ISSN : 0974 - 7532 Volume 6 Issue 2

Trade Science Inc.

Research & Reviews in

BioSciences

Regular Paper

RRBS, 6(2), 2012 [44-58]

LTR regulatory sequences and primate lentiviruses evolution

M.L.J.Moncany[1*], K.Dalet[1], P.R.R.Courtois[1], J.S.Le Brizaut[2]

[1]Laboratoire de Biologie Cellulaire et Moléculaire, UFR Sciences, Université de La Rochelle, Avenue Marillac, 17042 - La Rochelle Cedex 1, (FRANCE).

[2]Département Informatique et Mathématique, École Centrale de Nantes, BP9210, 44321-Nantes, (FRANCE).

E-mail : maurice.moncany@free.fr

Received: 14th December, 2011 ; Accepted: 16th January, 2012

ABSTRACT

Lentiviruses are characterised by high genomic flexibility due to multiple recombinations. The genomes are composed of many parts from different strains and show an alternation of long variable and short conserved domains which can evolve differently and independently. For primate lentiviruses conserved domains correspond to LTR-promotor regulatory sequences and it is of particular interest to understand how important genetic elements such as lentiviral-derived LTRs evolve. The analysis of 382 LTRs extracted from 298 primate lentivirus genomes including HIV-1, HIV-2, Macaque, African Green Monkey, Sooty Mangabey, Chimpanzee, Mandrill and Syke's viruses is reported in this paper. We compared the U3 promoter sites responsible for cellular-related regulatory proteins binding from upstream USF/Ets1 sequences to the TATA-box. The present investigation revealed a clustering of the different regulatory sequences that clearly appeared as belonging to distinct HIV-1 or HIV-2 like-groups. The nucleotidic sequences from simian viruses showed characteristics that were shared by the two generic groups with occasionally slight modifications peculiar to the concerned viral species. This situation allowed a better subtyping distinction of the mosaic simian lentiviruses. More striking was the occurrence of modified arrangements of these conserved sequences revealing either a total or partial HIV-1 or HIV-2 organization. Some sequences exhibited a dual HIV-1/HIV-2 organization, which was representative of Chimpanzee and African Green Monkey viruses. Thus the question arises about ancestral lentiviral sequences from which the recombinations produced new emerging viruses to form evolution groups. Then we suggest a model in which the AGM genomes represent a key crossroad in the herd evolution of primate lentiviruses.

KEYWORDS

LTR;
HIVs;
SIVs;
Primate lentiviruses;
Regulatory sequences;
Recombination;
Evolution crossroad.

INTRODUCTION

The lentiviral genome is under continuous recombination leading to the now generally admitted notions of "mosaic virus" and "quasi species" (for a review see Krebs et al[1]), which integrate the non-uniform variation occurring along the genome between different genes and even parts of a single gene (*e.g.*[2]). These viruses

can adapt to changes in the environment to allow continued reproduction and ensure infection spread. The conserved domains of the lentiviral genomes are essential for viral expression and virulence. Their preservation is important for the group evolution through a permanent genomic reshaping. Numerous investigations on lentivirus evolution based on the comparison of a single or a few genes, and parts of genes raised the difficulty to establish clear evolution trees that vary among the different studies (*e.g.*[3-6]).

A reliable attempt was possible when using detailed comparisons between key sequences which should be conserved through evolution or modified in a conservative way, *i.e.* in an analogous approach such as defined by Bonhoeffer et al[7]. These modifications should be coherent in the studied viral sub-group and should form families characteristic of the proper group. A first approach was to identify Conserved Lentiviral Sequences (CLS) as landmarks of genomic flexibility which evidenced a viral sub-typing[8]. From that study, it appeared that the simian and other mammal lentiviruses could be related either to HIV-1 or HIV-2 group and it was of interest to examine such a distribution through other well conserved domains.

The loss of phylogenetic information observed in recombinating sequences[9, 10] led us to the subsequent step which was to check the stable sites binding the cellular-related regulatory factors in the LTR promoter U3 domain and to compare the mapped organization of the primate lentivirus LTRs. This paper presents the computer analysis for the promoter regulatory sequences of 382 LTRs belonging to 298 primate lentiviruses followed by the studies of their organization. We reconstituted the U3 part from the TATA-box to the upstream USF/Ets1 sequences (*i.e.* the core and modulator elements). For this purpose 223 HIV-1, 38 HIV-2 and 121 simian LTRs were analyzed, either single or extracted from complete genomes. Such detailed comparisons allowed to determine modifications which led to clearly distinguish between the HIV-1 and HIV-2 type sequences. A global survey showed an increasing divergence and complexity of the structures from HIV-1 to HIV-2 up to simian ones. More than the confirmation of the viral affiliation previously reported for CLSs[8], this permitted a subtyping to avoid the confusion induced by the lentiviral genomic mosaic structure. The simian LTRs generally presented either a HIV-1 or a HIV-2 belonging with some unexpected dual HIV-1/HIV-2 one mainly for African Green Monkey (AGM) viruses. The combined studies of nucleotidic lexical variations and of U3 general organization led to consider the results obtained as identifiers of the viral strains. These characteristics may point to either an incomplete convergence phenomenon or a potential herd evolution scheme when considering common ancestral sequences. Such data led us to consider the AGM viruses as an early landmark in the suggested evolution model.

MATERIALS AND METHODS

Complete genomes and single LTRs

The referenced lentiviral complete genomes and single LTRs were extracted from Genbank database. The list of analyzed sequences is reported in "Supplemental data" available on line. These sequences represent 223 LTRs belonging to 176 HIV-1 strains (noticed 223/176, comprising 10 HIV-1-O and 213 HIV-1 other strains), 38/27 for HIV-2s and 47/39, 34/27, 19/11, 12/10, 6/6 and 3/2 for Macaque, African Green Monkey (AGM), Sooty Mangabey, Chimpanzee (CPZ), Mandrill and Syke's viruses, respectively.

Methodology

Several published papers presented different schemes for the HIV-1 LTR organization (e.g.[1, 11-14]). In the present work we chose for HIV-1s the sequences described in the review by Jones and Peterlin[12] as many of the regulatory domains reported in other papers corresponded to split parts of these sequences. For identical reasons and to allow a harmonization between all regulatory lentiviral sequences, we selected for HIV-2s those described by Guyader et al[15]. The reference sequences checked for detection in human and simian viruses were:

- *HIV-1*: TATA-box: TATAA, SP1/1: GGGGAGTGGC, SP1/2: TGGGCGGGACT, SP1/3: GGAGGCGTGGC, NF-κB1: GGGGACTTTCC, NF-κB2: GGGACTTTCCG, LEF: TTCAAGAACTG, Ets1: GCATCCGG, USF: CACATG;

- *HIV-2*: TATA-box: TATAAA, SP1/1: GGTGGGGAAC, SP1/2: TGGGAGGAGC, SP1/3: AAGGGAGGGAC, NF-κB1: GGGGCTGTAAC,

Regular Paper

NF-κB2: GGGACTTTCCA, LEF: GGAA GTAGCTA, Ets1/1: GTACCCAG, Ets1/2: ACATCCAG, USF: CACAAG.

The crude viral sequences were extracted from Genbank and the Los Alamos Laboratory data bank and corresponded either to LTR domains or complete genomes from which only LTRs were considered. They were listed and analyzed following a specific program based on an algorithm that we designed, which permitted an automatic research of the desired sequence[2, 8, 16]. The algorithm, used to match the sequences, was appropriate to find reference-related sequences, up to 30-50 bases. Some of the latter corresponded to regulatory sites also characterized by their respective positions on the viral genome. This method is not a usual alignment process but a research of sequences bearing particular properties. Thus, the studies considered both the nucleotide sequence (lexical determination) and the relative position of the site in the LTR. Some of the detected sequences had a variability superior to 30%, which is thought to delimit generally admitted jump from one family to another one and might fit with a possible biological significant divergence[8]. For comparison, another step studied sequences with a variability inferior or equal to 30% to the reference sequences as discussed in the first chapter of "Results-Discussion". The regulatory sequences were distributed under the HIV-1 or the HIV-2 type, those presenting the same number of transitions to be part of the HIV-1 or HIV-2 types were named "dual sequences".

RESULTS - DISCUSSION

The analysis of the LTR U3 part of primate lentiviruses considered the number and the nature of the conserved regulatory nucleotidic sequences (Figure 1, supplementary data available on line; TABLES 1 and 2) and their mapping organization (Figures 2 to 7). The type of association between the regulatory sequences was investigated in order to define either HIV-1 (yellow boxes) or HIV-2 (blue boxes) as identifier types. Sequences which presented the same number of nucleotide transitions to correspond to HIV-1

TABLE 1 : LTR regulatory sequences of human and simian genomes. The first line indicates, from right to left in order to reproduce the organization of the LTR, the name of the studied sequences from the TATA box upstream to the USF domain. The yellow, blue, green and white areas represent HIV-1, HIV-2, dual HIV-1/HIV-2 and non-sequences, respectively. These areas are proportional to the percentage of the respective detections. The HIV-1* quotation indicates the viral strains which do not belong to the HIV-1-O type.

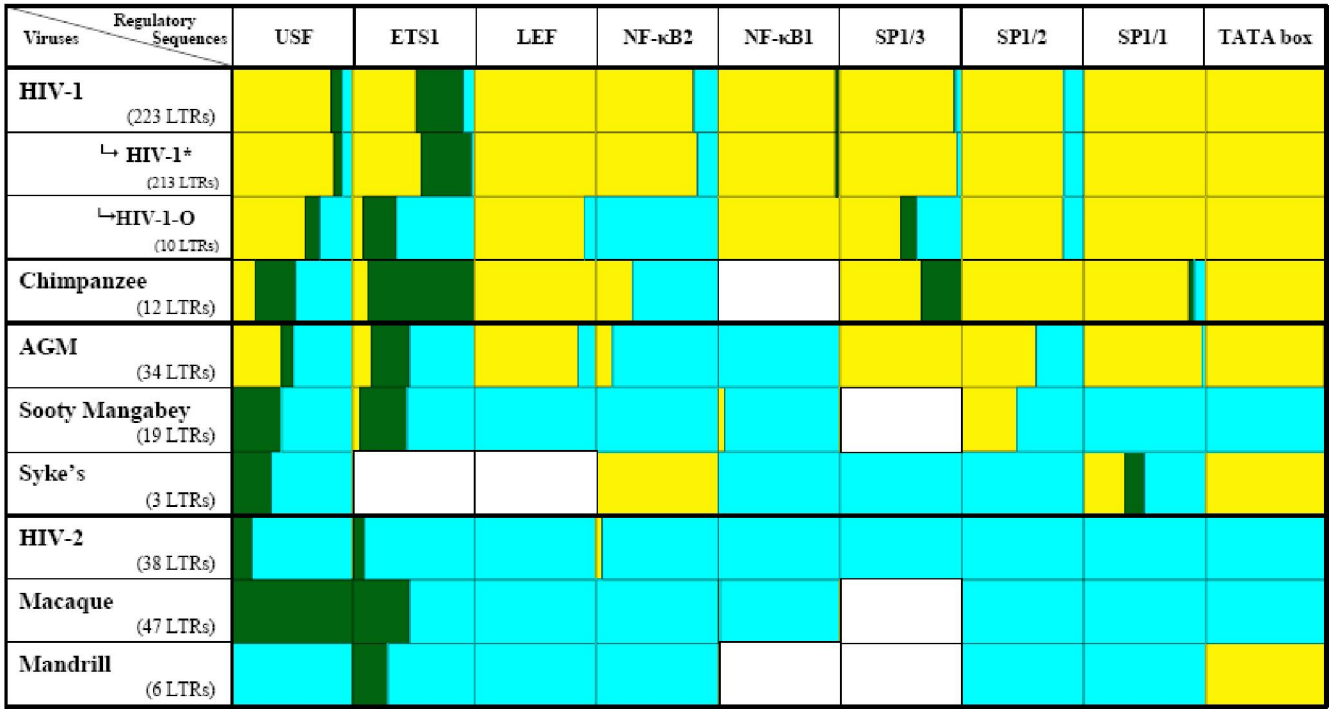

TABLE 2 : Number and percentage of LTR regulatory sequences presenting variations ≤ 30%. For each sequence, the three columns indicated the number of sequences presenting variations ≤ 30%, the total number of studied sequences and the percentage of sequences presenting variations ≤ 30%, respectively. The colors are as for Table 1. The HIV-1* quotation indicates the viral strains which do not belong to the HIV-1-O type.

	USF			Ets1			LEF			NF-kB/2			NF-kB/1			SP1/3			SP1/2			SP1/1			TATA box		
	ab seq (≤30% transitions)	total ab seq	% seq (≤ 30% transitions)	ab seq (≤30% transitions)	total ab seq	% seq (≤ 30% transitions)	ab seq (≤30% transitions)	total ab seq	% seq (≤ 30% transitions)	ab seq (≤30% transitions)	total ab seq	% seq (≤ 30% transitions)	ab seq (≤30% transitions)	total ab seq	% seq (≤ 30% transitions)	ab seq (≤30% transitions)	total ab seq	% seq (≤ 30% transitions)	ab seq (≤30% transitions)	total ab seq	% seq (≤ 30% transitions)	ab seq (≤30% transitions)	total ab seq	% seq (≤ 30% transitions)	ab seq (≤30% transitions)	total ab seq	% seq (≤ 30% transitions)
HIV-1	165	194	85	119	121	98	185	221	84	191	202	95	222	229	97	203	212	96	183	190	96	217	218	100	223	223	100
	18	20	90	19	21	90	2	2	100	52	52	100	0	2	0	9	12	75	35	36	97						
	4	25	16	91	92	99				2	2	100	0	6	0	1	2	50									
HIV-1*	154	178	87	117	119	98	181	212	85	191	202	95	213	219	97	201	208	97	176	180	98	210	210	100	213	213	100
	11	13	85	7	7	100	1	1	100	42	42	100	0	2	0	9	9	100	32	34	94						
	3	22	14	86	86	100				2	2	100	0	6	0	1	1	100									
HIV-1-O	11	16	69	2	2	100	4	9	44				9	10	90	2	4	50	7	10	70	7	8	88	10	10	100
	7	7	100	12	14	86	1	1	100	10	10	100				0	3	0	2	2	100						
	1	3	33	5	6	83										0	1	0									
HIV-2				1	1	100				2	2	100															
	62	87	71	62	67	93	36	38	95	36	36	100	38	38	100	39	39	100	38	38	100	38	38	100	38	38	100
	0	18	0	8	8	100																					
MACAQUES+ SHIV				50	50	100	3	47	6	54	54	100	84	91	92				47	47	100	47	47	100	23	23	100
	31	94	33	44	44	100							0	2	0												
AGM	51	53	96	9	9	100	22	22	100	7	7	100				9	9	100	36	38	95	33	33	100	33	33	100
	47	61	77	22	31	71	0	4	0	45	45	100	1	1	100				24	24	100	1	1	100	1	1	100
	3	11	27	16	19	84																					
SOOTY MANGABEY				2	2	100							1	1	100				17	17	100						
	20	21	95	20	20	100	0	19	0	34	34	100	18	18	100				21	21	100	19	19	100	19	19	100
	0	15	0	14	14	100																					
SYKES's										1	1	100										1	2	50	3	3	100
	2	2	100										3	3	100	1	1	100	2	2	100	3	3	100			
	1	1	100																			0	1	0			
MANDRILL																									4	4	100
	5	15	33	7	10	70	1	1	100	9	9	100							1	1	100	3	3	100			
				3	3	100																					
CHIMPANZEES	3	3	100	2	2	100	8	12	67	7	7	100				4	4	100	6	7	86	17	19	89	12	12	100
	3	7	43							16	16	100										2	2	100			
	0	5	0	14	14	100										0	2	0				0	1	0			

or HIV-2 type (green boxes) were named "dual sequences".

Determination mode of LTR classification

In the present paper, the viral character was observed using as reference sequences the regulatory ones reported by Jones and Peterlin[12] for HIV-1s and by Guyader et al[15] for HIV-2s. They were compared to those reported by Krebs et al[1] which corresponded to more split sequences. As an example, the AGM characteristics were studied using the two groups of references and the results are shown in a supplementary TABLE available on line, Sheets 3 and 4. When using the Jones and Peterlin/Guyader references, a classical HIV-1 or HIV-2 belonging was detected together with a dual HIV-1/HIV-2 characteristic. The latter was not totally identified when referring to Krebs et al[1] since more than one third of the sequences could not be classified. Thus the sequences could belong either to HIV-1 or HIV-2 according to the considered modifications of the nucleotidic domain reference. The split sequences permitted refined studies of the viral expression regulation[1], whereas the sequences we chose allowed the harmonization of the human and simian domains together with the HIV-1 and HIV-2 classification, the two types of studies being complementary.

In addition to the compared lexical construction of the domains (nucleotide alignment), the short regulatory sequences were retained when localized at a position corresponding to that of the reference one. A HIV-type connection was possible and a global belonging (HIV-1, HIV-2 or HIV-1/HIV-2 types) was established

Regular Paper

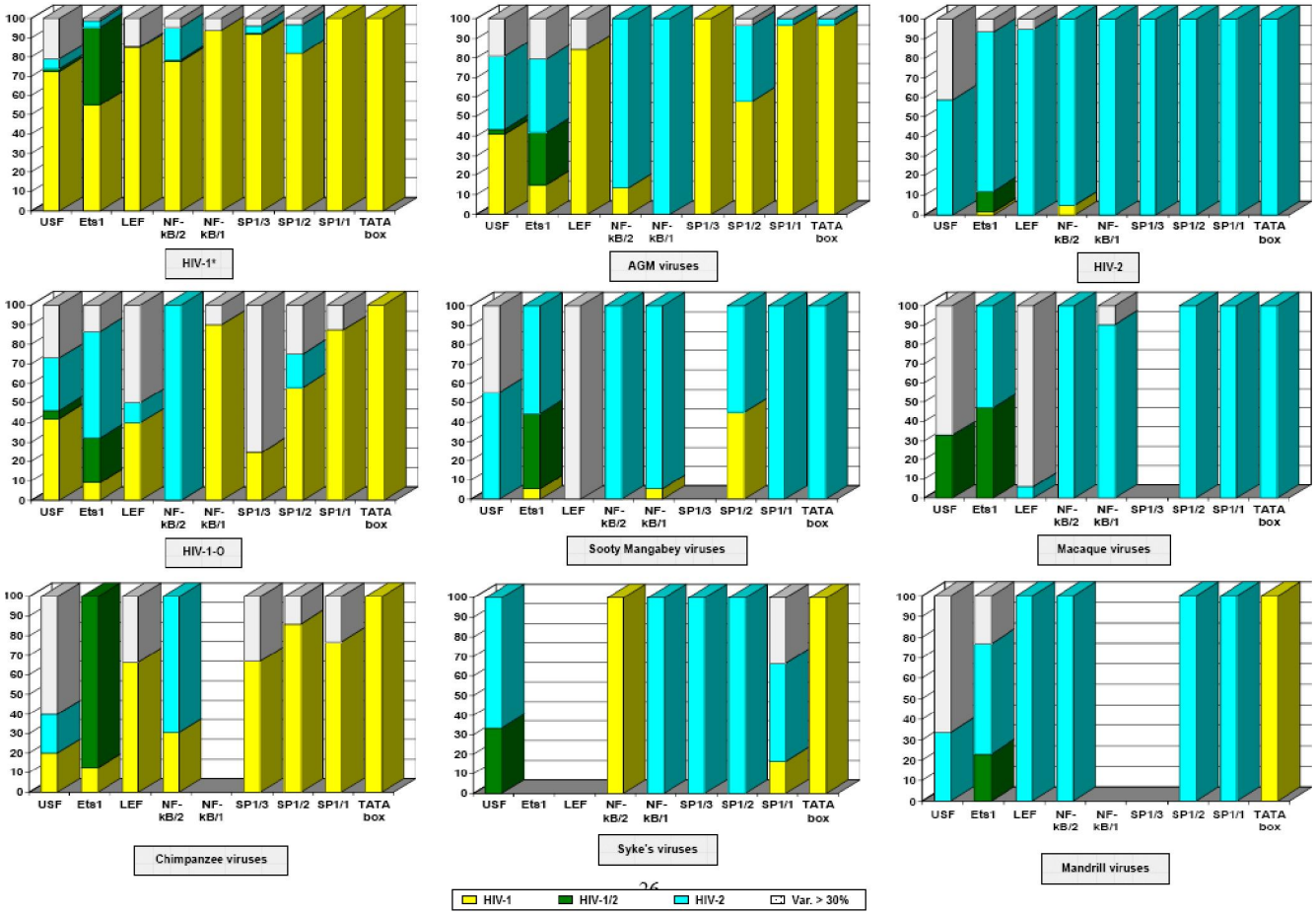

Figure 1 : Distribution of LTR regulatory sequences of simian viruses according to HIV belonging. Regulatory sequences are depicted from the TATA upstream to the USF sequence, from right to left in order to reproduce the organization of the LTR. The bars represent the percentage of each considered sequences. Colored bars correspond to sequences presenting a variation ≤ 30% and yellow, blue and green bars are for HIV-1, HIV-2 and dual HIV-1/HIV-2 sequences, respectively. Spotted bars show the percentage of sequences with variations > 30%. HIV-1* quotation indicates the viral strains which did not belong to HIV-1-O type.

for all the sequences for each virus family as reported on TABLE 1. Further we studied the sequences when presenting variations from the reference sequences inferior to the threshold leading to the generally admitted species jump, *i.e.* inferior or equal to 30%. The number and the percentage of these sequences are detailed in TABLE 2, which allowed to draw Figure 1. The results obtained under these conditions were close to those shown with all the sequences in TABLE 1. Only part of the sequences presented variations superior to 30% for a few simian viruses (see spotted bars in Figure 1). They were localized at the right position for each considered sequence and were mainly situated in the distal promoter (LEF, Ets1, USF). These sequences did not intervene in the promoter basal activity but principally in the viral host adaptation as discussed below. Consequently, our studies were pursued considering all the sequences when positioned at the right place.

Regulatory domain connections

U3 conserved domains are specific when considering the nature of the nucleotidic sequences (supplementary data available on line). The reference sequences were detected in the human and simian viral LTRs among families characteristic of the HIV-1 or HIV-2 reference group or both of them (dual HIV-1/HIV-2 belonging). The human viral LTRs presented exceptions when HIV-1 sequences were found in HIV-2 LTRs and conversely (mixed belonging). TABLE 1 and Figure 1 summarize the connection of the simian LTRs to the human ones, upstream the USF sequences to the TATA-box. Generally the simian LTRs belonged

Regular Paper

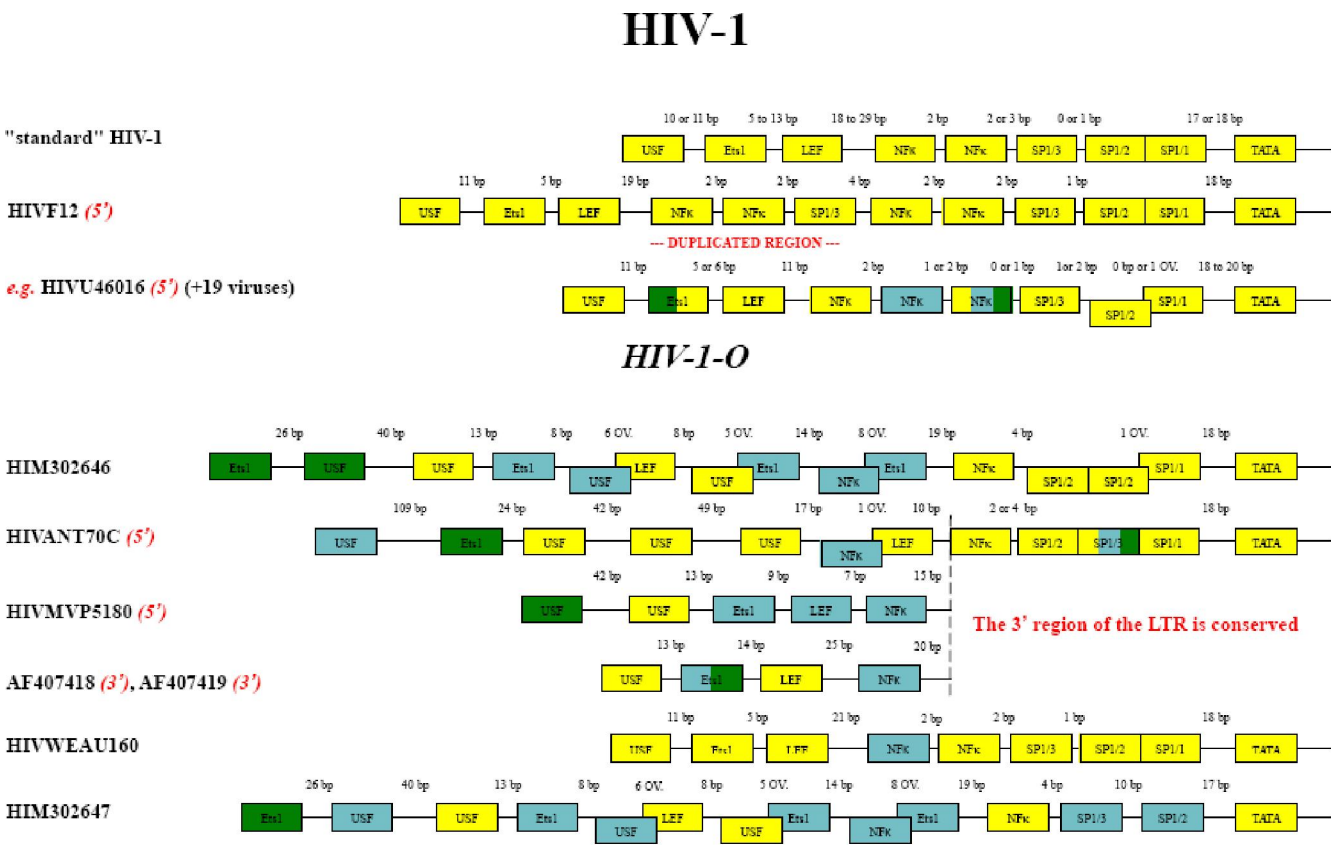

Figure 2 : Scheme of the HIV-1 LTRs Relative positions of regulatory sequences from the TATA upstream to the USF sequence are reported from right to left in order to reproduce the organization of the LTR. The numbers in red italic and between parentheses indicate the single LTR recovered for the strains when they usually presented the 3' and 5' ones. The colors are as in Figure 1 and double or tripled colored boxes indicate that they belong to HIV-1 and/or HIV-2 and/or HIV-1/HIV-2 type. The numbers of indicated base pairs represent the distance between the regulatory sequences. The "standard" HIV-1 organization, 2 of the more representative modified LTRs and specific HIV-1-O ones are reported. "OV." means overlapping concerning the number of indicated base pairs. The right part of the schemes from the dashed line corresponded to the conserved 3' region of the LTR.

either to HIV-1 or HIV-2's upstream from the TATA-box. The CPZ and about half AGM viruses were mainly related to HIV-1, while Macaque, Sooty Mangabey, Mandrill (except for half of the TATA-box) and Syke's viruses were close to HIV-2.

In the HIV-1s and HIV-2s the regulatory sequences were homogeneous except for the HIV-1-O viruses which exhibited a mixed belonging. Two types could be distinguished in simian viruses: either they were of the HIV-2 reference type (Macaque, Sooty Mangabey and Mandrill viruses) or they presented a preferential connection with many uncertainties, CPZ to HIV-1s or AGM to HIV-2s. Generally, the proximal regulatory sequences which mainly direct the viral expression (SP1, NF-κB) were clearly discriminated while the distal ones (Ets1, USF) presented a high rate of dual sequences. Such characteristic showed that these viruses had no rigid organization but large adaptation flexibility to the host or external events. When all the regulatory sequences were distributed without taking into account the 30% step of variations (see Materials and Methods), the scheme expressed in TABLE 1 was conserved when compared to that designed on Figure 1. However two main differences appeared for the LEF sequences which were typical of HIV-2 in Macaque and Sooty Mangabey viruses while they were not attributed in Figure 1. This sequence ambiguity corresponded to the viral host-adaptation phenomenon through ETs1[13, 17], LEF[18] or/and USF[19]. Cell-type specific viral replication can be influenced by either SP1/3 or NF-κB sites[14, 20] respectively, as well as both[21]. The overlap of SP1 and NF-κB sites complicated also the analysis[22] as did some dual belonging forms (*e.g.* SP1s in HIV-1 or Chimpanzee viruses). The use of LTR se-

Regular Paper

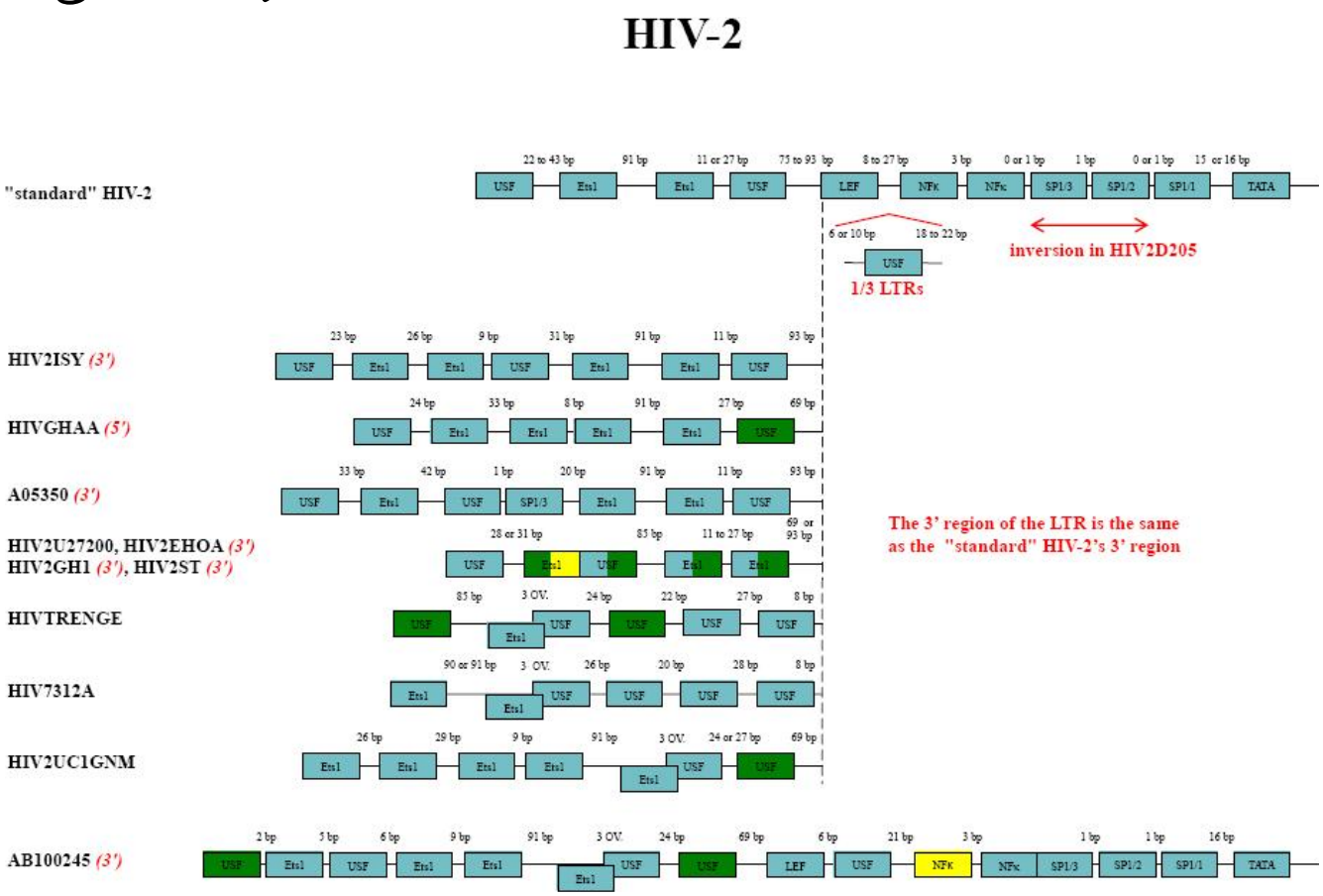

Figure 3 : Scheme of the HIV-2 LTRs. See legend Figure 2. The "standard" HIV-2 organization and 8 of the more representative modified LTRs are sketched which only differed upstream the LEF domain.

quences for subtyping[23] through the control of gene expression by host factors[21] illustrated the "herd evolution" as for example HIV-1 adaptation to its human host even after independent cross-species transmissions of SIV[24].

Global LTR organization

When considering the U3 organization, the HIV-1 standard organization (Figure 2) was generally conserved except for some noticeable cases. For example, the HIVF12 strain showed a repetition of the SP1/3-NF-κB2 domain between NF-κB2 and LEF sequences, while the HIV-1-O subtype (ANT70C and HIM302646 strains) was modified upstream the NF-κB1 sequence. The inversion of the LEF/NF-κB2 part in the ANT70C together with the above insertion in the HIVF12 strain revealed recombination events previously reported (*e.g.*[1, 8]). The HIV-2 standard organization (Figure 3) appeared well conserved from the TATA-box to LEF domain, whereas the upstream region exhibited specific combinations between the regulatory sequences. One third of the studied LTRs presented an extra USF box between NF-κB and LEF without noticeable connection with virulence. The global organization of HIV-1 and HIV-2 LTRs was generally maintained when the modifications occurred after the NF-κB/LEF proximal region. They consisted mostly in discrepancies in the position and the number of occurrences of ETs1, LEF and USF distal domains. Only a few overlaps were noticed indicating that the evolution is still ongoing: the more overlaps, the less the structure can evolve.

For the simian viruses, the standard organization was noted only for Macaque and Sooty Mangabey ones, which represented 85% and 42% of the studied LTRs, respectively (Figures 4 to 7). More complex were the observations for the other simian viruses as many divergences were revealed inside the species. The general organization was conserved from TATA-box to NF-

Regular Paper

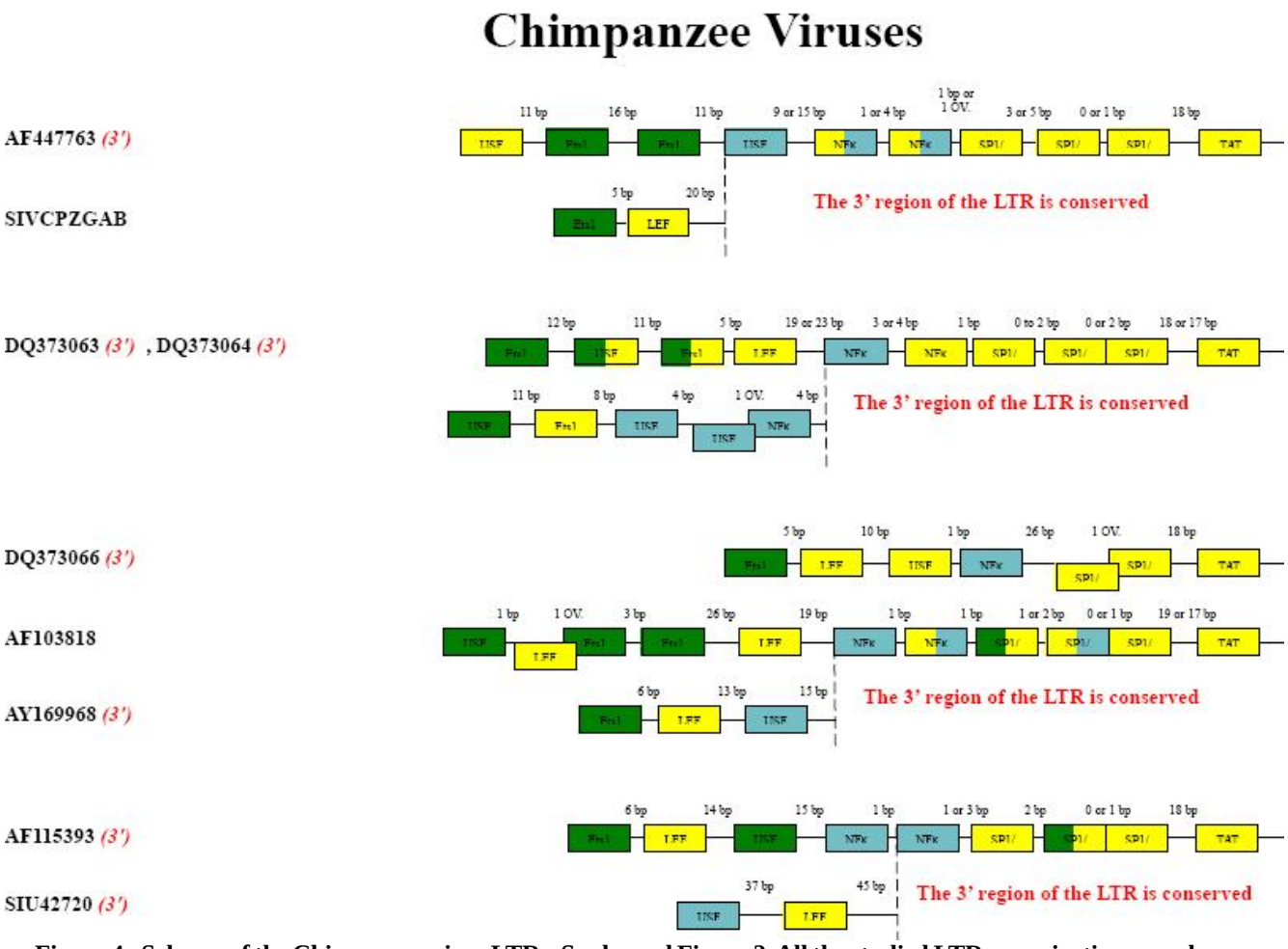

Figure 4 : Scheme of the Chimpanzee virus LTRs. See legend Figure 2. All the studied LTR organizations are shown.

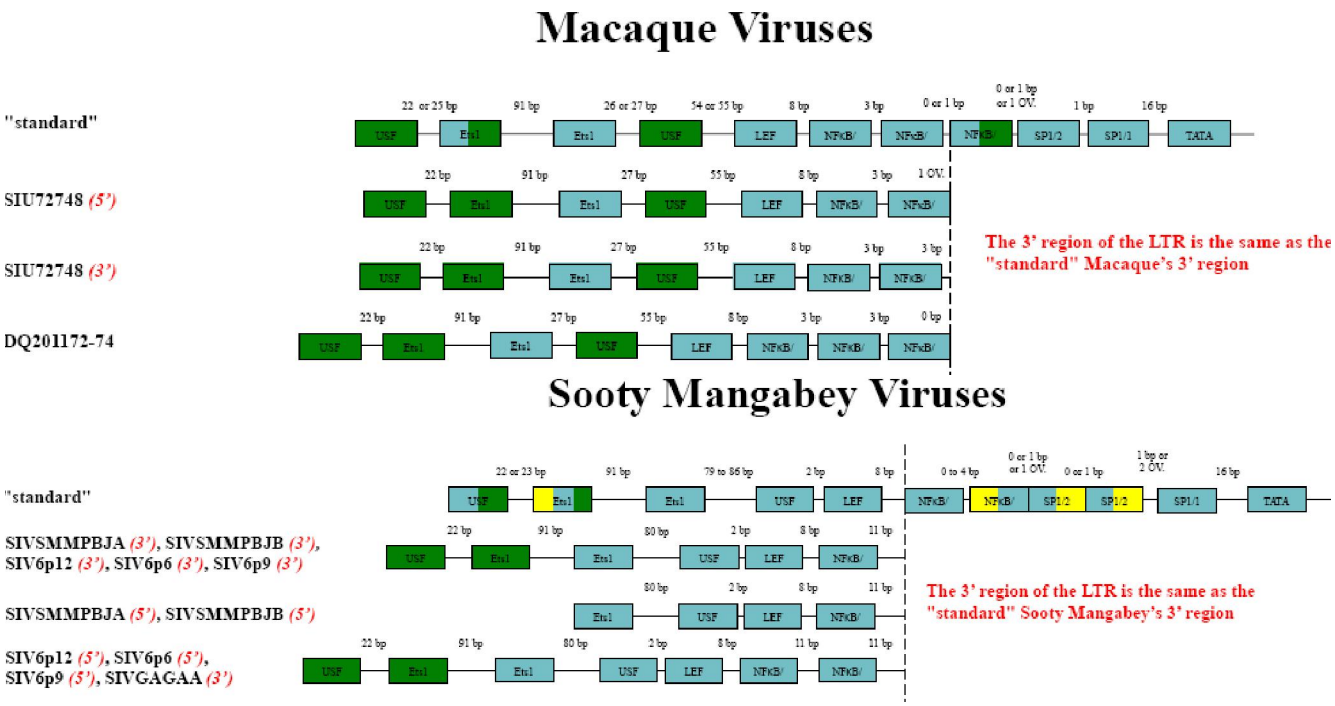

Figure 5 : Schemes of Macaque and Sooty-Mangabey viruses LTRs. See legend Figure 2. All the studied LTR organizations are shown.

Regular Paper

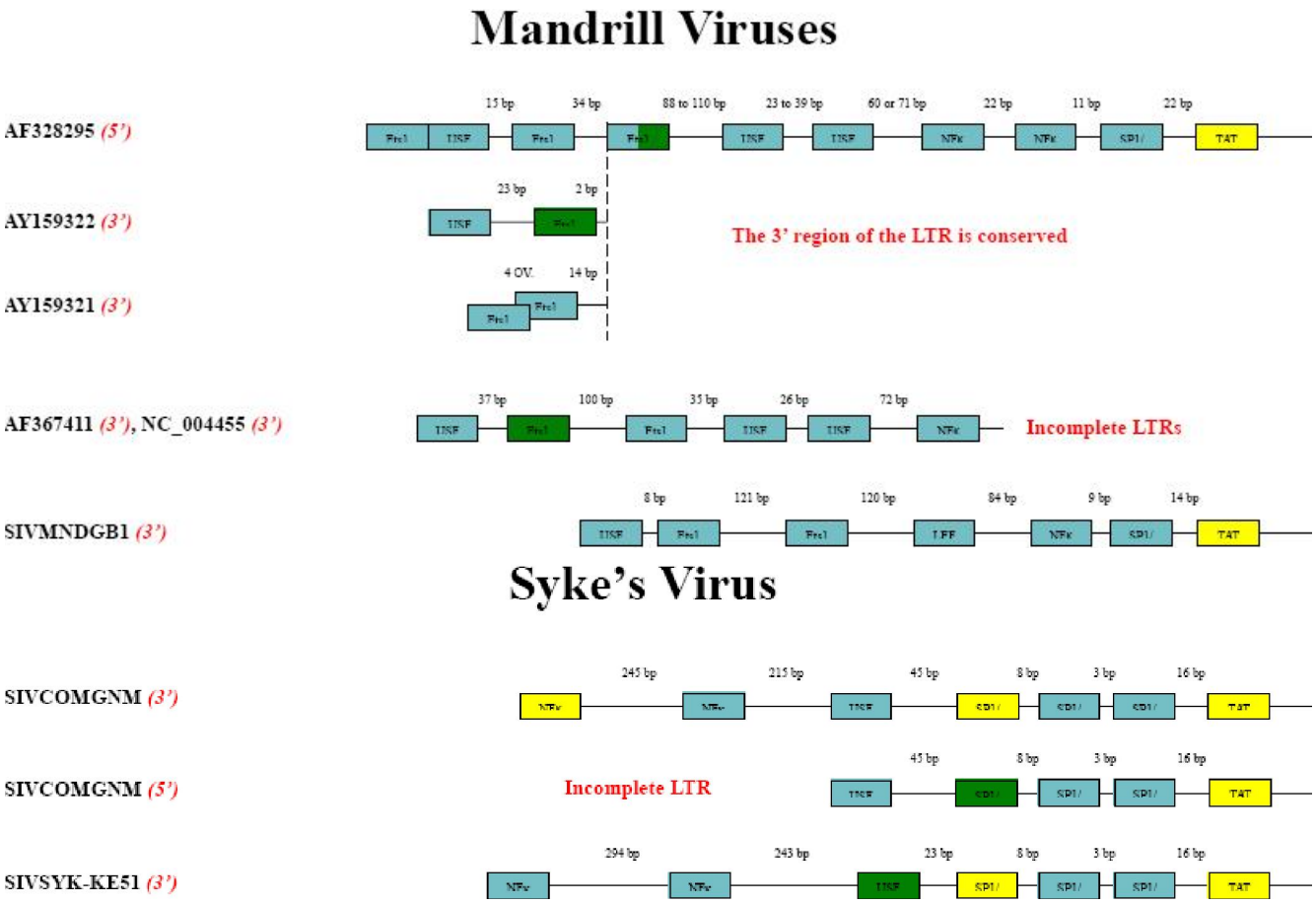

Figure 6 : Schemes of Mandrill and Syke's LTRs. See legend Figure 2. All the studied LTR organizations are shown.

κB1 with minor modifications and then specific structures intervened upstream. The CPZ viruses were related to the HIV-1s (Figure 4) while the Macaque, Sooty Mangabey, Mandrill (either mnd1 or 2 subtypes) and Syke's viruses were to HIV-2s (Figures 5 and 6). The Syke's viruses presented two SP1/1 sequences framing either a SP1/2 or a SP1/3 one and they exhibited typical inverted NF-κB/USF domains (Figure 6). AGM viruses offered the most diversified structure revealing differences between the four host sub-species (Figure 7A and B). Generally the host specificity was clearly noticed : Sabaeus (SAB 1C, SIU 21093, SIU 04008-04015) and Tantalus (SIU 58991, SIU 04016) on Figure 7A, Grivet (SIU 040006, 04007) and Vervet (SIU 04017, SIVAGM3, SIVAGM155) on Figure 7B but some strains had a common organization as SIVREV (Sabaeus) and SIVAGM90 (Vervet) without particular host specificity (Figure 7B). This recalled the AGM viruses subtyping reported by Jin et al[25]. Such LTRs showed a highly variable organization sometimes host-specific or belonging either to HIV-1, HIV-2 or both, which could imply a dual mapping organization.

Specificity of simian LTR organization

On the whole, it was clear that each simian virus owned a promoter presenting a characteristic organization for SP1 and NF-κB domains (TABLE 3). The SP1 domain was generally triplicated and the NF-κB duplicated as in HIV-1s and HIV-2s. When referring to Figures 2 to 7, these sites presented a variable distribution of SP1/1, SP1/2 and SP1/3 or NF-κB_1 and NF-κB_2. Three SP1 sequences were found in CPZ (sometimes only two), Sooty Mangabey, Syke's and AGM viruses. The latter had a large range of exceptions with 4, 2 or 1 SP1 sites. The Macaque viruses exhibited only two SP1 sequences when the Mandrill viruses had mainly one. The NF-κB_2 sequences were duplicated in CPZ, Mandrill and AGM viruses, which sometimes presented a single NF-κB_2 site or a classical NF-κB_1 plus NF-κB_2 organization for the Y00295 AGM virus

Regular Paper

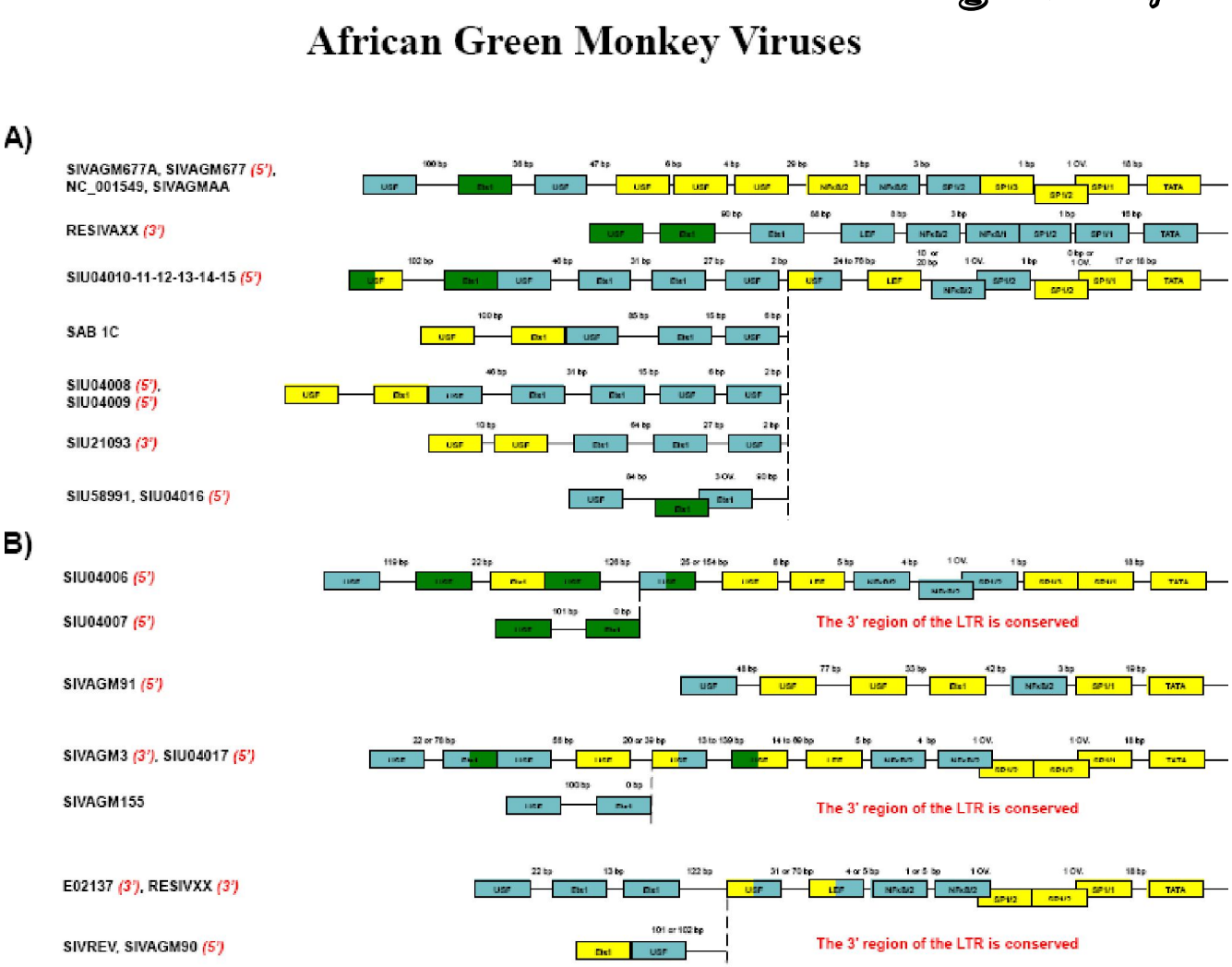

Figure 7 : Scheme of African Green Monkey LTRs. See legend Figure 2. All the studied LTR organizations are shown.

(former RESIVAXX). However, the latter result corresponding to a mixed organization between AGM and Macaque viruses may be linked to the production of virions after an infection of AGM with a Macaque virus[26]. Such organization was seen for the Sooty Mangabey viruses which exceptionally showed one or two extra NF-κB$_2$ sequences. The Macaque viruses usually exhibited two NF-κB$_1$ plus one NF-κB$_2$ sites, but sometimes had the opposite organization with one NF-κB$_1$ and two NF-κB$_2$ sites (SIU 72748 strain). The DQ 201172-74 strains presented a double NF-κB$_1$ plus a double NF-κB$_2$ adjacent positions. The Syke's viruses showed NF-κB$_1$ plus NF-κB$_2$ or double NF-κB$_1$ sequences.

The distribution of USF, LEF and Ets1 domains was difficult to unravel, the number and the position varying highly in each viral category and from a category to another one. One should note that some constant model was maintained in spite of the variations. As an example, the Macaque viruses LTR presented variations only after the third NF-κB sequence. The total length between this NF-κB and the last detected sequence was globally constant when considering the different regulatory sequences and spacers, (Figure 5). These observations plus the sequence overlappings emphasized the genomic reshaping occurring in that region, which was also seen for the other simian viruses, the Syke's one being the simplest. Nevertheless, AGM viruses exhibited the most divergent organization (Figure 7) when a nearly similar number of LTRs were analyzed for Macaque and AGM viruses. The limiting factor of these studies was the number of retrieved LTR nucleotidic sequences, especially for Mandrill and Syke's viruses.

Regulatory domain organization and virulence

Regular Paper

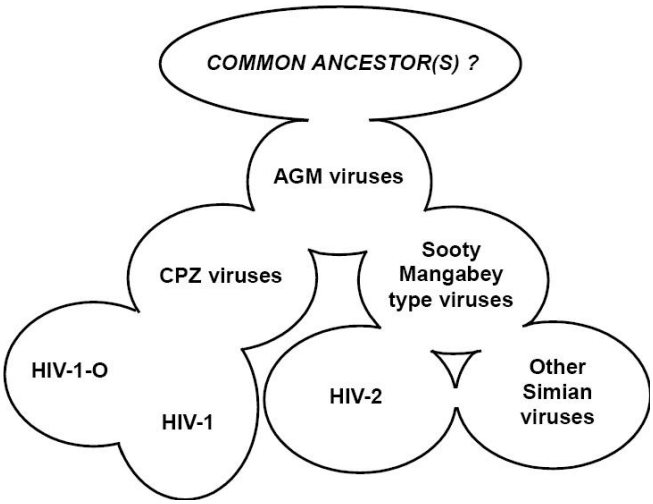

Figure 8 : Proposal for primate lentivirus evolution

TABLE 3 : Distribution of SP1 and NF-KB in simian viruses LTRs. Numbers represent the major frequencies of SP1 and NF-KB detection, the minor ones being indicated in parentheses.

Regulatory Sequences / Simian Viruses	NF-KB	SP 1
Chimpanzee	2(1 or 3)	3 (2)
Macaque	3 (4)	2
Sooty Mangabey	2 or 3 (4)	3
Mandrill	2 or 1	1 (2)
Sykes's	2 (0)	3
African Green Monkey	2 or 1	3 (1, 2 or 4)

Our results indicated that the LTRs are under a permanent recombination process. However any reshaping obtained should always be functional and present a maximal efficiency. In spite of these structural modifications, the global length of the enhancer domain was maintained when only the viable and major structures were detected, which biased the statistical interpretations. From another point of view, it was difficult to establish the relationship between both the number and type of the regulatory sequences and the subsequent infectivity of the considered viruses up to their clinical consequences. As an example of such complexity analysis, the Sooty Mangabey AF077017 that presented a "standard organization" (Figure 5) has been described as a neuropathogenic strain[27] when the other "standard" virus SIVMM 251 appeared to be non-pathogenic[26].

The increased number of NF-κB domains has been suggested to augment the viral virulence. For example, the Sooty Mangabey PBj14 strain (SIVSMMPBJA) which presented three NF-κB sites induced an acute immunodeficiency disease[28] and the PBj6.6 (SIV6p6, 5') one which had four NF-κB sites appeared as acutely lethal[29]. On the other hand, it is worth noting the HIV-1 HIVF12 which presented a duplicated SP1/3-NF-κB/1-NF-κB/2 region has been described as a non-producer clone which did not release particles in the medium[30]. This can indicate a specific behavior related to the host (human or not). It must be connected to the fact that a SIVmac was able to replicate in absence of the core enhancer[31] when it was not possible for a human virus type 1 or 2 to do so[32, 33]. These facts raised the question of adaptive changes in simian viruses to confer pathogenicity in humans (reviewed in[24, 34]). They reinforced the numerous data observed for a host-virus adaptation during cross-infections with simian viruses. Studies on virulence mainly involve regulatory sequences and associated proteins of the proximal promoter. However the viral expression is the result of complex cascade-combinations in which the distal promoter plays important roles as in host specificity and virus-host interactions.

Leading to a model of putative evolution

The difficulty to build lentiviral phylogenetic trees currently reported in the literature was due particularly to multiple-originated genomic segments or non-satisfactory computing methods. The mosaic recombinant viral structure was early considered as making complex the establishment of evolution trees depending on the genomic segment origin[6, 35-40]. Moreover, recombination should induce loss of phylogenetic information[9, 10]. Another difficulty was the choice of computing methods which can underestimate the variants[41] and also be time-consuming[42]. The main problem concerns the composition of databases[43] and their definition as the choice of parameters used for computation[44-46]. The methods can lead to false positive rate[47], incoherent data[3], underscore variants by ignoring some recombinations in phylogenetic analysis[48] or even incomplete data interpretation[5, 49, 50] as the recombinating HIV-1 genomes contained segments that cannot be classified[10, 51].

The approach considered in this paper gave to understand about the discriminating process leading to the

dual characteristics described below and observed for example in the AGM viruses. The literature reported that recombinant segments bore less phylogenetic information than non-recombinant one[10]. Consequently, it was necessary to build evolution studies on stable domains which have to be necessarily conserved. So, the data reported in TABLE 1 and Figure 1 were relevant to both the viral sequences and the viral host, the combination of these criteria establishing viral identifiers. The HIV-1s and HIV-2s well-defined strains generally showed sequences according to their type with some exceptions, mainly for HIV-1-Os. As to the simian viruses, Macaque, Sooty Mangabey, Mandrill and Syke's mostly exhibited HIV-2 sequences when CPZ and AGM presented mixed forms. Some differences occurred for the TATA box that could belong either to HIV-1, HIV-2 or rarely to both, the areas being proportional to the percentage of the respective detections. Moreover, CPZ and AGM viruses clearly displayed a dual HIV-1 and HIV-2 structure as they owned regulatory sequences of the two types. The duality concerned both the nature of sequences itself (*e.g.* SP1/2 for HIV-1 and HIV-2) and the presence of only one type of sequences (*e.g.* SP1/3 for HIV-1 and NF-κB1 for HIV-2).

CPZ and HIV-1-O viruses exhibited expected very similar schemes and this can lead to consider the HIV-1-Os as an evolution step between CPZ and HIV-1s. It was worth noting the similarity of these profiles with the AGM one (TABLE 1 and Figure 1). These results led us to present the data under three different groups showing a distribution according to three evolution potentialities. A first one exhibited the intermediate position of the HIV-1-O between HIV-1 and CPZ viruses. Another group reported the HIV-2 type and connected simian viruses. The in-between one revealed simian viruses with variable connections: *i.e.* the AGM viruses had ambivalent HIV-1/HIV-2 sequences while in Sooty Mangabey viruses HIV-2 sequences predominated. The interpretation of Syke's viruses was difficult due to their low number of retrieved sequences.

The viruses had U3 conserved sequences which can belong either to one or several types. Their high level of coexistence in the AGM viruses showed that they could represent a primary stage in the constitution of the present lentiviral genomes. This led us to suggest in Figure 8 a model for putative evolution from simian to human lentiviruses. We took into account the number of regulatory sequences that could be attributed to both types and the amount of the highly-potentially evolutive dual ones. As the evolutions did not concern individual viruses, the way of representation was to evoke inside circles the group or "herd" evolutions. The notions of reservoir, mosaic virus and quasi-species rendered contradictory the classical phylogenetic results. The continuous recombinating genome reshaping permitted the emergence of new lentiviral forms from which the only viable ones were conserved. Consequently, the ensuing produced viruses predominantly appertained to one of the families of the evolutional branch from which they emanated[6]. A strain cannot produce a simple progeny which can have consequently several direct ancestors[10]. This orientation should be directed by the selective pressure due to the homeostasis of the contaminated organism. Deep recombinations could eventually allow the issue of new virulent strains and even new viral families. All these criteria led us to consider the AGM group as the common ancestral one from which evolved the other lentiviruses. The HIV-1 branch could have originated through CPZ and HIV-1-O viruses. Alternatively, considering the dual sequences, a branch could have led to other simian viruses and HIV-2s with the Sooty-Mangabey viral group as possible intermediate common ancestral group.

CONCLUSIONS

The general scheme of the LTRs was conserved due to the selective pressure which directs the emergence of infecting strains in spite of multiple intra- and/or inter-LTR recombinations. The present study evidenced that AGM viruses could be distinguished from other simian ones, since they bore sequences belonging both to HIV-1 and HIV-2 viruses in the LTR core promoter that governs the viral expression. Moreover, as proposed in Figure 8, the AGMs can be considered as the common ancestral group of simian and human lentiviruses. This confirmed the sub-typing we showed when considering the CLS that evidenced all-directional cross-evolutions for all mammal

Regular Paper

lentiviruses[8], which was in agreement with many studies on simian and human lentiviruses. For example, CPZ/HIV-1 lineage could be not obviously specific[52] and Mandrill's SIVmnd-2 resulted from ancient recombinations between the Mandrill's SIVmnd-1 and Red-Capped Mangabey's SIVrcm[53]. The AGMs, which represent the largest reservoir of SIVs[54], also followed the host-adaptation phenomenon[55] using for instance the CXCR4 co-receptor to enter the human cells[56]. On the other hand the serum from Syke's virus-positive monkey immunoprecipitated *env* antigens from HIV-1 as well as from HIV-2, SIVmm and SIVagm[57]. The assumption of ancestral sequences expressed only for AGM viruses[58] was strengthened by the dual HIV-1/HIV-2 organization or by the common Sabaeus/REV-Vervet/AGM90 common structure as examples reported in this paper. Jin *et al*[25] and Sharp *et al*[59] also proposed a specific host-dependent evolution from common ancestor.

However, different interpretations and corrections to the evolution proposals appeared in the literature : an underestimation of the evolution clock[25], the observation that AGM viruses co-divergence does not follow the host evolution and possibly reaches to other simian viruses[58], the supposition that CPZ and human lentiviruses could derive from lentiviruses infecting other hosts[60], the assumption of a similar evolution rate for human and simian viruses[61], as examples. These considerations extended the general concept of lentiviral evolution including host-dependent evolution[62], independent cross-species transmissions[63, 64] or recombination events involving divergent lentiviruses in the distant past[65]. Such facts are related to the notion of "herd evolution" inducing a phenotypic heterogeneity due to the genomic multistability, which can be influenced by numerous factors as environmental conditions inside or outside the host. Consequently, based on these findings we can consider that AGM viruses correspond to an important evolution crossroad pointing to common ancestral sequences for lentiviruses and particularly for HIV-1 and HIV-2 groups.

ACKNOWLEDGMENTS

We thank Pr. André Klier and Dr. Lisa Chakrabarti for their fruitful commentaries on this manuscript and Géraldine Klein for her help in collating a part of the sequences.

REFERENCES

[1] F.C.Krebs, T.Hogan, S.Quiterio, S.Gartner, B.Wigdahl; Los Alamos National Laboratory HIV databases, 1-25 **(2005)**.
[2] M.L.Moncany, P.R.Courtois; J.Mol.Evol., **43**, 152-160 **(1996)**.
[3] A.Morris, M.Marsden, K.Halcrow, E.S.Hughes, R.P.Brettle, J.E.Bell, P.Simmonds; J.Virol., **73**, 8720-8731 **(1999)**.
[4] F.E.McCutchan, J.L.Sankale, S.M'Boup, B.Kim, S.Tovanabutra, D.J.Hamel, S.K.Brodine, P.J.Kanki, D.L.Birx; AIDS Res Hum Retroviruses, **20**, 819-826 **(2004)**.
[5] K.Wilbe, C.Casper, J.Albert, T.Leitner; AIDS Res Hum Retroviruses, **18**, 849-856 **(2002)**.
[6] J.K.Carr, N.D.Wolfe, J.N.Torimiro, U.Tamoufe, E.Mpoudi-Ngole, L.Eyzaguirre, D.L.Birx, F.E.McCutchan, D.S.Burke; Retrovirology, **7**, 39 **(2010)**.
[7] S.Bonhoeffer, E.C.Holmes, M.A.Nowak; Nature, **376**, 125 **(1995)**.
[8] M.L.Moncany, K.Dalet, P.R.Courtois; C.R.Biol., **329**, 751-764 **(2006)**.
[9] M.Worobey; Mol.Biol.Evol., **18**, 1425-1434 **(2001)**.
[10] G.Magiorkinis, D.Paraskevis, H.A.Schmidt, A.Hatzakis; Infect Genet.Evol., **8**, 139-145 **(2008)**.
[11] K.T.Jeang, V.Yedavalli; Nucleic.Acids Res., **34**, 4198-4205 **(2006)**.
[12] K.A.Jones, B.M.Peterlin; Annu.Rev.Biochem., **63**, 717-743 **(1994)**.
[13] R.E.Jeeninga, M.Hoogenkamp, M.Armand-Ugon, M.De Baar, K.Verhoef, B.Berkhout; J.Virol., **74**, 3740-3751 **(2000)**.
[14] J.J.McAllister, D.Phillips, S.Millhouse, J.Conner, T.Hogan, H.L.Ross, B.Wigdahl; Virology, **274**, 262-277 **(2000)**.
[15] M.Guyader, M.Emerman, P.Sonigo, F.Clavel, L.Montagnier, M.Alizon; Nature, **326**, 662-669 **(1987)**.
[16] P.R.Courtois, M.L.Moncany; Comput.Appl.Biosci., **11**, 657-665 **(1995)**.
[17] M.Horiba, L.B.Martinez, J.L.Buescher, S.Sato, J.Limoges, Y.Jiang, C.Jones, T.Ikezu; J.Gen.Virol., **88**, 236-241 **(2007)**.

Regular Paper

[18] E.Ramirez De Arellano, C.Martin, V.Soriano, J.Alcami, A.Holguin; Virus Genes., **34**, 111-116 **(2007)**.
[19] N.M.Clark, M.C.Hannibal, D.M.Markovitz; J.Virol., **69**, 4854-4862 **(1995)**.
[20] S.Ranjbar, A.V.Tsytsykova, S.K.Lee, R.Rajsbaum, J.V.Falvo, J.Lieberman, P.Shankar, A.E.Goldfeld; PLoS.Pathog., **2**, e130 **(2006)**.
[21] J.C.Burnett, K.Miller-Jensen, P.S.Shah, A.P.Arkin, D.V.Schaffer; PLoS.Pathog., **5**, e1000260 **(2009)**.
[22] M.R.Nonnemacher, T.H.Hogan, S.Quiterio, B.Wigdahl, A.Henderson, F.C.Krebs; Biomed. Pharmacother., **57**, 34-40 **(2003)**.
[23] E.R.De Arellano, V.Soriano, A.Holguin; AIDS Res Hum Retroviruses, **21**, 949-954 **(2005)**.
[24] L.V.Wain, E.Bailes, F.Bibollet-Ruche, J.M.Decker, B.F.Keele, F.Van Heuverswyn, Y.Li, J.Takehisa, E.M.Ngole, G.M.Shaw, M.Peeters, B.H.Hahn, P.M.Sharp; Mol.Biol.Evol., **24**, 1853-1860 **(2007)**.
[25] M.J.Jin, H.Hui, D.L.Robertson, M.C.Muller, F.Barre-Sinoussi, V.M.Hirsch, J.S.Allan, G.M.Shaw, P.M.Sharp, B.H.Hahn; Embo.J., **13**, 2935-2947 **(1994)**.
[26] G.Franchini, C.Gurgo, H.G.Guo, R.C.Gallo, E.Collalti, K.A.Fargnoli, L.F.Hall, F.Wong-Staal, M.S.Reitz, Jr.; Nature, **328**, 539-543 **(1987)**.
[27] F.J.Novembre, J.De Rosayro, S.P.O'Neil, D.C.Anderson, S.A.Klumpp, H.M.McClure; J.Virol., **72**, 8841-8851 **(1998)**.
[28] S.Dewhurst, J.E.Embretson, D.C.Anderson, J.I.Mullins, P.N.Fultz; Nature, **345**, 636-640 **(1990)**.
[29] F.J.Novembre, P.R.Johnson, M.G.Lewis, D.C.Anderson, S.Klumpp, H.M.McClure, V.M.Hirsch; J.Virol., **67**, 2466-2474 **(1993)**.
[30] M.Federico, F.Titti, S.Butto, A.Orecchia, F.Carlini, B.Taddeo, B.Macchi, N.Maggiano, P.Verani, G.B.Rossi; AIDS Res.Hum.Retroviruses., **5**, 385-396 **(1989)**.
[31] F.Kirchhoff, H.W.Kestler.3rd, R.C.Desrosiers; J.Virol., **68**, 2031-2037 **(1994)**.
[32] J.Y.Kim, F.Gonzalez-Scarano, S.L.Zeichner, J.C.Alwine; J.Virol., **67**, 1658-1662 **(1993)**.
[33] S.L.Zeichner, J.Y.Kim, J.C.Alwine; Gene.Expr., **1**, 15-27 **(1991)**.
[34] P.M.Sharp, G.M.Shaw, B.H.Hahn; J.Virol., **79**, 3891-3902 **(2005)**.
[35] S.H.Eshleman, M.J.Gonzales, G.Becker-Pergola, S.C.Cunningham, L.A.Guay, J.B.Jackson, R.W.Shafer; AIDS Res.Hum.Retroviruses, **18**, 507-511 **(2002)**.
[36] M.A.Papathanasopoulos, G.M.Hunt, C.T.Tiemessen; Virus.Genes., **26**, 151-163 **(2003)**.
[37] E.Castro, G.Echeverria, L.Deibis, B.Gonzalez De Salmen, A.Dos Santos Moreira, M.L.Guimaraes, F.I.Bastos, M.G.Morgado; J.Acquir.Immune. Defic.Syndr., **32**, 338-344 **(2003)**.
[38] G.M.Hunt, M.A.Papathanasopoulos, G.E.Gray, C.T.Tiemessen; Virus.Genes., **26**, 49-56 **(2003)**.
[39] M.Salemi, T.De Oliveira, V.Courgnaud, V.Moulton, B.Holland, S.Cassol, W.M.Switzer, A.M.Vandamme; J.Virol., **77**, 7202-7213 **(2003)**.
[40] S.Tovanabutra, C.Beyrer, S.Sakkhachornphop, M.H.Razak, G.L.Ramos, T.Vongchak, K.Rungruengthanakit, P.Saokhieo, K.Tejafong, B.Kim, M.De Souza, M.L.Robb, D.L.Birx, J.Jittiwutikarn, V.Suriyanon, D.D.Celentano, F.E.McCutchan; AIDS Res.Hum.Retroviruses., **20**, 465-475 **(2004)**.
[41] D.Posada; Mol.Biol.Evol., **19**, 708-717 **(2002)**.
[42] G.Didier, L.Debomy, M.Pupin, M.Zhang, A.Grossmann, C.Devauchelle, I.Laprevotte; BMC Bioinformatics, **8**, 1 **(2007)**.
[43] G.Myers, K.MacInnes, B.Korber; AIDS Res.Hum. Retroviruses., **8**, 373-386 **(1992)**.
[44] E.Op De Coul, A.Van Der Schoot, J.Goudsmit, R.Van Den Burg, W.Janssens, L.Heyndrickx, G.Van Der Groen, M.Cornelissen; Virology, **270**, 267-277 **(2000)**.
[45] A.C.Siepel, B.Korber; Los Alamos National Laboratory, **III**, 35-65 **(1995)**.
[46] A.C.Siepel, A.L.Halpern, C.Macken, B.T.Korber; AIDS Res.Hum.Retroviruses., **11**, 1413-1416 **(1995)**.
[47] J.M.Carlson, Z.L.Brumme, C.M.Rousseau, C.J.Brumme, P.Matthews, C.Kadie, J.I.Mullins, B.D.Walker, P.R.Harrigan, P.J.Goulder, D.Heckerman; PLoS.Comput.Biol., **4**, e1000225 **(2008)**.
[48] J.Graham, B.McNeney, F.Seillier-Moiseiwitsch; Bioinformatics, **21**, 589-595 **(2005)**.
[49] N.C.Grassly, E.C.Holmes; Mol.Biol.Evol., **14**, 239-247 **(1997)**.
[50] J.K.Carr, M.O.Salminen, C.Koch, D.Gotte, A.W.Artenstein, P.A.Hegerich, D.St Louis, D.S.Burke, F.E.McCutchan; J.Virol., **70**, 5935-5943 **(1996)**.
[51] G.Magiorkinis, D.Paraskevis, E.Magiorkinis, A.M.Vandamme, A.Hatzakis; Infect.Genet.Evol., **5**, 225-229 **(2005)**.
[52] P.Sharp, D.L.Robertson, F.Gao, B.H.Hahn; AIDS, 8 **(1994)**.
[53] T.Takemura, M.Hayami; Front.Biosci., **9**, 513 **(2004)**.

Regular Paper

[54] I.Pandrea, C.Apetrei, J.Dufour, N.Dillon, J.Barbercheck, M.Metzger, B.Jacquelin, R.Bohm, P.A.Marx, F.Barre-Sinoussi, V.M.Hirsch, M.C.Muller-Trutwin, A.A.Lackner, R.S.Veazey; J.Virol., **80**, 4858-4867 **(2006)**.

[55] M.C.Muller, N.K.Saksena, E.Nerrienet, C.Chappey, V.M.Herve, J.P.Durand, P.Legal-Campodonico, M.C.Lang, J.P.Digoutte, A.J.Georges, et al.; J.Virol., **67**, 1227-1235 **(1993)**.

[56] S.M.Owen, S.Masciotra, F.Novembre, J.Yee, W.M.Switzer, M.Ostyula, R.B.Lal; J.Virol., **74**, 5702-5708 **(2000)**.

[57] P.Emau, H.M.McClure, M.Isahakia, J.G.Else, P.N.Fultz; J.Virol., **65**, 2135-2140 **(1991)**.

[58] J.O.Wertheim, M.Worobey; PLoS.Pathog., **3**, e95 **(2007)**.

[59] P.M.Sharp, E.Bailes, D.L.Robertson, F.Gao, B.H.Hahn; Biol.Bull., **196**, 338-342 **(1999)**.

[60] M.M.Vanden Haesevelde, M.Peeters, G.Jannes, W.Janssens, G.Van Der Groen, P.M.Sharp, E.Saman; Virology, **221**, 346-350 **(1996)**.

[61] J.O.Wertheim, M.Worobey; PLoS.Comput.Biol., **5**, e1000377 **(2009)**.

[62] J.Hu, W.M.Switzer, B.T.Foley, D.L.Robertson, R.M.Goeken, B.T.Korber, V.M.Hirsch, B.E.Beer; J.Virol., **77**, 4867-4880 **(2003)**.

[63] B.H.Hahn, G.M.Shaw, K.M.De Cock, P.M.Sharp; Science, **287**, 607-614 **(2000)**.

[64] K.Hiebenthal-Millow, S.Pohlmann, J.Munch, F.Kirchhoff; Virology, **324**, 501-509 **(2004)**.

[65] M.C.Georges-Courbot, C.Y.Lu, M.Makuwa, P.Telfer, R.Onanga, G.Dubreuil, Z.Chen, S.M.Smith, A.Georges, F.Gao, B.H.Hahn, P.A.Marx; J.Virol., **72**, 600-608 **(1998)**.