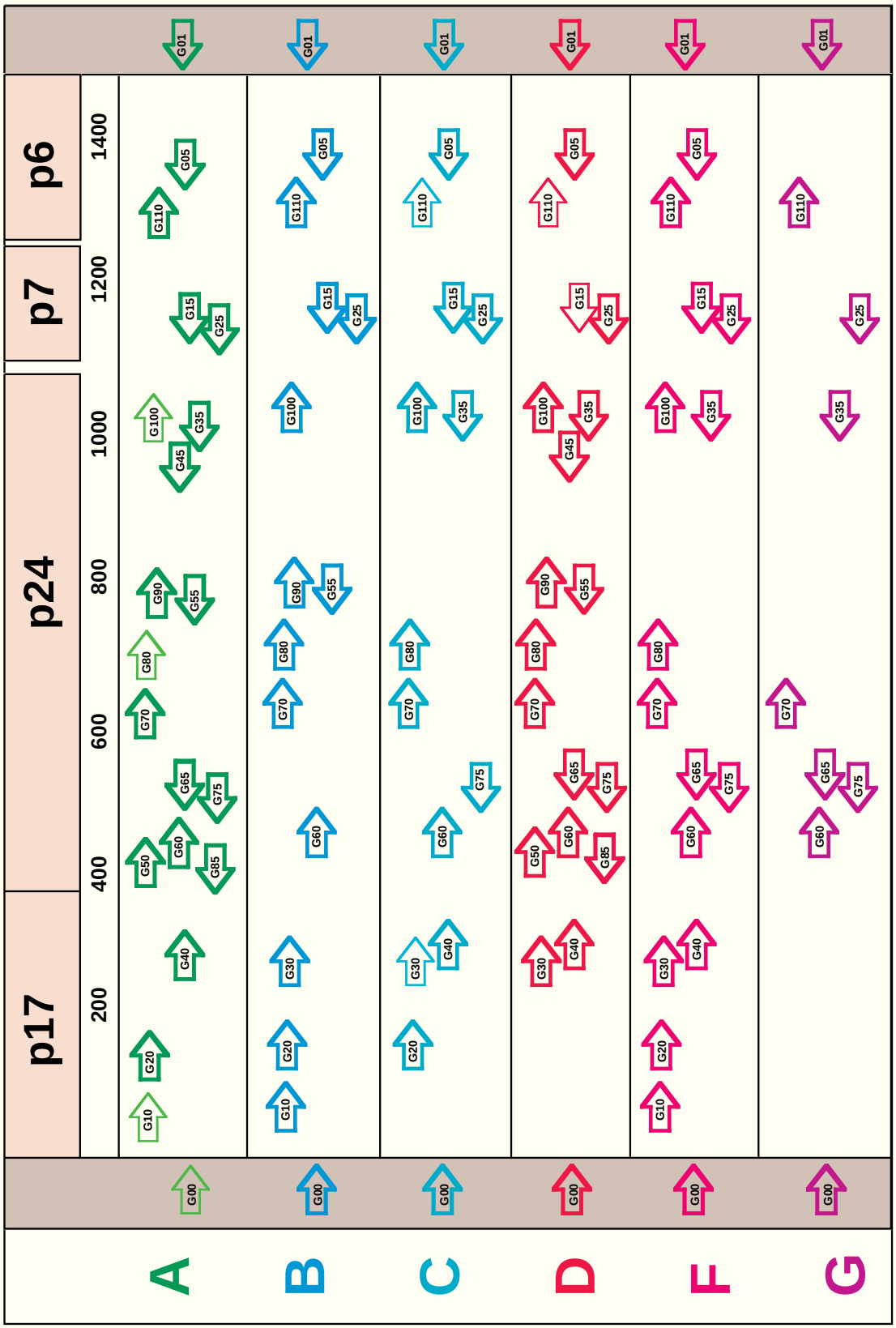
Primer	Location[1]	Sequence (5′–3′)	Primer	Location[1]	Sequence (5′–3′)
G 00	-26 >>	GACTAGCGGAGGCTAGAAG	G 01	<< 1493	AGGGGTCGTTGCCAAAGA
G 10	17 >>	CAGTATTAAGCGGGGGAGAATT	G 05	<< 1368	TGTTGGCTCTGGTCTGCTCT
G 20	103 >>	GTATGGGCAAGCAGGGAGCTAGAA	G 15	<< 1192	CTTTGCCACAATTGAAACACTT
G 30	242 >>	CAGTAGCAACCCTCTATTGTGT	G 25	<< 1100	ATTGCTTCAGCCAAAACTCTTGC
G 40	286 >>	GACACCAAGGAAGCTTTAGA	G 35	<< 1049	CATGCTGTCATCATTTCTTCTA
G 50	344 >>	CACAGCAAGCAGCAGCTG	G 45	<< 972	TTGGACCAACAAGGTTTCTGTC
G 60	384 >>	CAGCCAAAATTACCCTATAGTGCAG	G 55	<< 782	ATTTCTCCCACTGGGATAGGTGG
G 70	617 >>	ATGAGGAAGCTGCAGAATGGG	G 65	<< 523	ATGCTGAAAACATGGGTA
G 80	682 >>	ATGAGAGAACCAAGGGGAAGTGA	G 75	<< 481	CTTCTATTACTTTTACCCATGC
G 90	755 >>	ATAATCCACCTATCCCAGTAGGAGAAAT	G 85	<< 407	TGCACTATAGGGTAATTTTG
G100	1028 >>	TAGAAGAAATGATGACAG			
G110	1289 >>	AGGCTAATTTTTTAGGGA			

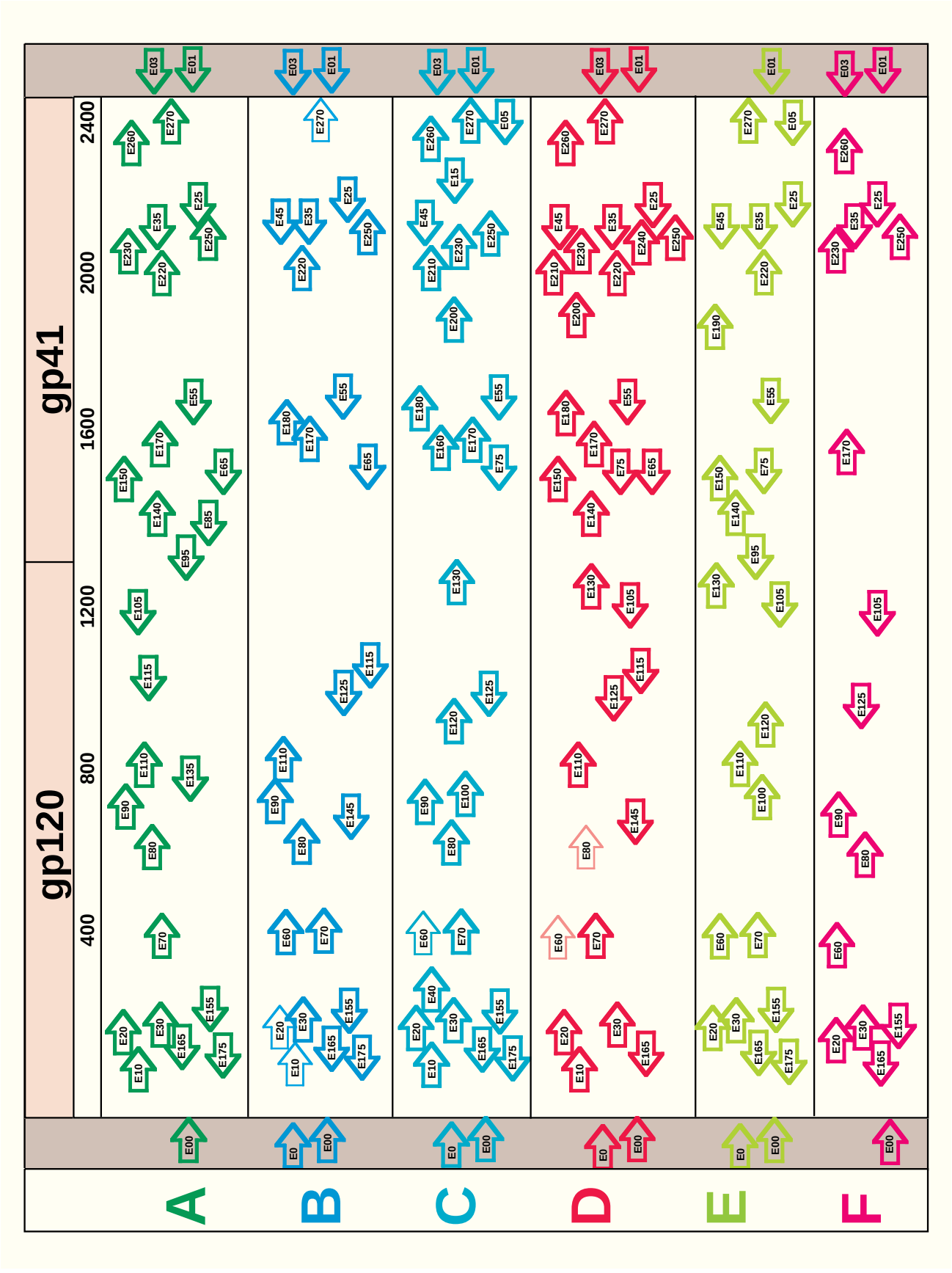
1 Nucleotide positions 1–1503 of the gag gene of reference isolate HXB2 (Genbank accession # K03455) are used to indicate the 5′ nucleotide of each primer. Positions upstream of the gene are indicated by "-".

Table 3. Sequencing Primers for the HIV-1 Env Gene

Primer	Location[1]	Sequence (5′–3′)	Primer	Location[1]	Sequence (5′–3′)
E0	-372 >>	TAGAGCCCTGGAAGCATCCAGGAAGTCAGCCTA	E01	<< +269	TCCAGTCCCCCCTTTTCTTTTAAAAA
E00	-24 >>	TAGAAAGAGCAGAAGACAGTGGCAATGA	E03	<< +218	TAAGTCATTGGTCTTAAAGGTACCTG
E10	100 >>	TTGTGGGTCACAGTCTATTATGGGGT	E05	<< 2363	TATTTGAGGGCTTCCCACCCCC
E20	206 >>	GGGCCACACATGCCTGTGTACCCACAG	E15	<< 2200	CTCTCTCTCCACCTTCTTCTTC
E30	221 >>	GTGTACCCACAGACCCCAGCCCACAAG	E25	<< 2122	GGTGAGTATCCCTGCCTAAC
E40	282 >>	CATGTGGAAAAATGACATGGTGGATCA	E35	<< 2116	GGTGAGTATCCCTGCCTAACTCTATT
E50	297 >>	CATGGTAGAGCAGATGCAGGAGGATG	E45	<< 2113	CCTGCCTAACTCTATTCAC
E60	323 >>	TAATCAGTTTATGGGATCAAAGC	E55	<< 1688	GCCCCAGACTGTGAGTTGCAACAGATG
E70	335 >>	GGGATCAAAGCCTAAAGCCATGTGTAA	E65	<< 1568	AGTGCTTCCTGCTGCTCC
E80	634 >>	CCAATTCCCATACATTATTGTG	E75	<< 1567	GCGCCCATAGTGCTTCCTGCTGCTCCC
E90	729 >>	CACAGTACAATGTACACATGGAAT	E85	<< 1404	GTCCCTCATATCTCCTCCTCCAGGTCT
E100	742 >>	ACACATGGAATTAAGCCAGT	E95	<< 1299	GATGGGAGGGGCATACAT
E110	778 >>	CTGTTAAATGGCAGTCTAGCAGAA	E105	<< 1277	GCTTTTCCTACTTCCTGCCAC
E120	874 >>	GTAGAAATTAATTGTACAAGACCC	E115	<< 1127	AGAAAAATTCCCCTCCACAATTAA
E130	1262 >>	ACAAATTATAAACATGTGGCAGG	E125	<< 1091	CAATTTCTGGGTCCCCTCCTGAGG
E140	1441 >>	GTGAATTATATAAATATAAAGTAG	E135	<< 836	AGCTGTACTATTATGGTTTTAGCATTGT
E150	1494 >>	CCAGGGCAAAGAGAAGAGTGGTG	E145	<< 757	CAGCAGTTGAGTTGATACTACTGG
E160	1535 >>	GTGGGAATAGGAGCTGTGTTCCTTGGG	E155	<< 281	CTGTTCTACCATGTTATTTTTCCACATGT
E170	1573 >>	AGCAGGAAGCACTATGGG	E165	<< 211	GGGGTCTGTGGGTACACAGGCATGTGT
E180	1633 >>	GTCTGGTATAGTGCAACAGCA	E175	<< 154	TTTAGCATCTGATGCACAAAATAG
E190	1824 >>	CCTGGAACTCCACTTGGAG			
E200	1850 >>	GGGATAACATGACCTGGATGCAGTGGG			
E210	2024 >>	TAACAAATTGGCTGTGGTATATAA			
E220	2024 >>	TATCAAAATGGCTGTGGTATATAA			
E230	2049 >>	AATATTCATAATGATAGTAGGAGG			
E240	2056 >>	ATAATGATAGTAGGAGGCTTGATAGGC			
E250	2068 >>	GGAGGCTTGATAGGTTTAAGAATA			
E260	2296 >>	TTCAGCTACCACCGCTTGAGAGACT			
E270	2344 >>	GTGGAACTTCTGGGACGCAG			

1 Nucleotide positions 1–2571 of the envelope glycoprotein gene of reference isolate HXB2 (Genbank accession # K03455) are used to indicate the 5′ nucleotide of each primer. Positions upstream and downstream of the gene are indicated by "-" and "+", respectively.


p17
p24
p7
p6
200
400
600
800
1000
1200
1400
A
B
C
D
F
G
G00
G01
G05
G10
G110
G15
G20
G25
G30
G35
G40
G45
G50
G55
G60
G65
G70
G75
G80
G85
G90
G100


gp120
gp41
400
800
1200
1600
2000
2400
A
B
C
D
E
F

References

[1] Myers G, Korber B, Wain-Hobson S, Jeang KT, Henderson LE, and Pavlakis GN. Human Retroviruses and AIDS, Los Alamos National Laboratory, Los Alamos, NM, 1994.

[2] Louwagie J, McCutchan F, Peeters M, Brennan TP, Sanders-Buell E, Eddy G, van der Groen G, Fransen K, Gershy-Damet GM, Deleys R, Burke DS. Phylogenetic Analysis of *gag* Genes from 70 International HIV-1 Isolates Provides Evidence for Multiple Genotypes. AIDS 7:769–780, 1993.

[3] WHO Network for HIV Isolation and Characterization. HIV Type 1 Variation in World Health Organization-Sponsored Vaccine Evaluation Sites: Genetic Screening, Sequence Analysis, and Preliminary Biological Characterization of Selected Viral Strains. AIDS Res. and Human Retroviruses 10:1327–1433, 1994.

[4] Louwagie J, Janssens W, Mascola J, Heyndricks L, Hegerich P, van der Groen G, McCutchan FE, Burke DS. Genetic Diversity of the Envelope Glycoprotein from Human Immunodeficiency Virus Type-1 (HIV-1) Isolates of African Origin. J. Virol. 69:263–271, 1995.

[5] Pau CP, Kai M, Holloman-Candal DL, Luo CC, Kalish ML, Schochetman G, Byers B, George JR, and the WHO Network for HIV Isolation and Characterization. Antigenic Variation and Serotyping of HIV Type 1 from Four World Health Organization-Sponsored HIV Vaccine Sites, AIDS Res. and Human Retroviruses 10:1369–1377, 1994.

[6] Mascola JR, Louwagie J, McCutchan FE, Fischer CL, Hegerich PA, Wagner KF, Fowler AK, McNeil JG, Burke DS. Two Antigenically Distinct Subtypes of HIV-1; Viral Genotype Predicts Neutralization Immunotype. J. Inf. Dis. 169:48–54, 1994.

[7] Moore JP, McCutchan FE, Poon S, Mascola J, Liu J, Cao Y, Ho DD. An Exploration of Antigenic Variation in gp120 from Clades A–F of Human Immunodeficiency Virus Type 1 Using Monoclonal Antibodies. J. Virol. 68:8350–8364, 1994.

[8] Nkengason JN, Peeters M, vanden Haesevelde M, et.al. Antigenic Evidence of the Presence of the Aberrant HIV-1ANT70 Virus in Cameroon and Gabon AIDS 7:1536–1538, 1993.

[9] Gurtler L, Eberle J, can Brunn A, et.al. A New Subtype of Human Immunodeficiency Virus Type 1 (MVP5180) from Cameroon. J. Virol. 68:1581–1585, 1994.

[10] Vanden Haesevelde M. DeCourt JL, de Leys RJ, at. al. Genomic Cloning and Complete Sequence Analysis of a Highly Divergent African Human Immunodeficiency Virus Isolate. J. Virol. 68:1586–1596, 1994.

[11] Hahn BH, Robertson DL, McCutchan FE, Sharp PM. Recombination and Diversity of HIV: Implications for Vaccine Development. Neuvième Colloque des Cent Gardes, Paris, 1994.

[12] Sharp P, Robertson D, McCutchan F, Hahn B. Recombination in HIV-1. Nature 374:124–126, 1995.

[13] Salminen MO, Carr JK, McCutchan FE. Genotyping of HIV-1. in: Human Retroviruses and AIDS, (Myers G, Korber B, et.al., eds.) Los Alamos National Laboratory, Los Alamos, NM, 1995 edition.

[14] Salminen MO, Koch K, Sanders-Buell E, Ehrenberg PK, Michael NK, Carr JK, Burke DS, McCutchan FE. Recovery of Full-Length HIV-1 Provirus of Diverse Subtypes from Primary Virus Cultures Using Polymerase Chain Reaction. Virology: in press, 1995.