

Full Length Research Paper

Dearth of full-length HIV-1 sequences obscures the true HIV-1 genetic subtypes distribution in sub-Saharan Africa

Kemebradikumo Pondei*, Abdulrasheed Abdu and Langley Orutugu

Department of Medical Microbiology and Parasitology, Faculty of Basic Medical Sciences, College of Health Sciences, Niger Delta University, Amassoma, Wilberforce Island, Bayelsa State, Nigeria.

Received 9 January 2014, Accepted 8 May, 2014

HIV infection is still a public health problem in sub-Saharan Africa. The broad diversity exhibited by HIV-1 may impact on transmission, disease progression, drug resistance and vaccine development. Most analyses of HIV-1 subtype distribution have been on partial *HIV-1* gene sequences, which may not adequately reflect the circulating subtypes. The objective of this study was to estimate the HIV-1 subtype distribution in sub-Saharan Africa using only full-length genome sequences. Using available HIV-1 full-length genome sequences from sub-Saharan Africa, the HIV-1 distribution in the region was analysed and compared with a previous global analysis which was not based entirely on full-length sequences. A total of 934 HIV-1 full-length genome sequences were available from 27 sub-Saharan countries. There was a disproportionate distribution of HIV-1 subtypes among countries with Cameroon having all the four HIV-1 groups. The subtype C was the most available in addition to a large proportion of circulating and unique recombinant forms (CRFs/URFs) especially in Central and West African countries, with frequencies of 32.6 to 90%. There was decreased representation of subtypes A and G in regions where CRFs/URFs were common compared with previous analysis using partial sequences. There is a need for more HIV-1 full-length genome sequences from sub-Saharan Africa for the true distribution of HIV-1 subtypes to be known, as analysis of partial sequences is not truly representative of the circulating subtypes.

Key words: Africa, distribution, genetic diversity, HIV sequence variability, subtypes, recombination.

INTRODUCTION

Infection with the Human Immunodeficiency Virus (HIV) continues to be a global public health problem with devastating consequences in developing countries especially in sub-Saharan Africa, even though antiretroviral therapy has improved the quality of life of

those infected. HIV exists in two genetically distinct forms (HIV-1 and HIV-2), with HIV-2 being restricted to West Africa and HIV-1 having a global spread and being responsible for the HIV pandemic.

HIV-1 exhibits genetic diversity in the form of viral

*Corresponding authors. E-mail: kemepondei@hotmail.com. Tel: +2348030940882.

quasispecies (Meyerhans et al., 1989) described as a heterogeneous viral population of related genomes (Domingo et al., 1997). This genetic diversity of HIV-1 is believed to result from a high mutation rate due to the infidelity or error-prone characteristic of reverse transcriptase during replication (Roberts et al., 1988; Boyer et al., 1992), a high replication rate of about 10^9 virions per day (Ho et al., 1995) and genomic recombination (Hu and Temin, 1990; Jetzt et al., 2000; Zhuang et al., 2002).

HIV-1 is classified into three main genetic groupings each representing independent cross-species transmission, although a fourth group (Group P) (Plantier et al., 2009) has been suggested. The three groups are: Group M (major), Group O (outlier) and Group N (non-M, non-O). Groups O and N are mainly restricted to Cameroon and the Democratic Republic of Congo. The Group M has a global distribution and is further divided into nine subtypes and some sub-subtypes: subtypes A, B, C, D, F, G, H, J and K, with sub-subtypes A1 to A5 (Gao et al., 2001; Meloni et al., 2004; Vidal et al., 2006, 2009), and F1 and F2 (Triques et al., 1999). Combinations of two or more subtypes and/or sub-subtypes exist, and when these mosaic forms become widely spread and fixed in the population, they are known as circulating recombinant forms (CRFs). A CRF is therefore defined as "intersubtype recombination for which at least three epidemiologically unlinked variants are monophyletic and share identical genetic structure along their full genomes" (Yebra et al., 2012), while a unique recombinant form (URF) is a variant that has not been isolated from three or more individuals.

Presently, about 58 CRFs have been characterized and are in the public domain (<http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>). HIV-1 subtypes may exhibit phenotypic differences. Subtypes are believed to impact on tropism, with some studies associating increased CXCR4 usage with infections with subtype C (Johnston et al., 2003; Connell et al., 2008), and others finding decreased CXCR4 usage in subtype C infections (Bjorndal et al., 1997; Abebe et al., 1999; Peeters et al., 1999; Esbjornsson et al., 2010). Subtypes may have an important effect on transmission of HIV-1, as the subtype B was associated with homosexual transmission and the subtype C with heterosexual transmission (van Harmelen et al., 1997; van Harmelen et al., 2001), but a heterosexually driven subtype B epidemic has been observed in Trinidad and Tobago (Cleghorn et al., 2000). Also, the subtype C is believed to be more likely to be transmitted vertically than the subtypes A and D (Blackard et al., 2001; Renjifo et al., 2001; Renjifo et al., 2003); and infection with the subtype D has been associated with faster CD4 T cell decline and a faster rate of disease progression (Alaeus et al., 1999; Kanki et al., 1999; Kaleebu et al., 2001; Vasan et al., 2006; Baeten et al., 2007; Kaleebu et al., 2007; Easterbrook et

al., 2010). Subtypes could also be important in vaccine design and development (Hemelaar et al., 2011).

Comprehensive understanding of disease process and effective interventions could arise from correlating genomic profile of HIV with that of patients (Sampathkumar et al., 2012). With HIV subtypes playing important roles in transmission and outcome of disease, it becomes imperative for the actual distribution of subtypes to be known. HIV subtype distribution has largely been determined using partial genome sequences. Since recombinant forms are relatively common, there is the possibility that the actual subtype distribution is not accurately represented using these partial genome sequences, and recombinants could be artificially scored as pure subtypes. This analysis therefore sought to firstly determine the HIV-1 subtype distribution in sub-Saharan Africa using only full-length or near full-length HIV-1 genome sequences available in the public domain. Secondly, it intended to document available full-length HIV-1 sequences from the region.

MATERIALS AND METHODS

All available full-length or near full-length HIV-1 genome sequences of sub-Saharan African origin were obtained from GenBank (<http://www.ncbi.nlm.nih.gov>) and the Los Alamos HIV Sequence Database (<http://www.hiv.lanl.gov>). The full-length sequence is one that contains the entire protein coding region as well as the non-coding regions, while a near full-length sequence contains almost all of the coding region.

Duplications were rectified with only one sequence per patient, except for those with infection with viruses of different subtypes. Information was transferred to an Excel spreadsheet and analysed. The countries in sub-Saharan Africa were grouped into four regions using the United Nations geoscheme for Africa: Central Africa, East Africa, Southern Africa and West Africa.

For convenience, Malawi, Zambia and Zimbabwe were grouped under Southern Africa.

Central Africa includes Angola, Cameroon, Central African Republic (CAR), Chad, Democratic Republic of Congo (DRC), Equatorial Guinea, Gabon, Republic of Congo and Sao Tome and Principe.

East Africa includes Burundi, Djibouti, Ethiopia, Eritrea, Kenya, Madagascar, Mauritius, Mozambique, Rwanda, Somalia, Sudan, Tanzania and Uganda.

Southern Africa includes Botswana, Lesotho, Malawi, Namibia, South Africa, Swaziland, Zambia and Zimbabwe.

West Africa includes Benin, Burkina Faso, Cape Verde, Gambia, Ghana, Guinea, Guinea-Bissau, Ivory Coast, Liberia, Mali, Mauritania, Niger, Nigeria, Senegal, Sierra Leone and Togo.

Results from this analysis were then compared with those obtained in an earlier analysis which did not discriminate between partial and full-length HIV genome sequences (Hemelaar et al., 2006).

RESULTS

A total of 1084 full-length or near full-length HIV-1 genome sequences of sub-Saharan Africa origin, were retrieved from GenBank (<http://www.ncbi.nlm.nih.gov>)

and the Los Alamos HIV Sequence Database (<http://www.hiv.lanl.gov>). The sequences had been submitted to the databases between 1983 and 2011. In contrast, there were 127,798 partial HIV-1 sequences from the region.

With the removal of duplications, there were 934 unique full-length HIV-1 sequences from patients within the sub-Saharan Africa region. Sequences were available from 27 of the 45 sub-Saharan countries. An overwhelming majority of the sequences (96.7%) were of the HIV-1 Group M, with Groups N and O accounting for 1.1 and 2.1%, respectively of the sequences (Table 1). The subtype C was the most common subtype accounting for almost half of the Group M sequences, with recombinants responsible for 29% of the Group M sequences. A global HIV-1 subtype analysis was done in 2004 (Hemelaar et al., 2006) and some of the results are compared with the present analysis (Table 1).

Regional analysis

Central Africa

There were available sequences from five of nine countries. Over three-quarters of the sequences were from Cameroon. Recombinant forms made up 64.2% of the sequences (Table 2). All the HIV groups (M, N, O and P) were present among the Cameroonian sequences, and also six of the nine Group M subtypes. Recombinants made up 66.7% of the sequences from Cameroon. The Group O was present among sequences from Gabon, while recombinants accounted for 66.7% of the sequences from the Democratic Republic of Congo.

East Africa

Sequences were retrieved for seven out of 13 countries in the region. Most of the sequences (94.4%) were from three countries: Kenya, Uganda and Tanzania. The subtype distribution of sequences from East Africa was spread between recombinants (37.2%), subtype A (32%), subtype D (18.4%) and subtype C (11.2%). About half of the sequences from Kenya were of the subtype A, with recombinants making up 41.8%. Of the Ugandan sequences, 44.2 and 32.6% were subtype D and recombinants respectively, while subtype A accounted for 23.2%. Of the sequences from Tanzania, 44% each were recombinants and the subtype C.

Southern Africa

Sequences were available for four out of eight countries in the region. Most of the sequences were from South Africa (78.4%) and 97.6% of the sequences from South

Africa were of the subtype C, with recombinants contributing just 1.2% of the sequences. All the sequences from Malawi and 94.7% of those from Zambia were also of the subtype C.

West Africa

Sequences were obtained for 10 out of 16 countries in the region. Over 75% of the sequences were from just three countries: Nigeria, Ghana and Senegal. In this region, 71.83% of the sequences were recombinants. The HIV-1 Group O was surprisingly present among the sequences from Senegal (Table 2). Among the Nigerian sequences, subtype G and recombinants were each responsible for 47.6% of the sequences. Recombinants were responsible for 90% of the sequences from Ghana and 46.1% of those from Senegal.

Circulating and unique recombinant forms (CRFs/URFs)

Recombinants accounted for a substantial proportion of the available sequences from certain countries in the West, Central and Eastern African regions (Table 3) ranging from 32.6 to 90%. Five countries had no recombinants (Chad, Djibouti, Ethiopia, Somalia and Malawi). Twenty-one out of the characterized 58 CRFs were present among the sequences. In terms of proportion, Ghana had the highest proportion of recombinant sequences, but Cameroon numerically had more CRFs/URFs.

Cameroon also had the greatest diversity of recombinants with nine different CRFs and 23 different URF types, followed by Kenya with 11 URF types and 2 CRFs; Democratic Republic of Congo (4 CRFs, 5 URFs) and Ghana (3 CRFs, 5 URFs) (Table 4). CRF02_AG was the most prevalent recombinant form accounting for 47.6% of all CRFs and 21.8% of all recombinants. It was common in Cameroon, Nigeria and Ghana. An inter-group recombinant, 02O was among the Cameroonian sequences.

All the countries in the Southern African region had no CRFs but a few URFs, while the countries in East Africa had more URFs than CRFs. In contrast, there were more CRFs than URFs in the countries in Central and West Africa (Table 4). Also, more recombinants were detected in our analysis compared to earlier analysis using partial sequences.

Table 5 shows regional differences in the distribution of some subtypes when partial and full-length sequences were used in the analysis. The detection rate of the subtypes A and G were less when full-length genome sequences are used in analysis compared with partial sequences, especially in regions where recombinants are common.

Table 1. Frequency of the HIV-1 groups and subtypes using full-length genome sequences. Figures from an earlier study (Hemelaar et. al are indicated).

HIV-1 variant	Number of full-length sequences	% of total HIV-1 subtypes	Hemelaar et al. (2006) (%)
Group M	903	96.7	
Subtype A	89	9.5	14%
Subtype B	3	0.3	N/S
Subtype C	446	47.8	56%
Subtype D	60	6.4	N/S
Subtype F	7	0.7	N/S
Subtype G	23	2.5	10%
Subtype H	1	0.1	N/S
Subtype J	1	0.1	N/S
Subtype K	2	0.2	N/S
Recombinants	271	29	16%
(CRF)	(124)	(13.3)	N/S
(URF)	(147)	(15.7)	N/S
Group N	10	1.1	
Group O	20	2.1	
Group P	1	0.1	
TOTAL	934	100	

N/S - Not stated.

DISCUSSION

Despite Africa bearing the brunt of HIV infection, there is limited information on the molecular epidemiology of HIV-1 due to the paucity and uneven availability of both partial and full-length genome sequences across the continent. Results from previous analyses (Hemelaar et al., 2006), had shown that the subtype A accounted for 21% of HIV-1 infections in West Africa, but in our analysis, it represented only 4.2% of sequences. Also, the subtype G previously observed to represent 35% of infections, accounted for only 16.9% of sequences. Though the subtype A was projected to represent 29% of infections in Nigeria, there were no available full-length or near full-length subtype A sequences from Nigeria. These analysed partial sequences in their survey, might really be part of CRF02_AG recombinants, but on their own appear as subtypes A or G. The earlier analysis had used sequences irrespective of the length, but weighted the distribution according to the number of HIV-infected people in each country. Disparities in subtype frequencies between our analysis and the Hemelaar study were also observed in some countries from the different sub-regions. This implies that the true HIV-1 subtype distribution might not have been captured using partial sequences.

Our analysis shows that recombinants (CRFs/URFs) constituted a substantial proportion of HIV-1 genotypes in sub-Saharan Africa. Our estimates of CRFs were higher than those obtained in a comprehensive audit of HIV distribution in 2004 (Hemelaar et al., 2006).

In that analysis, recombinants were projected to account for 42.6% of HIV-1 infections for West Africa, while we observed 71.8% for the same region. The differences in estimates might be due to the different approaches used, and also the timing because the Hemelaar study evaluated 2003/2004, while we analysed all available full-length sequences at the time of analysis. Our analysis sought to present the subtype distribution as based on full-length HIV-1 sequences available in the Los Alamos database. Whilst not a perfect approach, it presents the genetic diversity as determined by the unambiguity of full-length genome sequences.

There has been a consistent increase in the reporting of CRFs and URFs (Vidal et al., 2000; Nyombi et al., 2008). This is further buttressed by the fact that 21 new CRFs have been characterized between 2008 and 2013, implying that the already complex genetic diversity of HIV-1 is evolving further.

The increasing number of CRFs and their relative spread is also a reason for more full-length sequencing and analysis. This is important because the clinical implications of subtype variation with regards to recombinants are yet to be established. The spatial distribution of CRFs and URFs needs clarification as CRFs were common in Central and West Africa, while URFs were common in East Africa.

Our analysis further reveals the dearth of HIV-1 sequence information from sub-Saharan Africa, as there were only 934 full-length sequences from 27 countries having millions of people living with HIV. Hemelaar et al. (2011) in a later review had noted that the available

Table 2. Regional distribution of HIV-1 subtypes using full-length genome sequences.

Region	Country	*People living with HIV 2012	HIV* prevalence	GROUP M SUBTYPES										Group N	Group O	Group P	Total	
				A	B	C	D	F	G	H	J	K	CRF/URF					
Central Africa	Angola	250 000	2.3			1		1										3
	Cameroon	600 000	4.5	1			4	6	8		1	1	96	10	16	1		144
	CAR	N/A	N/A								1		3					4
	Chad	210 000	2.7				4											4
	DRC	480 000	1.1	4			4						1	18				27
	Gabon	41 000	4		1									2			2	5
East Africa	Djibouti	9 700	1.2			1												1
	Ethiopia	760 000	1.3			2												2
	Kenya	1 600 000	6.1	45		2	3		3					38				91
	Rwanda	240 000	2.9	8										2				10
	Somalia	29 000	0.5			1												1
	Tanzania	1 500 000	5.1	5		22	1							22				50
	Uganda	1 500 000	7.2	22				42						31				95
Southern Africa	Botswana	340 000	23			51							2					53
	Malawi	1 100 000	10.1			20												20
	South Africa	6 100 000	17.9	1	2	326	1						4					334
	Zambia	1 100 000	12.7			18								1				19
West Africa	Benin	72 000	1.1										1					1
	Gambia	14 000	1.3										3					3
	Ghana	240 000	1.4						2				18					20
	Guinea Bissau	41 000	3.9										2					2
	Ivory Coast	450 000	3.2										6					6
	Liberia	22 000	0.9										1					1
	Mali	100 000	0.9										1					1
	Niger	46 000	0.5										3					3
	Nigeria	3 400 000	3.1			1			10				10					21
	Senegal	43 000	0.5	3		1	1						6			2		13
Total		20 287 700		89	3	446	60	7	23	1	1	2	271	10	20	1	934	

* UNAIDS Report on the Global AIDS epidemic 2013; DRC, Democratic Republic of Congo; CAR, Central African Republic; N/A Not available.

Table 3. Distribution of circulating and unique recombinant forms.

Region	Country	Circulating recombinant forms (CRFs)																			Unique Recombinant Forms (URFs)			
		1	2	6	9	10	11	13	16	18	21	22	25	26	27	30	32	36	37	45	49	Total	(Number of URFs indicated in brackets)	Total
Central Africa	Angola	1																			1	0222 (6), 0213 (1), 0225 (1), 0102A (1), 0122F (1), 01ADF2 (1), 01A1G (1), 01F2 (1), 01F2G (1), 02A1U (1), 02B (1), 02F2G (1), 02G (1), 02O (1), A1F2 (1), A1G (1), A1H (1), AF2 (1), AF2G (1), AGU (1), AHJU (1), 22A1U (1), 22DU(1)	0	
	Cameroon	36					7	6		1		10	3				2	2		1	68		28	
	CAR	3																			3		0	
	DRC			1										4	2						2	9	26C (4), U (2), AGKU (1), A2G (1), 26CU (1)	9
	Gabon																				1	1	A1GHU (1)	1
East Africa	Kenya								1		3										4	A1D (9), A1C (7), CD (4), A1CD (3), A1A2D (3), A2CD (2), A1CG (2), A1G (1), A1A2CD (1), AC (1), DG (1)	34	
	Rwanda																				0	A1C (1), CD(1)	2	
	Tanzania					3															3	A1C (12), A1D (4), CD (2), ACD (1)	19	
	Uganda																				0	A1D (26), A1C (2), A1CD (1), A1G (1), CD (1)	31	
Southern Africa	Botswana																				0	CD (1), A1GJ (1)	2	
	South Africa																				0	A1CDGKU (1), A2C (1), AC (1), A1C (1)	4	
	Zambia																				0	A2C (1)	1	
West Africa	Benin															1					1		0	
	Gambia																				3	3	0	
	Ghana		9	1	1																	11	02D (2), 02A1 (2), 02A (1), 02G (1), 0206 (1)	7
	Guinea Bissau		2																			2		0
	Ivory Coast				1																	1	0209 (4), 02GK (1)	5
	Liberia		1																			1		0
	Mali			1																		1		0
	Niger															3						3		0
	Nigeria		8																			8	02A1 (1), A1G (1)	2
Senegal		2	1	1																	4	02A1 (2)	2	
Total		3	59	4	3	3	7	6	1	1	3	10	3	4	2	3	1	2	2	4	3	124		147

1, CRF01_AE; 2, CRF02_AG; 6, CRF06_cpx; 9, CRF09_cpx; 10, CRF10_CD; 11, CRF11_cpx; 13, CRF13_cpx; 16, CRF16_A2D; 18, CRF18_cpx; 21, CRF21_A2D; 22, CRF22_01A1; 25, CRF25_cpx; 26, CRF26_AU; 27, CRF27_cpx; 30, CRF30_0206; 32, CRF32_06A1; 36, CRF36_cpx; 37, CRF37_cpx; 45, CRF45_cpx; 49, CRF49_cpx.

Table 4. Frequency of recombinant forms in some countries.

Country	% Recombinants	% CRFs	% URFs
Cameroon	66.7	47.2	19.4
DRC	66.7	33.3	33.3
Ghana	90	55	35
Kenya	41.8	4.4	37.4
Nigeria	47.6	38.1	9.5
Tanzania	44	6	38
Uganda	32.6	0	32.6

Table 5. Regional distribution of Recombinants. The proportion of regional sequences that are recombinants in this analysis and the Hemelaar et. Al. (2006) study are indicated.

Region	Country	CRFs	URFs	Total recombinants	Hemelaar et al. (2006) (%)	
Central Africa	Angola	1	0	1	64.2%	18.6
	Cameroon	68	28	96		
	CAR	3	0	3		
	DRC	9	9	18		
	Gabon	1	1	2		
	Regional Total	82	38	120		
East Africa	Kenya	4	34	38	37.2%	28.6
	Rwanda	0	2	2		
	Tanzania	3	19	22		
	Uganda	0	31	31		
	Regional Total	7	86	93		
Southern Africa	Botswana	0	2	2	1.2%	0.25
	South Africa	0	4	4		
	Zambia	0	1	1		
	Regional Total	0	7	7		
West Africa	Benin	1	0	1	71.8%	42.6
	Gambia	3	0	3		
	Ghana	11	7	18		
	Guinea Bissau	2	0	2		
	Ivory Coast	1	5	6		
	Liberia	1	0	1		
	Mali	1	0	1		
	Niger	3	0	3		
	Nigeria	8	2	10		
	Senegal	4	2	6		
Regional Total	35	16	51			
Total		124	147	271		

sequences were not representative of the HIV-1 distribution in the countries of origin, and that some countries harbouring large numbers of infected individuals with high subtype diversity had a small amount of HIV data. This is particularly true of a country

like Nigeria that has a relatively high HIV-1 burden, but has only 21 available full-length HIV-1 sequences.

In a review of selected studies that have documented HIV subtype diversity in East, West and Southern Africa, only two studies were observed to have used near full-

length and full-length genome sequences (Lihana et al., 2012). There is thus a need for more full-length sequencing. Cost and lack of requisite equipment and manpower are probably responsible for the gross under-representation of sub-Saharan HIV-1 full-length sequences. Worrisome is the fact that of over 2148 high-throughput sequencing machines in the world, there are only 17 in Africa (3 in Kenya, 14 in South Africa) (<http://omicsmaps.com/>). Due to cost or other limitations, most studies in Africa are limited to sequencing of partial HIV genomes, and even these studies are identifying recombination within partial gene sequences (Kiwelu et al., 2013), so the full extent of genetic variability and recombination could be obscured unless the full-length genome is sequenced and analysed. Studies using partial sequences, have shown an increase in the detection of URFs and drug resistant viruses in sub-Saharan Africa (Ragupathy et al., 2011; Jacobs et al., 2014). Analysis of full-length sequences could possibly lead to greater identification of these recombinant forms.

Analysis of full-length sequences could also help in the accurate identification of low frequency viral variants (Henn et al., 2012) and the use of multiple genes rather than single gene to identify HIV-1 subtypes can reduce the chances of false identification (Neogi et al., 2012). The fact that the full-length subtypes E and I isolates were never found, and have now been re-designated as circulating recombinant forms, CRF01_AE and CRF04_cpx respectively (Carr et al., 1996; Gao et al., 1998; Paraskevis et al., 2001), and the suggestion that the subtype G, was actually a recombinant, whose parental subtype included the CRF02_AG (Abecasis et al., 2007), justify the calls for subtype classification to be based only on analysis of full-length or near full-length genomes.

It follows that analysis of partial HIV-1 sequences could be misinterpreted and may not reveal the true picture of HIV-1 biology and pathogenesis. Therefore, there is the need to know the current incidence/distribution of HIV-1 and also the need to expand the subtype database as these may impact on diagnosis, therapy and vaccine design. Full-length sequences are probably the most accurate representation of HIV genetic diversity.

Conclusion

This analysis brings to light the need for more sequences of full-length genomes from the sub-Saharan Africa region. This is a herculean task because even partial sequences are difficult to come by in most countries in the region. It will require an understanding of the importance of sequencing, commitment from governments within the region and continuous hard work from scientists to achieve this objective. The periodic monitoring of HIV variants could help determine the extent of virus evolution.

Conflict of Interests

The author(s) have not declared any conflict of interests.

REFERENCES

- Abebe A, Demissie D, Goudsmit J, Brouwer M, Kuiken CL, Pollakis G, Schuitemaker H, Fontanet AL, Rinke de Wit TF (1999). HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and coreceptor usage among Ethiopian patients with AIDS. *AIDS* 13:1305-1311.
- Abecasis AB, Lemey P, Vidal N, de Oliveira T, Peeters M, Camacho R et al (2007). Recombination Confounds the Early Evolutionary History of Human Immunodeficiency Virus Type 1: Subtype G Is a Circulating Recombinant Form. *J. Virol.* 81:8543-8551.
- Alaeus A, Lidman K, Bjorkman A, Giesecke J, Albert J (1999). Similar rate of disease progression among individuals infected with HIV-1 genetic subtypes A-D. *AIDS* 13:901-907.
- Baeten JM, Chohan B, Lavreys L, Chohan V, McClelland RS, Certain L et al (2007). HIV-1 Subtype D Infection Is Associated with Faster Disease Progression than Subtype A in Spite of Similar Plasma HIV-1 Loads. *J. Infect. Dis.* 195:1177-1180.
- Bjorndal A, Deng HK, Jansson M, Fiore JR, Colognesi C, Karlsson A et al (1997). Coreceptor usage of primary human immunodeficiency virus type 1 isolates varies according to biological phenotype. *J. Virol.* 71:7478-7487.
- Blackard JT, Renjifo B, Fawzi W, Hertzmark E, Msamanga G, Mwakagile D, Hunter D, Spiegelman D, Sharghi N, Kagoma C, Essex M (2001). HIV-1 LTR Subtype and Perinatal Transmission. *Virology* 287:261-265.
- Boyer JC, Bebenek K, Kunkel TA (1992). Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. *Proc. Natl. Acad. Sci.* 89:6919-6923.
- Carr JK, Salminen MO, Koch C, Gotte D, Artenstein AW, Hegerich PA et al (1996). Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J. Virol.* 70:5935-5943.
- Cleghorn FR, Jack N, Carr JK, Edwards J, Mahabir B, Sill A et al (2000). A distinctive clade B HIV type 1 is heterosexually transmitted in Trinidad and Tobago. *Proc. Natl. Acad. Sci.* 97:10532-10537.
- Connell BJ, Michler K, Capovilla A, Venter, WDF, Stevens WS, Papathanasopoulos MA et al (2008). Emergence of X4 usage among HIV-1 subtype C: evidence for an evolving epidemic in South Africa. *AIDS* 22:896-899
- Domingo E, Menendez-Arias L, Holland JJ (1997). RNA virus fitness. *Rev. Med. Virol.* 7:87-96.
- Easterbrook P, Smith M, Mullen J, O'Shea S, Chrystie I, de Ruiter A et al (2010). Impact of HIV-1 viral subtype on disease progression and response to antiretroviral therapy. *J. Intl. AIDS Soc.* 13:4.
- Esbjornsson J, Mansson F, Martinez-Arias W, Vincic E, Biague A, da Silva Z et al (2010). Frequent CXCR4 tropism of HIV-1 subtype A and CRF02_AG during late-stage disease - indication of an evolving epidemic in West Africa. *Retrovirology* 7:23.
- Gao F, Robertson DL, Carruthers CD, Li YY, Bailes E, Kostrikis LG et al (1998). An isolate of human immunodeficiency virus type 1 originally classified as subtype I represents a complex mosaic comprising three different group M subtypes (A, G, and I). *J. Virol.* 72:10234-10241.
- Gao F, Vidal N, Li YY, Trask SA, Chen YL, Kostrikis LG et al. (2001). Evidence of two distinct subsubtypes within the HIV-1 subtype A radiation. *AIDS Res. Hum. Retroviruses* 17:675-688.
- Hemelaar J, Gouws E, Ghys PD, Osmanov S (2006). Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS* 20: W13-W23.
- Hemelaar J, Gouws E, Ghys PD, Osmanov S, WHO-UNAIDS Network for HIV Isolation and Characterisation (2011). Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS* 25:679-689.
- Henn MR, Boutwell CL, Charlebois P, Lennon NJ, Power KA, Macalalad AR (2012). Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog* 8:e1002529.

- Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*. 373:123.
- Hu W, Temin HM (1990). Genetic consequences of packaging two RNA genomes in one retroviral particle: Pseudodiploidy and high rate of genetic recombination. *Proc. Natl. Acad. Sci.* 87:1556-1560.
- Jacobs GB, Wilkinson E, Isaacs S, Spies G, de Oliveira T, Seedat S et al (2014). HIV-1 Subtypes B and C Unique Recombinant Forms (URFs) and Transmitted Drug Resistance Identified in the Western Cape Province, South Africa. *PLoS ONE* 9:e90845.
- Jetzt AE, Yu H, Klarmann GJ, Ron Y, Preston BD, Dougherty JP (2000). High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J. Virol.* 74:1234-1240.
- Johnston ER, Zijenah LS, Mutetwa S, Kantor R, Kittinunvorakoon C, Katzenstein DA (2003). High frequency of syncytium-inducing and CXCR4-tropic viruses among human immunodeficiency virus type 1 subtype C-infected patients receiving antiretroviral Treatment. *J. Virol.* 77:7682-7688.
- Kaleebu P, Nankya IL, Yirrell DL, Shafer LA, Kyosimire-Lugemwa J, Lule DB et al (2007). Relation Between Chemokine Receptor Use, Disease Stage, and HIV-1 Subtypes A and D: Results From a Rural Ugandan Cohort. *JAIDS* 45:28-33
- Kaleebu P, Ross A, Morgan D, Yirrell D, Oram J, Rutebemberwa A et al (2001). Relationship between HIV-1 Env subtypes A and D and disease progression in a rural Ugandan cohort. *AIDS*. 15:293-299.
- Kanki PJ, Hamel DJ, Sankale JL, Hsieh CC, Thior I, Barin F et al (1999). Human immunodeficiency virus type 1 subtypes differ in disease progression. *J. Infect. Dis.* 179:68-73.
- Kiwelu IE, Novitsky V, Margolin L, Baca J, Manongi R, Sam N et al (2013). Frequent Intra-Subtype Recombination among HIV-1 Circulating in Tanzania. *PLoS ONE* 8:e71131.
- Lihana RW, Ssemwanga D, Abimiku A, Ndembu N (2012). Update on HIV diversity in Africa: A decade in review. *AIDS Rev.* 14:83-100.
- Meloni ST, Kim B, Sankale J-L, Hamel DJ, Tovanabutra S, Mboup S et al (2004). Distinct Human Immunodeficiency Virus Type 1 Subtype A Virus Circulating in West Africa: Sub-Subtype A3. *J. Virol.* 78: 12438-12445.
- Meyerhans A, Cheyner R, Albert J, Seth M, Kwok S, Sninsky J et al (1989). Temporal fluctuations in HIV Quasispecies *in vivo* are not reflected by sequential HIV isolations. *Cell* 58:901-910.
- Neogi U, Gupta S, Shet A, De Costa A, Laishram RL, Wanchu A et al (2012). Emergence of unique recombinant forms (URFs) in Indian HIV-1 epidemic: data from nationwide clinical cohort between 2007 and 2011. *Retrovirology*. 9(Suppl 2):151.
- Nyombi BM, Kristiansen KI, Bjune G, Müller F, Holm-Hansen C (2008). Diversity of Human immunodeficiency virus type 1 subtypes in Kagera and Kilimanjaro regions, Tanzania. *AIDS Res. Hum. Retroviruses*. 24:761-769.
- Paraskevis D, Magiorkinis M, Vandamme AM, Kostrikis LG, Hatzakis A (2001). Re-analysis of human immunodeficiency virus type 1 isolates from Cyprus and Greece, initially designated 'subtype I', reveals a unique complex A/G/H/K/? mosaic pattern. *J. Gen. Virol.* 82:575-580.
- Peeters M, Vincent R, Perret J-L, Lasky M, Patrel D, Liegeois F et al (1999). Evidence for differences in MT2 cell tropism according to genetic subtypes of HIV-1: Syncytium-inducing variants seem rare among subtype C HIV-1 viruses. *J. Acquir. Immune. Defic. Syndr. Hum. Retrovirol.* 20:115-121.
- Plantier J, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemeé V et al (2009). A new human immunodeficiency virus derived from gorillas. *Nat. Med.* 15:871.
- Ragupathy V, Zhao J, Wood O, Tang S, Lee S, Nyambi P et al (2011). Identification of new, emerging HIV-1 unique recombinant forms and drug resistant viruses circulating in Cameroon. *Virol. J.* 8(185).
- Renjifo B, Chung M, Gilbert P, Mwakagile D, Msamanga G, Fawzi W, Essex M (2003). In-utero transmission of quasispecies among human immunodeficiency virus type I genotypes. *Virology* 307:278-282.
- Renjifo B, Fawzi W, Mwakagile D, Hunter D, Msamanga G, Spiegelman D et al (2001). Differences in perinatal transmission among human immunodeficiency virus type 1 genotypes. *J. Hum. Virol.* 4:16-25.
- Roberts JD, Bebenek K, Kunkel TA (1988). The accuracy of reverse transcriptase from HIV-1. *Science* 242:1171-1173.
- Sampathkumar R, Shadabi E, Luo M (2012). Interplay between HIV-1 and Host Genetic Variation: A Snapshot into Its Impact on AIDS and Therapy Response. *Advances Virol.* 2012:508967(doi:10.1155/2012/508967).
- Triques K, Bourgeois A, Saragosti S, Vidal N, Mpoudi-Ngole E, Nzilambi N et al (1999). High diversity of HIV-1 subtype F strains in Central Africa. *Virology* 259:99-109.
- van Harmelen J, Williamson C, Kim B, Morris L, Carr J, Karim SSA, McCutchan F (2001). Characterization of full-length HIV type 1 subtype C sequences from South Africa. *AIDS Res. Hum. Retroviruses* 17:1527-1531.
- van Harmelen J, Wood R, Lambrick M, Rybicki EP, Williamson AL, Williamson C (1997). An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa. *AIDS* 11:81-87.
- Vasan A, Renjifo B, Hertzmark E, Chaplin B, Msamanga G, Essex M et al (2006). Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. *Clin. Infect. Dis.* 42:843-852.
- Vidal N, Bazepeo SE, Mulanga C, Delaporte E, Peeters M (2009). Genetic Characterization of Eight Full-Length HIV Type 1 Genomes from the Democratic Republic of Congo (DRC) Reveal a New Subsubtype, A5, in the A Radiation That Predominates in the Recombinant Structure of CRF26_A5U. *AIDS Res. Human Retroviruses* 25:823-832.
- Vidal N, Mulanga C, Bazepeo SE, Lepira F, Delaporte E, Peeters M et al (2006). Identification and molecular characterization of subsubtype A4 in Central Africa. *AIDS Res. Human Retroviruses* 22:182-187.
- Vidal N, Peeters M, Mulanga-Kabeya C, Nzilambi N, Robertson D, Ilunga W et al (2000). Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. *J. Virol.* 74:10498- 10507.
- Yebrá G, de Mulder M, Martín L, Rodríguez C, Labarga P, Viciano I et al (2012). Most HIV Type 1 Non-B Infections in the Spanish Cohort of Antiretroviral Treatment-Naïve HIV-Infected Patients (CoRIS) Are Due to Recombinant Viruses. *J.Clin. Microbiol.* 50:407-413.
- Zhuang JL, Jetzt AE, Sun GL, Yu H, Klarmann G, Ron Y et al (2002). Human immunodeficiency virus type 1 recombination: Rate, fidelity, and putative hot spots. *J. Virol.* 76:11273-11282.