

## Pattern of the evolution of HIV-1 *env* gene in Côte d'Ivoire

Sery Gonedelé Bi<sup>1, 2\*</sup>, Didier P Sokouri<sup>1</sup>, Kouakou Tiékoura<sup>1</sup>, Oulo Alla N'Nan<sup>1</sup>, Marcel Lolo<sup>1</sup>, Félix Gnangbé<sup>1</sup> & Assanvo SP N'Guetta<sup>1</sup>

<sup>1</sup>Laboratoire de Génétique, UFR Biosciences, Université Félix Houphouët Boigny, Abidjan, Côte d'Ivoire, 22 BP 582 Abidjan 22;

<sup>2</sup>Centre Suisse de Recherches Scientifiques en Côte d'Ivoire, 01 BP 1303 Abidjan 01; Sery Gonedelé Bi - Email: gonedelebi@yahoo.fr; Phone: 00(225) 03 48 89 23; Fax: 00(225) 23 45 12 11; \*Corresponding author

Received September 25, 2014; Revised November 11, 2014; Accepted November 12, 2014; Published November 27, 2014

### Abstract:

Côte d'Ivoire continues to have the highest HIV-1 prevalence rate in West Africa, although the infection number is in constant decline. The external envelope protein of the viruses is a likely site of selection, and responsible for receptor binding and entry into host cells, and therefore constitutes an ideal region with which to investigate the evolutionary processes acting on HIV-1. In this study, we analyse 189 envelope glycoprotein V3 loop region sequences of virus isolates from 1995 to 2009, from HIV-1 untreated patients living in Côte d'Ivoire, to decipher the temporal relationship between disease diversity, divergence and selection. Our analyses show that the nonsynonymous and synonymous ratio (dN/dS) was lower than 1 for viral populations analysed within 15 years, which showed the sequences did not undergo adequate immune pressure. The phylogenetic tree of the sequences analysed demonstrated distinctly long internal branches and short external branches, suggesting that only a small number of viruses infected the new host cell at each transmission. In addition to identifying sites under purifying selection, we also identified neutral sites that can cause false positive inference of selection. These sites presented form a resource for future studies of selection pressures acting on HIV-1 *env* gene in Côte d'Ivoire and other West African countries.

**Keywords:** HIV-1, *env* gene, diversity, selection, Côte d'Ivoire

### Background:

Since the first AIDS case detection in Côte d'Ivoire in 1985, the infection number is in constant decline with an actual estimated prevalence of 3.7%. Although this constant decrease, Côte d'Ivoire continues to have the highest HIV-1 prevalence rate in West Africa and 60% of HIV-infected patients are women, most of them of childbearing age [1]. Based on partial polymerase (*pol*) and/or envelope (*env*) sequences, the high prevalence of circulating recombinant form CRF02\_AG (82%) and cocirculation of subtype A (5%), CRF01\_AE (1%), CRF06\_cpx (4%), and complex intersubtype recombinants (11%) has been documented in Côte d'Ivoire [2]. One important feature of HIV-1 infection is the diversification and evolution of the viral genome over the course of infection. From all the protein encoding genes, the most variable is the *env* gene. It encodes for the envelope proteins associated with the host cell-HIV interaction [3]. Changes in this highly

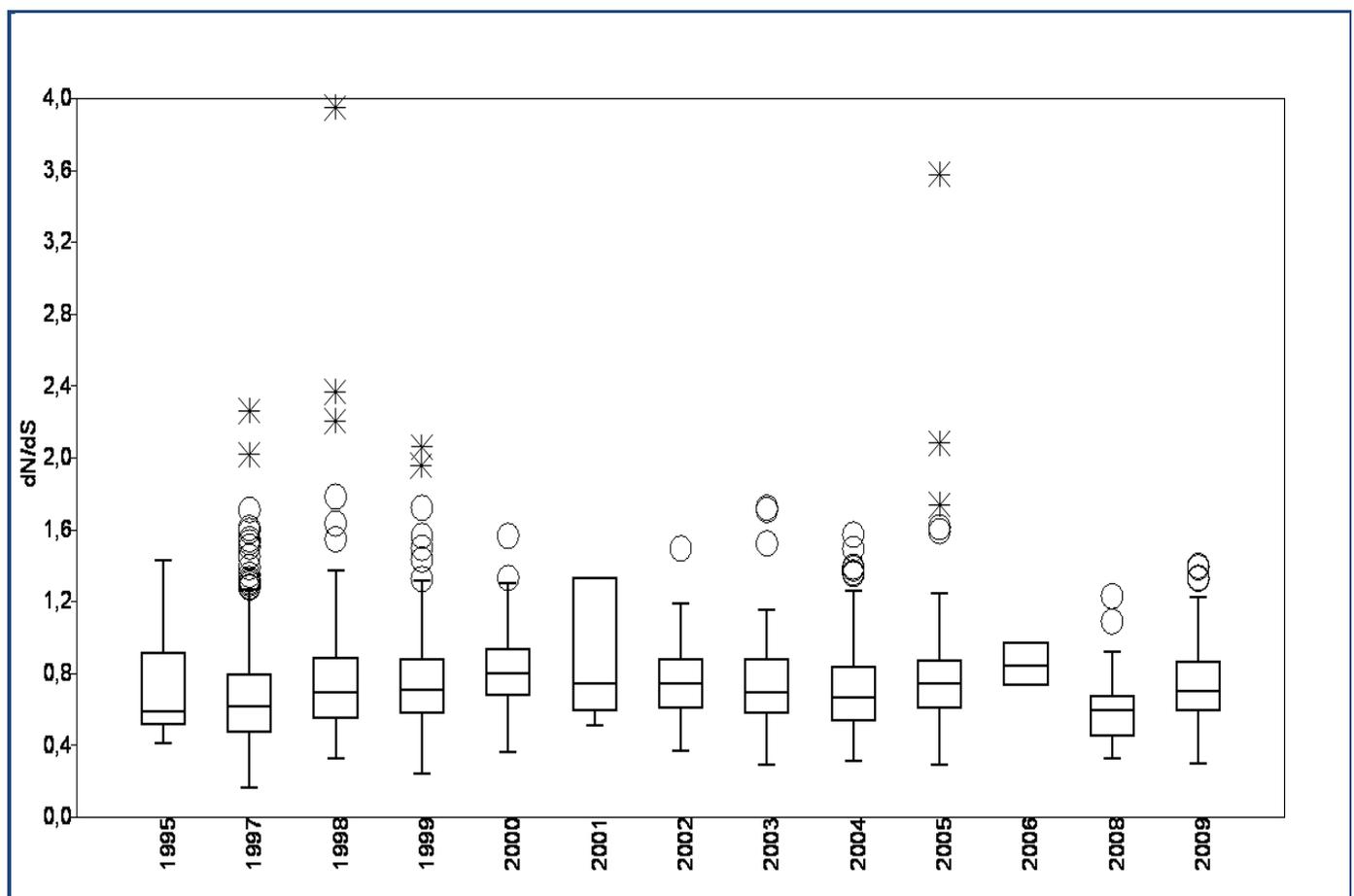
conserved residue provide an interesting case of study to test whether selective pressure was altered with the substitution.

Nevertheless, due to their functional relevance, several amino acid residues are extremely conserved among HIV-1 variants. The external envelope protein is a likely site of selection, being targeted by the patient's antibody response [4] and responsible for receptor binding and entry into host cells [5], and therefore constitutes an ideal region with which to investigate the evolutionary processes acting on HIV-1. The long term fate of these abundant genetic changes depends on the interplay of effective population size and natural selection, resulting in an extremely high rate of HIV genomic evolution [6]. Population level process such as selection, migration, population dynamic and recombination shape HIV genetic diversity both among and within hosts [6]. The ratio of nonsynonymous/synonymous substitution rates has proved useful in

investigating molecular adaptation; however, changes in the absolute rates of nonsynonymous and synonymous substitution should provide greater insight [7]. Changes in synonymous substitution rates can reflect changes in generation time or mutation rate, while nonsynonymous rates can also be affected by changes in selective pressure and effective population size. Previous studies of HIV evolution have typically assumed that the rate of neutral or synonymous change (per month or year) is approximately constant among patients [8].

Differences in the mutational profile among HIV subtypes have been reported [8]. Such high viral genetic diversity among subtypes is involved in difference in the rate of disease progression and response to antiretroviral therapy including the development of resistance [9]. Therefore it is crucial to acquire further knowledge concerning the real significance of

these differences; it may be important to determine strategies of initial treatment for infected individuals. Studying the evolutionary relationship of HIV-1 and characterizing the distinct adaptation patterns in different parts of the HIV-1 genome that interact with the immune system will be key to elucidate how HIV-1 overwhelms the immune system and leads to AIDS [10]. In this study, we present sequence analyses of envelope glycoprotein V3 loop region of virus isolates from HIV-1 untreated patients living in Côte d'Ivoire, to decipher the temporal relationship between diversity, divergence and selection, in the HIV-1 envelope gene. Understanding the process that determines viral genetic diversity will undoubtedly assist in the struggle against viral infections and will contribute to our knowledge of past epidemiologic events in Côte d'Ivoire.



**Figure 1:** Synonymoud and nonsynonimous ratio (dN/dS) plot along HIV-1 *env* gene. Interhost overall dN/dS ratios at all sites within the *env* gene fragments from HIV-1-infected individuals. Box plot showing the mean, standard error and 95% confidence interval for dN/dS ratios obtained for each codon of the HIV-1 alignments. Statistical significance was determined using the Kruskal-Wallis test.

## Methodology:

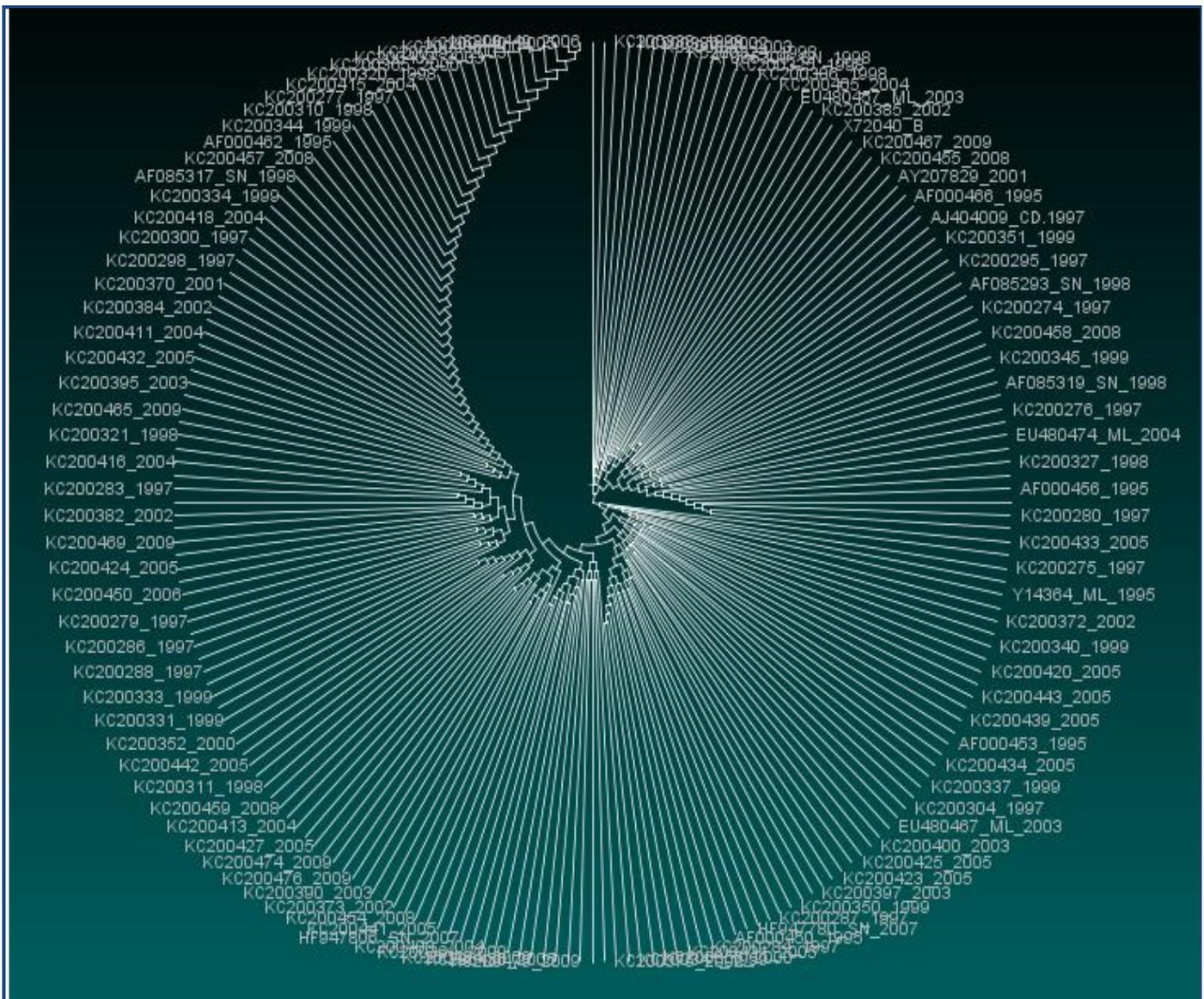
### Data sets compilation

All HIV-1 sequences classified as subtype A derived from Côte d'Ivoire were downloaded from the Los Alamos National Laboratory and GenBank databases. Pseudogenes (as noted in GenBank), clones and sequences with less than 250 bp were excluded from the following analyses. The sequences included in this work were from individuals in the asymptomatic phase

of infection and they were naïve to drug therapy. The description of data sets and the GenBank accession number of each sequence are summarized as supplementary material. The final set included the 189 subtype A sequences described and four non-A sequences, (subtypes B), which were used as out-groups. Fifteen (18) sequences originating from other African countries were also included for phylogenetic comparison: 4 from Mali, 8 from Senegal, and 6 from Congo Democratic

Republic. The sequences were first aligned using the ClustalX program [11]. All sites with deletions and insertions were then excluded in order to preserve the reading frames of the genes.

The final alignment was 406 bp long and is presented as supplementary information.



**Figure 2:** Radial phylogeny tree reconstructed by Maximum likelihood methods based on fragment of HIV-1 subtype A *env* V3 loop sequence isolated from HIV-1 untreated patients in Côte d'Ivoire from 1995 to 2009. The tree was generated using the GTR + I + G model of nucleotide substitution. Sequences were named according to their accession number and year of isolation

### Phylogenetic inference

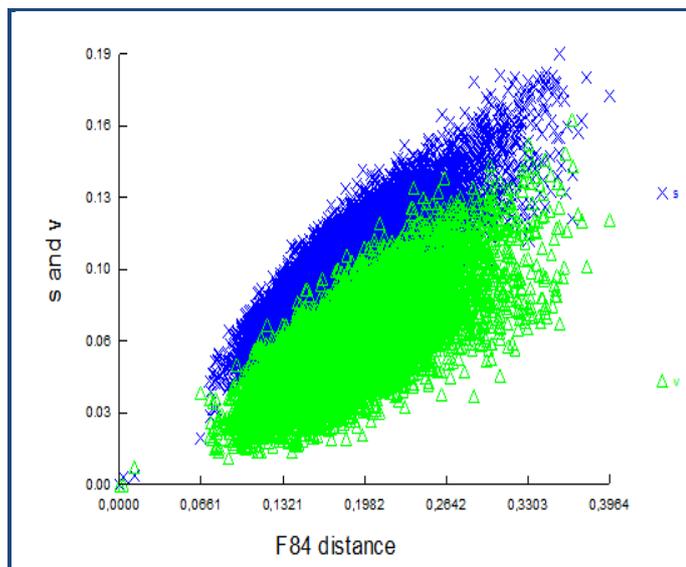
For the maximum likelihood analysis of selection pressures, phylogenetic trees were constructed. We first determined the most appropriate model of nucleotide substitution for each data set using the program jModeltest 2.1 [12]. Models GTR+G, and GTR+I+G were suggested to have better Likelihood scores. Then we reconstructed the phylogenetic tree using the ML method under GTR+G and GTR+I+G methods. We used Markov chain Monte Carlo (MCMC) methods as implemented in BEAST 1.7 to obtain a posterior distribution of trees under an uncorrelated relaxed clock [13]. In order to assess confidence in each of the internal nodes of the constructed phylogeny, a bootstrap resampling (1,000 replicates) of the data using the neighbor-joining method based on maximum-likelihood distances performed with FigTree [14]. To

investigate the diversity change, we inferred between-host mean diversities for the 1995s to 2009s, using the nucleotide diversity,  $p$ , implemented in MEGA 6 [15] under the GTR+G model again. For Setup Data, the viral sequences obtained from the same year were grouped as one subpopulation. Then, the within-year diversity was calculated by Mean Diversity within Subpopulations, whereas the between-year diversity was calculated by Mean Interpopulational Diversity.

### Analysis of selective pressures

Codon models of coding sequence evolution were used to detect positive selection operating on the HIV-1 *env* gene. In particular, we were interested in differences in positive selection pressure on the virus from the 1995s to the 2009s. Selective pressures were analyzed using two distinct

approaches that estimate the number of nonsynonymous (dN) and synonymous (dS) at all sites in the sequence alignments. This compares the fit to the data of various models of codon evolution, which differ in the distribution of nonsynonymous and synonymous ratio (dN/dS) among sites and takes into account the phylogenetic relationships of the sequences. HyPhy software [16] was used to generate simulated data under a neutral model with trees generated from the original alignments. The same sequence alignments used as input in the initial analysis were used and one hundred simulated datasets were generated for each alignment. Each simulated dataset was then analyzed using the Dual Model as described above. The minimum value of mean dS across all sliding windows of three adjacent codons, in all of the one hundred simulated datasets, was used as a conservative threshold to identify windows of reduced dS in the observed data. This stringent threshold and a less stringent one that included 95% of the values inferred from the simulated data are shown in the sliding window plots.



**Figure 3:** Transitions and transversions versus genetic divergence plots of *env* gene fragment isolated from HIV-1-infected individuals in Côte d'Ivoire (s: transitions, v: transversions). The estimated number of transitions and transversions for each pairwise comparison was plotted against the genetic distance.

### Codon usage analysis

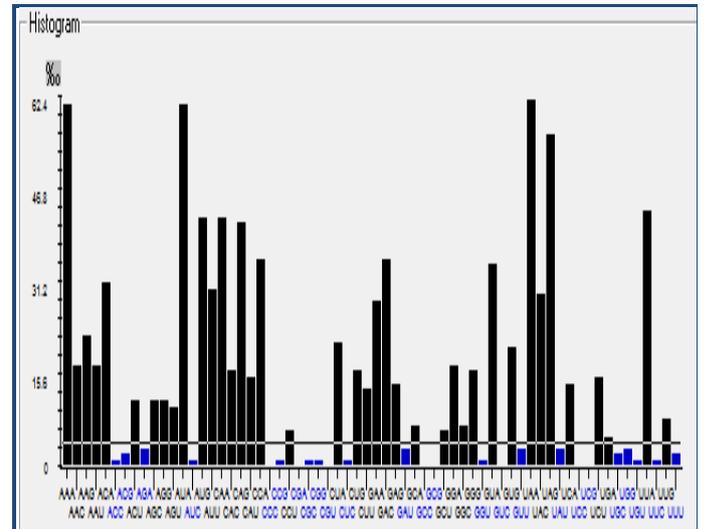
The Relative Synonymous Codon Usage (RSCU) values were calculated for the dataset. The RSCU statistics is calculated by dividing the observed usage of a codon by that expected if all codons were used equally frequently. Thus an RSCU of 1 indicates a codon is used as expected by random usage, RSCU > 1 indicates a codon used more frequently than expected randomly, and RSCU < 1 indicates a codon used less frequently than random. RSCU analysis was conducted using Mega 6 software [15]. Rare codon was computed by the improved implementation in DAMBE [17].

### Results:

#### Evolutionary analysis

The nucleotide frequencies are 45.59% (A), 26.29% (T/U), 10.41% (C), and 17.71% (G). The transition/transversion rate

ratios are  $k1 = 1.729$  (purines) and  $k2 = 8.445$  (pyrimidines). The overall transition/transversion bias is  $R = 1.596$ . The analysis involved 189 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair. There were a total of 82 positions in the final dataset.



**Figure 4:** Graph showing the relative codon frequency of portion of HIV-1 subtypes A *env* gene fragment isolated from HIV-1-infected individuals in Côte d'Ivoire. The blue and black color bar indicating codons that are used less than 5% and more than 5% respectively.

### Selection analyses

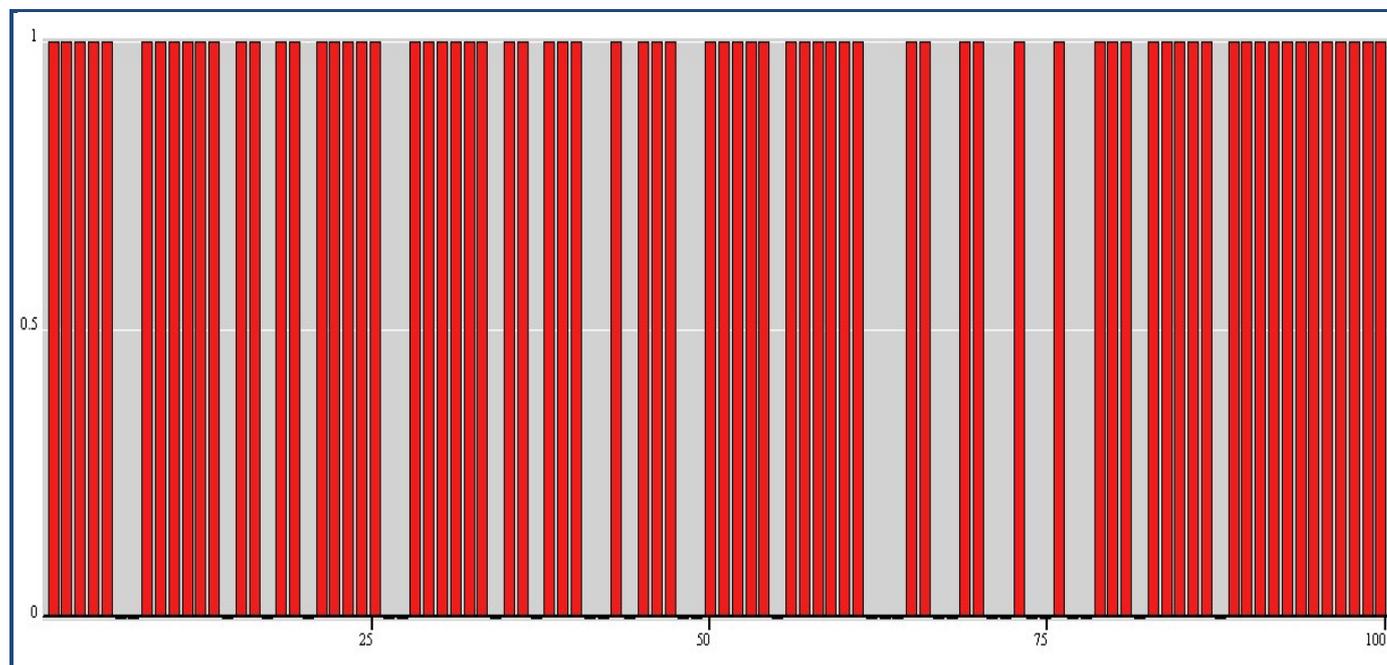
A faster increase of dS was detected with respect to dN through time, and a slowdown of dN/dS in *env* gene due to a slower increase through time of dN with respect to dS **Table 1 (see supplementary material)**. The dN/dS ratio of *env* regions fluctuates was less than 1. For the fragment of *env* region analysed, the dN/dS ratio was lower than 1 for viral populations analysed within each of the 15 years, which showed the lowest levels of divergence (**Figure 1**). A low dN/dS ratio indicates that, the sequences did not undergo adequate immune pressure to lead to changes in amino acids. The dN/dS ratio of the years 1997 and 2008 showed a significant difference ( $P < 0.022$ ) compared to the other years, except for year 2001 and 2006. Both the years 1997 and 2008 showed any significant difference ( $p = 0.215$ ). The synonymous substitution rate was always significantly higher ( $P = 0.001$ , Student's t) than the nonsynonymous substitution rate.

### Phylogenetic tree

In the Maximum Likelihood (ML) tree, viral sequences from the same year or from other West African countries do not form a distinct cluster (**Figure 2**). No significant change in sequence diversity was found after nearly two decades of evolution. The reconstructed phylogenetic tree of 211 sequences demonstrated distinctly long internal branches and short external branches, suggesting that only a small number of viruses infected the new host cell at each transmission so that these founder viruses usually are quite different among hosts. These results are compatible with a severe bottleneck at each new infection. The topology of the tree is notable in that the sequences sampled through different times are evenly

distributed among the terminal branches (**Figure 2**). This suggests that most of these mutations have occurred independently and have not been transmitted for sustained periods of time. Sequences did not cluster according to year or compartment. Using this model of evolution, the neighbor-joining tree for the entire data set shows that sequences cluster

predominantly by host individual. Furthermore, no sequences clustered strongly (bootstrap values 50) with known laboratory strains of HIV subtype B, an indication of no evidence for recombination.



**Figure 5:** Sliding-window analysis of the cumulative dN/dS across Bayes factor for the event of negative selection at a site along the *env* gene fragment isolated from HIV-1-infected individuals in Côte d'Ivoire.

### Phylogenetic Signal and Informativeness

The level of substitution saturation in *env* gene was measured by comparing the number of transitions and transversions with the size of the genetic distance for each pair of sequences (**Figure 3**). The analysis of these sequences showed that the amount of substitutions was increasing with the extension of genetic distance. Only the number of transversions within the *env* gene showed a lower increasing tendency. However, none of the plots took the form of a plateau, typical of the state of saturation with substitutions.

### Evolution of *env*-V3: nucleotide distances

Mean nucleotide distances per year of the *env*-V3 region of the viral genome of strains A1, are shown in **Table 1**. Overall nucleotide distances slowly rise over the years. This rise is mostly accounted for by an increase in synonymous substitutions, while non-synonymous nucleotide distances are more constant throughout the period investigated.

### Codon usage analysis

Relative synonymous codon usage (RCSU) patterns support the attenuation hypothesis as well **Table 2** (see **supplementary material**). Those with the greatest rate of positive change over time were UUU, UUA, CUG, AUA, GUA, GUG, UCU, UCA, CCA, ACU, ACA, GCU, GCC, GCA, UAC, CAU, AAU, GAC, GAA, AGU, AGC, AGA, AGG, GGC, GGG. These changes are due to simple transition mutations, possibly associated with some mutational bias. The variation of relative synonymous codon usage (RSCU) values not only indicated the different frequency of occurrence of each codon for a given amino acid

in different protein but also revealed the preference of either A + U or G + C codon usage as listed in **Table 2**. Preferential codon usage in the portion of *env* gene analysed indicates that the codons with A or U at the third position are more preferred compared to G or C ending codons.

### Codon Based Analysis

To examine how the variation of codon usage pattern over time reflects in the usage of individual codons in HIV-1 *env* genes in Côte d'Ivoire, the normalized frequency of each codon in each sequence was compared between the years 1995 and 2009. A graph of codon frequency distribution was plotted to identify the quantities of rare codons present in each sequence (**Figure 4**). Frequency of codon usage with a value of 100 indicates that the codons are highly used for a given amino acid. Conversely, the frequency of codon usage with a value of less than 5 is determined as low-frequency codon (blue bars) which is likely to affect the expression efficiency. Low-frequency codon are ACG, ACC, AUC, AGA, CCG, CGA, CCC, CGG, CGC, CGU, CUC, GAU, GCG, GGU, GUC, GUU, UAU, UCC, UCG, UGC, UGG, UGU, UUC, and UUG, respectively. This result suggested that the *env* gene analysed contain a large number of rare codons that may reduce translational efficiency of the gene. We detected fewer nonsynonymous substitutions than expected by chance and  $dN/dS < 1$ . Taken together, these results indicate that these regions are subject to very strong purifying selection. The location of the midpoints of the window showing negative selection is given in **Figure 5**. On average, we detected 68% of codon sites under negative selection and 28% neutral sites.

## Discussion:

In this study, we contrast the changes in genetic diversity and adaptive evolution of the HIV-1 *env* gene between samples collected during fifteen years. Since HIV-1 is an obligate pathogen on human for replication and assembly, codon usage bias, that affects the translational efficiency, is likely to be subjected to host selection pressure [18]. Thus, codon usage bias can play a significant role in host adaptation of HIV-1. For Côte d'Ivoire, no study has examined the *env* genes at the global scale over a long time period to address this issue.

## Phylogenetic tree

The *env* sequences analysed indicated that substitution saturation has not been reached, so that the data can be expected to provide reliable phylogenetic signal. No significant change in between-year-sequence diversity was found after more than a decade of evolution. The reconstructed phylogenetic tree of 211 sequences demonstrated distinctly short internal branches and long external branches, suggesting that a large number of viruses infected the new host cell at each transmission. Moreover, the viruses that successfully infected new host cells are not under strong selective pressure from the host immune system, which does not limited between-host diversification, as indicated by those large clusters on the tree. The selective pressure does not significantly vary between the early and the recent samples. These samples seem most likely infected with the virus representing the transmissions between the populations with different genetic backgrounds. This is supported by the intermixing of Ivorian strains to these isolated from other African countries (Senegal, Mali, RDC). Since, Côte d'Ivoire has for decades been the most important destination for migrants in West Africa, the exchange of HIV-1 gene pools between and the populations of Côte d'Ivoire and those of the neighboring countries may increasingly affect the diversity of HIV-1 gene pools in Côte d'Ivoire. The data indicate a maximum intragenotypic subtype A distance of 7.3%, lower than these reported by Janssens *et al.* [19] who observed a maximum intragenotypic subtype A distance of 14.1% in their limited number of samples collected during 1990±1991 in Abidjan. It is likely that the intragenotypic distance obtained by these authors is skewed on the basis of so few years analysed. The low intragenotypic distance obtained by our data is supported by a large number of silent (synonymous) mutations that cause no change in the amino acid sequence. The virus sequences were remarkably well conserved at the amino acid level, both within and among different individuals.

## Selective pressure

A low dN/dS ratio indicates that, the sequences did not undergo adequate immune pressure to lead to changes in amino acids and hence a reflection of the lower variability in the *env* gene when compared. The analyses indicate that the *env* sequences analysed are subject to purifying selection overall and that the derived proteins are not subject to positive selection favoring diversity at the amino acid level but actually tend to be conserved evolutionarily. Since the continuity of the various patients analysed are not known over time, the changes described may not reflect immune pressure. Indeed, purifying immune selection dominates evolution of HIV within hosts, but evolution between hosts is largely decoupled from within-host evolution [20]. The ratio of 0.701 found in our

study is lower than the ratio of 0.90 found by Yamaguchi-Kabata and Gojobori [21], and higher than that of 0.68 reported by Brown & Monaghan [22]. Although we did not analyze the four variable regions where insertions, deletions, and partial duplications might be very frequent, we think that the ratio in this study is realistic. We are aware that changes in the strength of the immune response may not result in predictable changes in the dN/dS ratio if the selection coefficient is on the same order of magnitude as the effective population size and hence providing only a little information about the status of the immune system [23].

## Codon usage bias

Although we are studying changes in codon usage pattern over a decade, the data were not collected throughout the time for each host analysed. Hence, our results may represent outcome of additional and may be even contradictory selective forces (e.g., effect of anti-retroviral therapies). Such a scenario can also give rise to results similar to this study. On studying the codons, Meintjes and Rodrigo [24] found that the early *env* sequences displayed a very biased codon usage pattern, where many codons occurred at very low frequency and the preferred codons were used at a very high frequency.

## Conclusion:

This study is the first that examine the selective pressures that governed the evolution of the subtypes of HIV-1 in Côte d'Ivoire, the most affected country in West Africa. No significant change in the HIV-1 *env* gene sequences diversity was found over one decade of evolution. We detected fewer nonsynonymous substitutions than expected by chance, indicating that the sequences analyzed are subject to very strong purifying selection. In addition to identifying sites under purifying selection, we also identified neutral sites that can cause false positive inference of selection. These sites presented form a resource for future studies of selection pressures acting on HIV-1 *env* gene in Côte d'Ivoire and other West African countries.

## Acknowledgment:

Authors are grateful to Laboratoire de Génétique at UFR Biosciences, Université Félix Houphouët Boigny for funding this research.

## References:

- [1] Toure S *et al.* *AIDS*. 2008 **22**: 873 [PMID: 18427206]
- [2] Toni T *et al.* *AIDS Res Hum Retroviruses*. 2005 **21**: 667 [PMID: 16060839]
- [3] Shioda T *et al.* *Nature*. 1991 **349**: 167 [PMID: 1986308]
- [4] Moore JP *et al.* *Philos Trans R Soc Lond B Biol Sci*. 1993 **342**: 83 [PMID: 7904352]
- [5] Wyatt R & Sodroski J, *Science*. 1998 **280**: 1884 [PMID: 9632381]
- [6] Lemey P *et al.* *AIDS Rev*. 2006 **8**: 125 [PMID: 17078483]
- [7] Seo TK *et al.* *Mol Biol Evol*. 2004 **21**: 1201 [PMID: 15014159]
- [8] Ganeshan S *et al.* *J Virol*. 1997 **71**: 663 [PMID: 8985398]
- [9] Frost SD *et al.* *Proc Natl Acad Sci U S A*. 2005 **102**: 18514 [PMID: 16339909]
- [10] Williamson S *et al.* *Mol Biol Evol*. 2005 **22**: 456 [PMID: 15509726]
- [11] Thompson JD *et al.* *Nucleic Acids Res*. 1997 **25**: 4876 [PMID: 9396791]

- [12] Darriba D *et al.* *Nat Methods*. 2012 **9**: 772 [PMID: 22847109]  
[13] Drummond AJ *et al.* *Mol Biol Evol*. 2012 **29**: 1969 [PMID: 22367748]  
[14] Rambaut A (2008) FigTree v1.1.1: Tree figure drawing tool.  
Available:<http://tree.bio.ed.ac.uk/software/figtree/>.  
[15] Tamura K *et al.* *Mol Biol Evol*. 2013 **30**: 2725 [PMID: 24132122]  
[16] Pond SK & Muse SV, *Mol Biol Evol*. 2005 **22**: 2375 [PMID: 16107593]  
[17] Xia X, *Mol Biol Evol*. 2013 **30**: 1720 [PMID: 23564938]  
[18] Pandit A & Sinha S, *PLoS ONE*. 2011 **6**: e28889 [PMID: 22216135]  
[19] Janssens W *et al.* *AIDS*. 1994 **8**: 21 [PMID: 8011235]  
[20] Keele BF, *Curr Opin HIV AIDS*. 2010 **5**: 327 [PMID: 20543609]  
[21] Yamaguchi-Kabata Y & Gojobori T, *J Virol*. 2000 **74**: 4335 [PMID: 10756049]  
[22] Brown AL & Monaghan P, *AIDS Res Hum Retroviruses*. 1988 **4**: 399 [PMID: 3219234]  
[23] Nielsen R & Yang Z, *Genetics* 1998 **148**: 929 [PMID: 9539414]  
[24] Meintjes PL & Rodrigo AG, *J Bioinform Comput Biol*. 2005 **3**: 157 [PMID: 15751118]

Edited by P Kanguane

Citation: Gonedelé Bi *et al.* *Bioinformatics* 10(11): 671-678 (2014)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

## Supplementary material:

**Table 1:** Estimates of Average Codon-based Evolutionary Divergence over HIV-1 *env* gene Sequence Pairs within year in Côte d'Ivoire

Year	N	d	dN	dS
1995	8	0.039	0.113	0.171
1997	31	0.065	0.132	0.222
1998	21	0.051	0.139	0.205
1999	20	0.036	0.165	0.234
2000	13	0.041	0.168	0.212
2001	5	0.043	0.089	0.112
2002	13	0.035	0.162	0.230
2003	14	0.073	0.153	0.224
2004	15	0.025	0.157	0.241
2005	24	0.046	0.157	0.218
2006	3	0.023	0.180	0.209
2008	8	0.065	0.169	0.301
2009	13	0.042	0.172	0.245

The number of synonymous substitutions per synonymous site from averaging over all sequence pairs within each group is shown. The analysis involved 188 nucleotide sequences. All ambiguous positions were removed for each sequence pair. Evolutionary analyses were conducted in MEGA 6 [15]. Mean nucleotide distance were estimated by Maximum Likelihood using the model GTR + I for the entire set HIV-1 envelope glycoprotein V3 loop gene region.

**Table 2:** Synonymous codons usage pattern in the coding sequences. Optimal codons were identified based on differences in relative synonymous codon usage. Frequency per thousand bases: use frequency per thousand bases in identified high-confidence coding sequences from partial HIV-1 *env* gene ( $P < 0.05$ ).

Codon	Count	RSCU									
UUU(F)	1.5	1.89	UCU(S)	2.9	2.06	UAU(Y)	0.2	0.34	UGU(C)	0.1	0.41
UUC(F)	0.1	0.11	UCC(S)	0	0.01	UAC(Y)	1.2	1.66	UGC(C)	0.2	1.59
UUA(L)	3.2	2.46	UCA(S)	1.5	1.04	UAA(*)	4.3	1.62	UGA(*)	0.5	0.2
UUG(L)	0.7	0.58	UCG(S)	0	0.01	UAG(*)	3.2	1.19	UGG(W)	1.3	1
CUU(L)	0.2	0.17	CCU(P)	0.6	1.24	CAU(H)	2.2	1.21	CGU(R)	0	0.07
CUC(L)	0.1	0.11	CCC(P)	0	0.07	CAC(H)	1.4	0.79	CGC(R)	0	0.13
CUA(L)	1.7	1.33	CCA(P)	1.2	2.58	CAA(Q)	3.6	1.02	CGA(R)	0.1	0.18
CUG(L)	1.7	1.34	CCG(P)	0	0.1	CAG(Q)	3.5	0.98	CGG(R)	0	0.09
AUU(I)	3.9	1.3	ACU(T)	1.7	0.91	AAU(N)	4.9	1.38	AGU(S)	2.6	1.83
AUC(I)	1	0.34	ACC(T)	0.4	0.21	AAC(N)	2.2	0.62	AGC(S)	1.5	1.04
AUA(I)	4.1	1.36	ACA(T)	5.2	2.75	AAA(K)	3.7	1.35	AGA(R)	1.5	4.38
AUG(M)	2.4	1	ACG(T)	0.2	0.13	AAG(K)	1.8	0.65	AGG(R)	0.4	1.14
GUU(V)	0.6	0.35	GCU(A)	0.3	0.37	GAU(D)	0.2	0.13	GGU(G)	1	0.65
GUC(V)	0.1	0.05	GCC(A)	0.8	1.08	GAC(D)	3.4	1.87	GGC(G)	1.7	1.13
GUA(V)	3.5	2.15	GCA(A)	1.8	2.34	GAA(E)	4.1	1.24	GGA(G)	1.2	0.77
GUG(V)	2.4	1.45	GCG(A)	0.2	0.22	GAG(E)	2.5	0.76	GGG(G)	2.2	1.45