

# Global trends in molecular epidemiology of HIV-1 during 2000–2007

Joris Hemelaar<sup>a</sup>, Eleanor Gouws<sup>b</sup>, Peter D. Ghys<sup>b</sup>,  
Saladin Osmanov<sup>c</sup> and WHO-UNAIDS Network for HIV Isolation  
and Characterisation

**Objective:** To estimate the global and regional distribution of HIV-1 subtypes and recombinants between 2000 and 2007.

**Design:** Country-specific HIV-1 molecular epidemiology data were combined with estimates of the number of HIV-infected people in each country.

**Methods:** Cross-sectional HIV-1 subtyping data were collected from 65 913 samples in 109 countries between 2000 and 2007. The distribution of HIV-1 subtypes in individual countries was weighted according to the number of HIV-infected people in each country to generate estimates of regional and global HIV-1 subtype distribution for the periods 2000–2003 and 2004–2007.

**Results:** Analysis of the global distribution of HIV-1 subtypes and recombinants in the two periods indicated a broadly stable distribution of HIV-1 subtypes worldwide with a notable increase in the proportion of circulating recombinant forms (CRFs), a decrease in unique recombinant forms (URFs) and an overall increase in recombinants. In 2004–2007, subtype C accounted for nearly half (48%) of all global infections, followed by subtypes A (12%) and B (11%), CRF02\_AG (8%), CRF01\_AE (5%), subtype G (5%) and D (2%). Subtypes F, H, J and K together cause fewer than 1% of infections worldwide. Other CRFs and URFs are each responsible for 4% of global infections, bringing the combined total of worldwide CRFs to 16% and all recombinants (CRFs along with URFs) to 20%.

**Conclusion:** The global and regional distributions of individual subtypes and recombinants are broadly stable, although CRFs may play an increasing role in the HIV pandemic. The global diversity of HIV-1 poses a formidable challenge to HIV vaccine development.

© 2011 Wolters Kluwer Health | Lippincott Williams & Wilkins

*AIDS* 2011, **25**:679–689

**Keywords:** circulating recombinant form, HIV, molecular epidemiology, recombinant, subtype, vaccine

## Introduction

HIV-1 remains a global health problem of unprecedented dimensions, with an estimated 33.4 million people living with HIV in 2008. The pandemic is dynamic, with 2.7 million new infections and 2 million deaths occurring in 2008 [1].

HIV originated from multiple zoonotic transmissions of simian immunodeficiency virus (SIV) from nonhuman primates to humans in west and central Africa in the early 1900s. HIV-1 group M, the pandemic branch of HIV, originates from SIVcpz in the chimpanzee *Pan troglodytes troglodytes* [2]. After transmission to humans, while still confined to western-central Africa, HIV-1 group M

<sup>a</sup>Nuffield Department of Obstetrics and Gynaecology, University of Oxford, John Radcliffe Hospital, Oxford, UK, <sup>b</sup>Epidemiology and Analysis Division, UNAIDS, and <sup>c</sup>WHO-UNAIDS HIV Vaccine Initiative, World Health Organisation, Geneva, Switzerland. Correspondence to Dr Joris Hemelaar, Nuffield Department of Obstetrics and Gynaecology, University of Oxford, The Women's Centre, John Radcliffe Hospital, Headington, Oxford, OX3 9DU, UK.

E-mail: joris.hemelaar@trinity.oxon.org

Received: 6 August 2010; revised: 11 November 2010; accepted: 22 November 2010.

DOI:10.1097/QAD.0b013e328342ff93

diversified into genetic subtypes (named A-D, F-H and J-K) in the first half of the twentieth century [3]. HIV genetic variability is the result of the high mutation and recombination rates of the reverse transcriptase enzyme, together with high rates of virus replication. Recombinants between subtypes are designated as circulating recombinant forms [CRFs; 48 different CRFs have been described so far (Los Alamos National Laboratory, <http://www.hiv.lanl.gov>)] if fully sequenced and found in three or more epidemiologically unlinked individuals, and unique recombinant forms (URFs) if not meeting these criteria [4]. In the second half of the twentieth century, the global spread of HIV-1 group M took place resulting in the differential global distribution of HIV-1 subtypes and recombinants [5].

HIV diversity impacts HIV diagnosis and viral load measurements [6,7] and may affect the response to antiretroviral treatment and the emergence of drug resistance [8,9]. Subtypes may differ in the rate of disease progression [10–12] and some evidence suggests that subtypes are transmitted at different rates [13,14].

HIV infection in humans induces humoral and cellular immune responses which control primary viraemia without completely eliminating infection due to the rapid selection of immune escape mutants [15,16]. Although the induction of neutralizing antibodies and effective cytotoxic T-lymphocyte (CTL) responses against HIV-1 through vaccination has proven extremely difficult, an even greater challenge is posed by the extreme genetic diversity of the virus and its continuing evolution [17].

Genetic variation within a subtype is in the order of 8–17%, whereas variation between subtypes is usually between 17 and 35%, depending on the subtypes and genome regions examined [18]. This degree of diversity is likely to limit the intrasubtype and intersubtype cross-reactivity of immune responses [19]. To increase the likelihood of vaccine-induced immune responses cross-reacting with circulating strains, immunogens should match as closely as possible with the viral sequences circulating in the target population [20]. Candidate HIV vaccines tested in efficacy trials to date have been based on primary sequences of subtype B and CRF01\_AE [21–23]. Recent encouraging results of a phase III HIV-1 vaccine efficacy trial in Thailand emphasize the importance of further studies on global and regional HIV diversity to inform rational vaccine design and development [23].

Data on the global distribution of HIV subtypes are limited. HIV sequences deposited in the *Los Alamos* sequence database are not representative for the distributions in the countries of origin (Los Alamos National Laboratory, <http://www.hiv.lanl.gov>). Some studies have relied on pooling published data from a wide

range of years, an approach which is hampered by publication bias and a lack of information on trends [24]. Here we present an analysis of the global and regional distribution of HIV-1 subtypes and recombinants for the periods 2000–2003 and 2004–2007 using a combination of cross-sectional country-specific HIV-1 molecular epidemiology data, derived from published and unpublished sources, and estimates of the number of HIV-infected people in each country.

## Methods

### Country-specific cross-sectional subtype distribution data

Data on the distribution of HIV-1 subtypes and recombinants in individual countries were obtained from researchers in the field and from a comprehensive literature review. Research laboratories across the globe specializing in subtyping of HIV-1 samples were solicited for cross-sectional HIV-1 subtyping data of samples collected between 2000 and 2007. The *MEDLINE* literature database was searched for HIV-1 subtyping data for each country using the terms ‘HIV’, ‘subtype’ and the relevant country names. All potentially relevant articles in English were retrieved and HIV-1 subtyping data from samples collected between 2000 and 2007 were included in our study. Data submitted to us and derived from the literature were combined to determine the overall proportions of HIV-1 subtypes and recombinants in each country. In addition to information on HIV-1 subtypes, the data sets included country (and region/city) of sample origin, sampling year, transmission route/risk group, detection method and the genome segment(s) analysed. Overall, subtyping data of 41.8% of the samples originated from previously unpublished data by members of the WHO-Joint United Nations Programme on HIV/AIDS Network for HIV Isolation and Characterisation, 35.6% were derived from the published literature and 22.6% were both submitted to us and published.

### Global HIV-1 epidemiology and regional country groupings

Country-specific HIV-1 epidemiology data were obtained from the UNAIDS/WHO estimates of the burden of HIV-1 in the years 2000 through 2007 [25]. For each country, the average number of HIV-1 infections in the periods of 2000–2003 and 2004–2007 was determined and used in further analysis.

Countries were grouped in geographical regions according to the UNAIDS classification [25], with some modifications. Sub-Saharan Africa was divided into five separate regions (west, east, central and southern Africa and Ethiopia) because the region has the largest number of HIV-1 infections and a high level of regional HIV-1 subtype diversity. India and Ethiopia were analysed

separately as most infections were caused by a single subtype (C), which would have skewed the distribution in their respective regions, where other subtypes predominate. These modifications resulted in the final grouping of all countries into 15 regions [25].

### Data analysis

The HIV-1 subtype distribution in each region was determined by first multiplying the proportions of all subtypes and recombinants in each country in each period (2000–2003 or 2004–2007) by the estimated (average) number of people living with HIV in the same country in the relevant period. The resulting numbers of each subtype in each country in each region were added up and used to calculate the proportions of the different subtypes and recombinants in each region. Countries for which no HIV-1 subtyping data had been obtained were not included in this part of the analysis.

To determine the global distribution of HIV-1 variants, the regional proportions of HIV-1 subtypes and recombinants were multiplied by the number of HIV-1-infected individuals in each region (including countries for which no subtyping data were obtained). The resulting total numbers of people living with each subtype in each region were added up and the global distribution of HIV-1 subtypes and recombinants was derived.

To determine the spread over the regions of individual subtypes, the estimated number of infections caused by a subtype in each region was taken as a proportion of the global number of infections caused by that same subtype.

## Results

### Primary HIV-1 subtype distribution data

HIV-1 subtype characterization data were collected from a total of 65 913 samples from HIV-infected individuals in 109 countries between 2000 and 2007. For our analysis the data were divided into two periods (2000–2003 and 2004–2007) and 15 geographical regions (Methods, Table 1, legend to Table 2). In 2000–2003, 39 148 samples from 95 countries were analysed, and in 2004–2007, 26 765 samples from 70 countries were available.

Worldwide, the countries for which HIV-1 subtype distribution data were collected accounted for 94 and 90% of individuals living with HIV-1 in 2000–2003 and 2004–2007, respectively (Table 1, last column). In nine out of the 15 regions, for both periods, the countries with subtype data represented more than 90% of people living with HIV in the region.

The number of samples analysed as a proportion of the number of people living with HIV was higher in 2000–2003 than 2004–2007 globally, as well as in all but two regions (Table 1, third last column). The proportion of the infected population sampled varied between regions and was higher in the Americas, western and central Europe, east Asia and Oceania, whereas India, Middle East, north Africa and sub-Saharan Africa were less well represented.

### Global distribution of HIV-1 subtypes and recombinants

The distribution of HIV-1 subtypes in individual countries was weighted according to the number of HIV-infected people in each country to generate estimates of regional and global HIV-1 subtype distribution for 2000–2003 and 2004–2007 (Fig. 1, Table 2a). In 2004–2007 subtype C accounted for nearly half (48%) of all global infections. Subtypes A and B caused 12 and 11% of infections, respectively, followed by CRF02\_AG (8%), CRF01\_AE (5%), subtype G (5%) and D (2%). Subtypes F, H, J and K together caused fewer than 1% of infections worldwide. Other CRFs and URFs are each responsible for 4% of global infections, bringing the combined total of worldwide CRFs (CRF01\_AE, CRF02\_AG, CRF03\_AB and other CRFs) to 16% and all recombinants (all CRFs and URFs) to over 20% (Fig. 1, Table 2a).

Overall, the global distributions of subtypes were similar between 2000–2003 and 2004–2007 and in line with previous estimates [5,24]. Three epidemiological trends were noted (Table 2a). First, an increase in the global proportion (and absolute growth) of the epidemics of subtypes A, F, G, H, CRF01\_AE, CRF02\_AG and other CRFs was observed. Second, the epidemics caused by subtypes D, J, K, CRF03\_AB and URFs decreased in size and thus their proportion of the global total became smaller. Third, the epidemics caused by subtypes B and C grew at a rate below the average, resulting in a decrease of their proportion of the global epidemic, although subtype C still caused the largest absolute increase in number of infections.

The global proportion of all CRFs combined increased by 4.5% which corresponds to a 50% increase in the number of infections. In contrast, the proportion of infections caused by URFs diminished by 3.1%, a 39% decrease of the burden of URFs. Together, all recombinants (CRFs and URFs) increased by 17%, resulting in a 1.4% increase in the proportion of recombinant infections to a total of 20.5% (Table 2a). These changes reflect a widespread trend, as an increase in the proportion of other CRFs and a decrease in URFs is widely observed in 10 of the 15 regions (Table 2b). It is further notable that, among the major subtypes, subtype A increased and subtype D decreased globally (Table 2a).

Table 1. Global and regional HIV-1 epidemiology and sample collection.

Region of the world	Number of individuals living with HIV (n) <sup>a</sup>		HIV infections as a proportion of global total (%) <sup>b</sup>		Change in the number of HIV infections (%) <sup>c</sup>	Number of samples collected (n)		Depth of sampling (%) <sup>d</sup>		Number of countries with HIV-1 subtype data (n) <sup>e</sup>		Coverage of region (%) <sup>f</sup>	
	2000–2003	2004–2007	2000–2003	2004–2007		2000–2003	2004–2007	2000–2003	2004–2007	2000–2003	2004–2007	2000–2003	2004–2007
North America	1 100 000	1 200 000	3.6	3.6	11.0	2097	4117	0.195	0.346	2	2	100.0	100.0
Caribbean	210 000	230 000	0.7	0.7	9.1	539	56	0.258	0.025	4	4	39.2	90.9
Latin America	1 400 000	1 500 000	4.6	4.7	10.7	5395	3899	0.392	0.256	13	7	77.9	75.2
Western and central Europe	620 000	700 000	2.1	2.1	13.2	14 108	7334	2.278	1.046	16	8	88.3	69.6
Eastern Europe and central Asia	780 000	1 400 000	2.6	4.3	81.1	2071	623	0.267	0.044	11	5	98.8	67.0
India	2 700 000	2 500 000	9.0	7.6	–7.4	685	449	0.025	0.018	1	1	100.0	100.0
South and south-east Asia (excluding India)	1 600 000	1 800 000	5.2	5.4	13.1	2137	2084	0.136	0.117	8	7	84.6	67.3
East Asia	510 000	690 000	1.7	2.1	34.7	1972	2142	0.384	0.310	3	3	100.0	99.9
Oceania	28 000	57 000	0.1	0.2	105.2	906	1105	3.286	1.953	2	3	56.1	97.7
Middle East and north Africa	310 000	360 000	1.0	1.1	16.2	283	363	0.091	0.101	5	3	4.1	5.5
West Africa	4 200 000	4 600 000	14.1	14.0	7.7	3252	1398	0.077	0.031	10	7	94.0	87.2
East Africa (excluding Ethiopia)	4 400 000	4 400 000	14.6	13.5	0.2	3098	2002	0.071	0.046	6	4	98.9	95.5
Ethiopia	920 000	930 000	3.1	2.8	0.3	235	39	0.025	0.004	1	1	100.0	100.0
Central Africa	910 000	1 100 000	3.0	3.3	17.1	979	388	0.107	0.036	6	6	100.0	100.0
Southern Africa	10 000 000	11 000 000	34.5	34.6	9.0	1391	766	0.013	0.007	7	9	96.0	96.0
Global	30 000 000	33 000 000	100.0	100.0	9.0	39 148	26 765	0.130	0.082	95	70	94.0	90.1

<sup>a</sup>The numbers of individuals living with HIV were obtained from the Joint United Nations Programme on HIV/AIDS/WHO estimates of the burden of HIV [25]. The average of the numbers of infections in the periods 2000–2003 and 2004–2007 are shown. Figures are rounded off.

<sup>b</sup>The numbers of individuals living with HIV in each region as a proportion of the global total, based on the averages of the numbers of infections in each period (%).

<sup>c</sup>The change in the number of HIV infections between the periods 2000–2003 and 2004–2007 as a proportion of the number of infections in 2000–2003 (%).

<sup>d</sup>The number of samples collected from a region as a proportion of the number of people living with HIV in the region (%).

<sup>e</sup>Countries for which HIV subtype data was collected are specified in the legend of Table 2.

<sup>f</sup>The combined number of people living with HIV in the countries for which HIV subtype distribution data were collected in a region (see legend of Table 2), as a proportion of the total number of people living with HIV in the region. If any data was available for a country (independent of the number of samples collected), the whole HIV-infected population in that country was deemed to be represented in this analysis.

Table 2. Distribution of HIV-1 subtypes and recombinants (a) globally and (b) in each region in 2000–2003 and 2004–2007.

a													
	Year	A	B	C	D	F	G	H	J	K	CRF01_AE	CRF02_AG	CRF03_AB
Global	2000–2003 <sup>e</sup>	11.56	11.57	49.62	2.87	0.44	4.50	0.10	0.24	0.02	4.52	5.45	0.04
	2004–2007 <sup>e</sup>	12.03	11.33	48.23	2.49	0.45	4.60	0.26	0.12	0.01	5.09	7.73	0.00
	Change <sup>f</sup>	0.47	-0.23	-1.40	-0.37	0.02	0.10	0.16	-0.12	-0.01	0.57	2.28	-0.04
	2000–2003 <sup>g</sup>	3 500 000	3 500 000	14 900 000	860 000	130 000	1 400 000	29 000	71 000	7100	1 400 000	1 600 000	12 000
	2004–2007 <sup>g</sup>	3 900 000	3 700 000	15 800 000	820 000	150 000	1 500 000	85 000	38 000	3400	1 700 000	2 500 000	100
b	Change <sup>h</sup>	13.4	6.8	5.9	-5.3	13.0	11.3	191.9	-46.4	-52.3	22.7	54.7	-99.3
	Other CRFs <sup>a</sup>												
	URFs <sup>b</sup>												
	Total CRFs <sup>c</sup>												
	Total CRFs and URFs <sup>d</sup>												
Region of the world	Year <sup>i</sup>	A	B	C	D	F	G	H	J	K	CRF01_AE	CRF02_AG	CRF03_AB
	2000–2003	0.88	95.55	1.08	0.01	0.00	0.00	0.00	0.00	0.00	0.49	1.73	0.00
	2004–2007	0.70	94.20	2.46	0.45	0.10	0.20	0.01	0.00	0.00	0.43	1.44	0.00
	Change	-0.18	-1.35	1.38	0.45	0.10	0.20	0.01	0.00	0.00	-0.06	-0.30	0.00
	2000–2003	0.03	89.67	0.28	0.00	0.01	0.13	0.01	0.00	0.00	0.00	0.02	0.00
North America	2004–2007	0.00	98.32	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Change	-0.03	8.66	-0.28	0.00	-0.01	-0.13	-0.01	0.00	0.00	0.00	-0.02	0.00
	2000–2003	0.00	74.90	8.99	0.06	4.69	0.00	0.00	0.00	0.00	0.00	0.03	0.00
	2004–2007	0.00	67.89	6.53	0.04	3.53	0.00	0.00	0.00	0.00	0.01	0.04	0.00
	Change	0.00	-7.01	-2.46	-0.02	-1.16	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Caribbean	2000–2003	2.22	84.13	2.90	0.73	0.41	2.82	0.06	0.10	0.02	0.79	2.94	0.01
	2004–2007	1.76	85.20	1.91	0.27	0.47	1.00	0.07	0.00	0.00	1.01	4.50	0.00
	Change	-0.45	1.09	-0.99	-0.46	0.07	-1.82	0.02	-0.10	-0.02	0.26	1.56	-0.01
	2000–2003	78.81	14.09	1.19	0.26	2.45	0.26	0.00	0.00	0.00	0.00	0.05	1.55
	2004–2007	80.01	15.20	1.51	0.00	1.14	0.01	0.00	0.00	0.00	0.76	0.01	0.01
Latin America	2000–2003	1.20	1.12	0.32	-0.26	-1.31	-0.25	0.00	0.00	0.00	0.50	-0.04	-1.55
	2004–2007	1.40	0.31	96.54	0.00	0.00	0.00	0.00	0.00	0.00	0.36	0.00	0.00
	Change	0.89	1.34	97.77	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2000–2003	-0.50	1.03	1.23	0.00	0.00	0.00	0.00	0.00	0.00	-0.36	0.00	0.00
	2004–2007	2.89	9.12	3.03	0.01	0.00	0.03	0.00	0.00	0.00	79.57	0.01	0.00
Western and central Europe	2000–2003	10.24	3.62	0.04	0.00	0.00	0.00	0.00	0.00	0.00	78.60	0.05	0.00
	2004–2007	7.35	-5.49	-2.99	-0.01	0.00	-0.03	0.00	0.00	0.00	10.58	0.04	0.00
	Change	0.66	33.06	0.08	0.00	0.00	0.05	0.00	0.00	0.00	27.62	0.46	0.00
	2000–2003	0.24	26.17	1.38	0.00	0.01	0.06	0.00	0.00	0.00	17.04	0.35	0.00
	2004–2007	-0.43	-6.89	1.30	0.00	0.01	0.01	0.00	0.00	0.00	4.39	1.25	0.00
Eastern Europe and central Asia	2000–2003	0.51	87.62	5.19	0.54	0.11	0.00	0.00	0.00	0.00	1.24	0.53	0.00
	2004–2007	0.24	31.17	66.34	0.00	0.06	0.09	0.00	0.00	0.00	-3.15	-0.72	0.00
	Change	-0.27	-56.45	61.15	-0.54	-0.05	0.09	0.00	0.00	0.00	0.00	12.98	0.00
	2000–2003	0.76	57.25	0.00	0.76	0.37	1.84	0.00	0.00	0.00	0.02	18.08	0.00
	2004–2007	0.98	53.59	0.45	2.23	0.37	2.84	0.00	0.00	0.00	0.02	14.84	0.00
South and south-east Asia (excluding India)	2000–2003	18.73	0.02	0.38	0.45	0.32	27.87	0.06	0.21	0.01	0.35	35.23	0.00
	2004–2007	4.36	0.35	0.66	1.87	0.84	27.59	0.12	0.15	0.06	0.07	50.07	0.00
	Change	-14.37	0.32	0.28	1.42	0.52	-0.28	0.06	-0.05	0.05	-0.28	14.84	0.00
	2000–2003	38.00	0.05	19.08	16.89	0.05	0.37	0.00	0.00	0.00	0.00	0.17	0.00
	2004–2007	50.80	0.01	22.97	13.80	0.11	0.29	0.00	0.00	0.00	0.17	0.22	0.00
East Asia	2000–2003	12.80	-0.03	3.89	-3.08	0.07	-0.08	0.00	0.00	0.00	0.00	0.06	0.00
	2004–2007	1.28	0.00	98.30	0.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Change	0.00	0.00	97.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.56	0.00
	2000–2003	-1.28	0.00	-0.86	-0.43	0.00	0.00	0.00	0.00	0.00	0.00	2.56	0.00
	2004–2007												

(continued overleaf)



Table 2 (continued)

	Year	A	B	C	D	F	G	H	J	K	CRF01_AE	CRF02_AG	CRF03_AB	Other CRFs <sup>a</sup>	URFs <sup>b</sup>	Total CRFs <sup>c</sup>	Total CRFs <sup>d</sup> and URFs <sup>e</sup>
Central Africa	2000–2003	28.74	0.14	7.98	8.81	2.45	12.35	2.90	6.73	0.72	1.69	5.15	0.00	5.54	16.78	12.38	29.17
	2004–2007	11.15	0.53	5.75	8.51	2.69	18.17	7.41	1.39	0.07	4.07	7.86	0.00	24.03	8.39	35.96	44.34
Southern Africa	Change	–17.59	0.39	–2.24	–0.30	0.23	5.82	4.51	–5.35	–0.66	2.37	2.71	0.00	18.49	–8.40	23.57	15.18
	2000–2003	0.06	0.69	98.77	0.06	0.06	0.12	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.17	0.06	0.23
	2004–2007	0.21	0.04	98.31	0.14	0.00	0.17	0.00	0.14	0.00	0.00	0.02	0.00	0.02	0.95	0.05	0.99
	Change	0.14	–0.65	–0.46	0.08	–0.06	0.05	0.00	0.14	0.00	0.00	–0.04	0.00	0.02	0.78	–0.02	0.76

Proportions of HIV-1 subtypes and recombinants in each region and the world (%). The countries making up each region are as follows: *North America*: Canada, USA; *Caribbean*: Bahamas, Barbados, Cuba, Dominican Republic, Haiti, Jamaica, Trinidad and Tobago; *Latin America*: Argentina, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Guyana, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Suriname, Uruguay, Venezuela; *Western and Central Europe*: Albania, Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Luxembourg, Malta, Montenegro, Netherlands, Norway, Poland, Portugal, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, The Former Yugoslav Republic of Macedonia, the United Kingdom; *Eastern Europe and Central Asia*: Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Republic of Moldova, Romania, Russian Federation, Tajikistan, Turkmenistan, Ukraine, Uzbekistan; *India, South and South-East Asia (excl. India)*: Afghanistan, Bangladesh, Bhutan, Brunei Darussalam, Cambodia, Indonesia, Iran (Islamic Republic of), Lao People's Democratic Republic, Malaysia, Maldives, Myanmar, Nepal, Pakistan, Philippines, Singapore, Sri Lanka, Thailand, Timor-Leste, Viet Nam; *East Asia, China*: Democratic People's Republic of Korea, Japan, Mongolia, Republic of Korea; *Oceania*: Australia, Fiji, New Zealand, Papua New Guinea; *Middle East and North Africa*: Algeria, Bahrain, Cyprus, Egypt, Iraq, Jordan, Kuwait, Lebanon, Libyan Arab Jamahiriya, Morocco, Oman, Qatar, Saudi Arabia, Sudan, Syrian Arab Republic, Tunisia, Turkey, United Arab Emirates, Yemen; *West Africa*: Benin, Burkina Faso, Cameroon, Côte d'Ivoire, Equatorial Guinea, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Mauritania, Niger, Nigeria, Senegal, Sierra Leone, Togo; *East Africa (excl. Ethiopia)*: Burundi, Djibouti, Eritrea, Kenya, Mauritius, Rwanda, Somalia, Uganda, United Republic of Tanzania, Ethiopia; *Central Africa*: Angola, Central African Republic, Chad, Congo, Democratic Republic of the Congo, Gabon; *Southern Africa*: Botswana, Comoros, Lesotho, Madagascar, Malawi, Mozambique, Namibia, South Africa, Swaziland, Zambia, Zimbabwe. Composition of regions: countries are underlined if HIV-1 subtype distribution data was obtained for the period 2000–2003 (hatched); 2004–2007 (single underlined); or both 2000–2003 and 2004–2007 (double underlined). CRF, circulating recombinant form; URF, unique recombinant form.

<sup>a</sup>CRFs other than CRF01\_AE, CRF02\_AG and CRF03\_AB.

<sup>b</sup>URFs.

<sup>c</sup>The combined proportions of CRF01\_AE, CRF02\_AG, CRF03\_AB and other CRFs.

<sup>d</sup>The combined proportions of all CRFs (combined proportions of CRF01\_AE, CRF02\_AG, CRF03\_AB and other CRFs) and URFs.

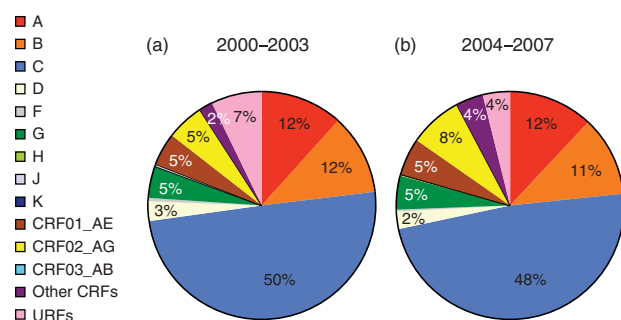
<sup>e</sup>Global proportions of HIV-1 subtypes and recombinants are presented for the periods indicated (%).

<sup>f</sup>Global changes between the two periods in the proportions of each subtype/recombinant (percentage points).

<sup>g</sup>The global number of infections caused by each subtype/recombinant in the periods indicated. Figures are rounded off.

<sup>h</sup>Changes between the two periods in the global numbers of each subtype/recombinant (%).

<sup>i</sup>Proportions of HIV-1 subtypes and recombinants are presented for each region for the periods indicated. Changes between the two periods in the proportions of each subtype in each region (percentage points).



**Fig. 1. Global distribution of HIV-1 subtypes and recombinants in 2000–2003 and 2004–2007.** The number of infections caused by HIV-1 subtypes and recombinants are represented as a proportion of the global total number of people living with HIV-1 in (a) 2000–2003 and (b) 2004–2007 (see Table 2a). The colours representing the different HIV-1 subtypes are indicated in the legend on the left-hand side of the figure. CRF, circulating recombinant form; URF, unique recombinant form.

### Regional distribution of HIV-1 subtypes and recombinants

The distribution of HIV-1 subtypes and recombinants in each region is strikingly different across the world [Fig. 2, Table 2b and Supplemental Digital Content (SDC) 1, <http://links.lww.com/QAD/A107>, which shows the subtype distribution in individual countries in 2000–2003 and 2004–2007]. Changes in regional subtype distributions are shown in Table 2b. The regional subtype distributions in 2004–2007 are discussed here.

The greatest diversity is found in central Africa where all subtypes and many CRFs and URFs are represented (Fig. 2, Table 2b and SDC 1, <http://links.lww.com/QAD/A107>). The six countries in this region all harbour a great subtype diversity, but differ in the dominant subtypes (see SDC 1, <http://links.lww.com/QAD/A107>). In the Democratic Republic of the Congo, all subtypes, CRF01\_AE, CRF02\_AG and many other CRFs and URFs, are found, except subtype B (see SDC 1, <http://links.lww.com/QAD/A107>).

In west Africa, all subtypes are detected with the dominant variants being CRF02\_AG and subtype G. In east Africa, the majority of infections are due to subtype A, with the remainder due to subtypes C and D and URFs. In southern Africa, Ethiopia and India, the epidemics are nearly exclusively caused by subtype C.

Subtype B dominates in north America, the Caribbean, Latin America, western and central Europe and Australia. In western and central Europe, all major subtypes and many CRFs and URFs are detected. The epidemic in eastern Europe and central Asia is dominated by subtype A and subtype B.

In south and south-east Asia, CRF01\_AE is responsible for the vast majority of infections. In this region, the combined proportion of all recombinant infections is 86% which is the highest in the world. In east Asia, the epidemic is dominated by CRF07\_BC, CRF08\_BC, CRF01\_AE and subtype B. The middle east and north Africa is mainly affected by subtype B and various CRFs.

### Global spread of individual HIV-1 subtypes and recombinants

The distribution of individual HIV-1 subtypes and recombinants across the globe is shown in Fig. 3 and SDC 2 and 3 (<http://links.lww.com/QAD/A107>). The majority of global dominant subtype C is present in southern Africa and India, with further infections in east Africa and Ethiopia (Table 2). Subtype A is mainly found in east Africa, eastern Europe and Central Asia, with the remainder in west and central Africa and south and south-east Asia. The subtype B epidemic is more widely and evenly spread than the other subtypes. For eight subtypes/CRFs analysed, 95% of infections are contained in only three or fewer regions. In contrast, 95% of subtype B is spread over seven regions (Fig. 3, SDC 2 and 3, <http://links.lww.com/QAD/A107>). Interestingly, hardly any subtype B infections are found in sub-Saharan Africa.

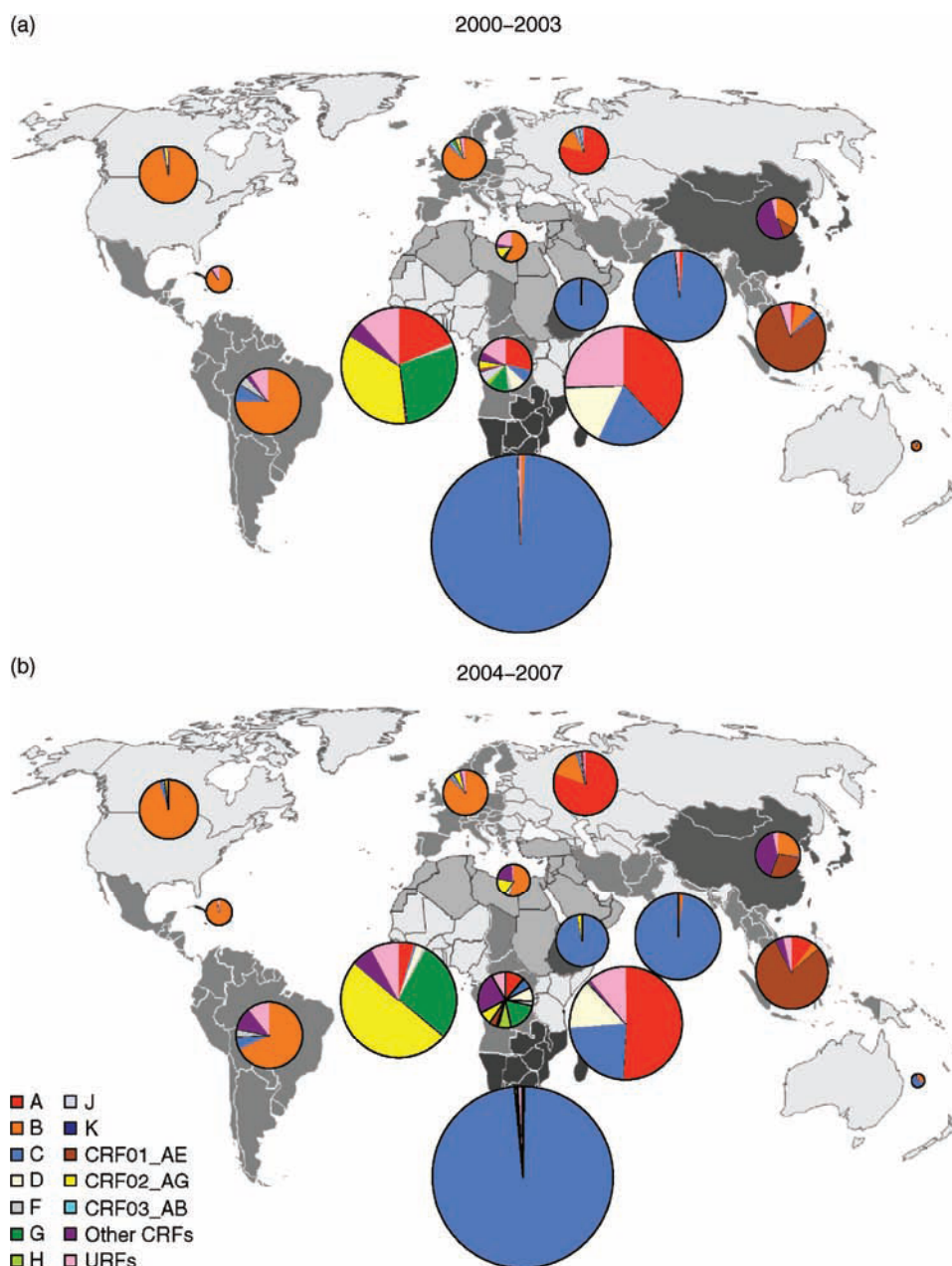
CRF02\_AG is the fourth largest variant globally and is concentrated in west Africa, with smaller numbers in central Africa, the Middle East and north Africa. CRF01\_AE is the fifth largest subtype and is found in south and south-east Asia, east Asia and a small number in central Africa.

Subtype G is concentrated in west and central Africa. Subtype D is present mainly in eastern Africa, with further infections in central and west Africa. Subtype F is widely and evenly spread worldwide, whereas subtype H, J and K are found in central, southern and west Africa. CRF03\_AB does not play a significant role globally or regionally.

Other CRFs are differentially distributed over west Africa (mainly CRF06\_cpx), east Asia (mainly CRF07\_BC and CRF08\_BC in China), central Africa (CRF11\_cpx among others), Latin America (CRF12\_BE, CRF28\_BE, CRF31\_BC, CRF38\_BE and others), Middle East and north Africa (mainly CRF06\_cpx) and south and south-east Asia (mainly CRF35\_AD and CRF07\_BC). A wide variety of URFs is distributed over sub-Saharan Africa, Latin America (mainly unique recombinants of subtypes B and F) and south and south-east Africa.

### Discussion

The global distribution of HIV-1 subtypes was broadly stable over the 2000–2007 period, with an overall

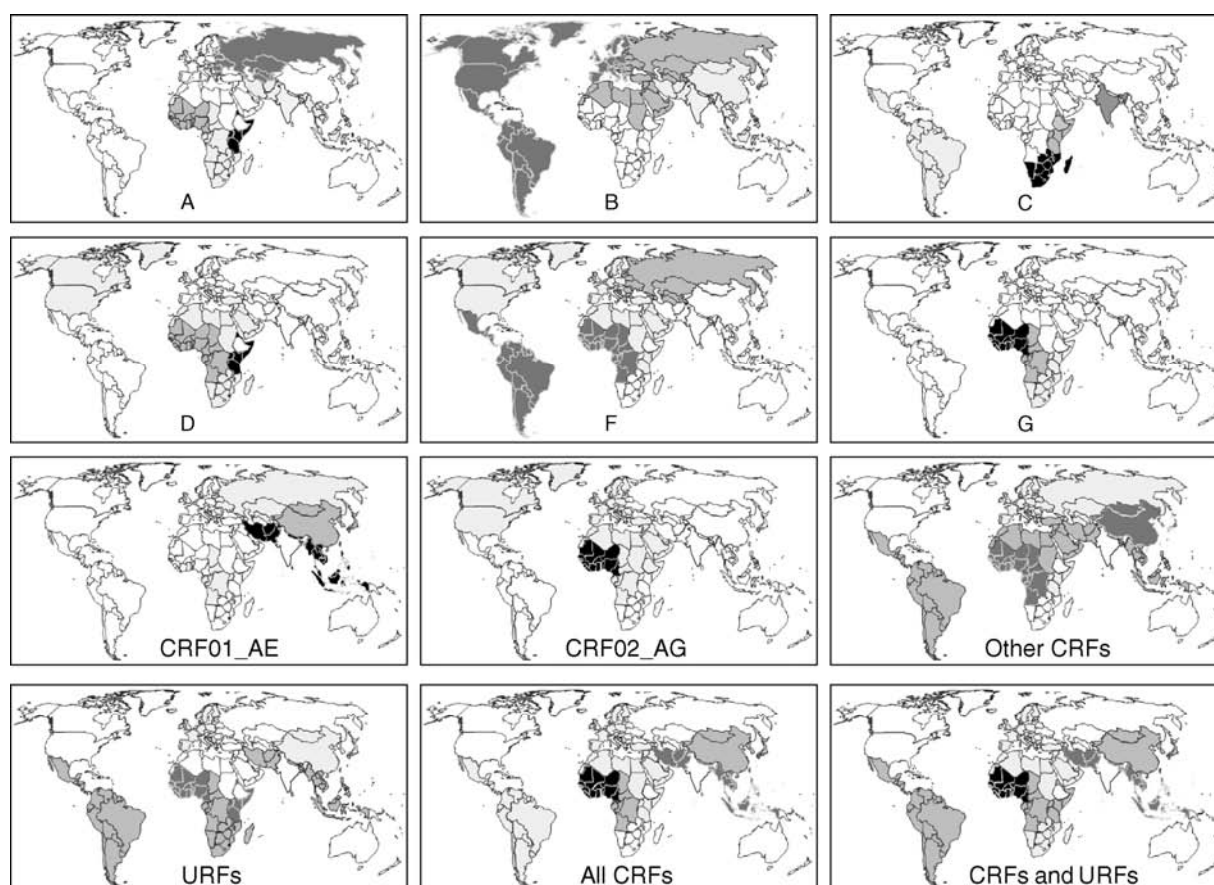


**Fig. 2. Regional distribution of HIV-1 subtypes and recombinants in 2000–2003 and 2004–2007.** The world was divided into 15 regions consisting of groups of countries as specified in the Methods and legend to Table 2. Countries forming a region are shaded in the same colour. Pie charts representing the distribution of HIV-1 subtypes and recombinants in each region in (a) 2000–2003 and (b) 2004–2007 are superimposed on the regions. The pie charts were prepared using the data presented in Table 2b. The colours representing the different HIV-1 subtypes are indicated in the legend on the left-hand side of the figure. The relative surface areas of the pie charts correspond to the relative numbers of people living with HIV in the regions (Table 1). CRF, circulating recombinant form; URF, unique recombinant form.

increase in recombinants (Fig. 1, Table 2a) and dynamic changes in some regions (Fig. 2, Table 2b, SDC 1, <http://links.lww.com/QAD/A107>). The global HIV subtype distribution was broadly similar to estimated distributions obtained using published data only, as well as estimates calculated by combining country HIV subtype distributions in the *Los Alamos* database with the country-specific numbers of HIV-infected people used in our

study (*Los Alamos National Laboratory*, <http://www.hiv.lanl.gov>) [5,24] (data not shown). The observed trends in subtype distribution between the periods were caused by an interplay between changes in subtype distribution in countries and in the numbers of HIV-infected people (Tables 1 and 2, SDC 1, <http://links.lww.com/QAD/A107>). These changes over an 8-year period are consistent with a stabilizing global epidemic with a





**Fig. 3. Global distribution of individual HIV-1 subtypes and recombinants in 2004–2007.** For each subtype/recombinant indicated, each of the 15 regions is shaded according to the proportion of the global number of infections caused by the subtype/recombinant present in each region. CRF, circulating recombinant form; URF, unique recombinant form. ■ 70–100; ■ 40–70; ■ 15–40; ■ 5–15; □ 0.5–5; □ <0.5%; Percentage of global total.

significant annual turnover due to new infections and deaths, and rapid growth of certain regional epidemics (Table 1) [1].

Our study has some limitations. Some regions had poor coverage because of lack of data (Caribbean, Oceania, middle east and north Africa; Table 1, last column), some countries and regions had small sample sizes, particularly those which harbour the largest number of infections and have the highest subtype diversity (Table 1) and only a small amount of data were obtained from representative national surveys. Moreover, the heterogeneous nature of the data sets from the two periods precluded a direct comparison and statistical analysis. Sampling biases may have occurred due to patient selection (risk groups, treatment failure and disease progression), limited geographical coverage and consistency within countries and unknown dates and places of infection. In addition, subtyping methods used and the type and number of genome segments analysed may have affected results. Finally, publication bias may have occurred in the data derived from the literature.

We found a notable global increase in the proportion of CRFs, a decrease in URFs and an overall increase in recombinants, although no formal statistical test for trend was used (Table 2a). Detailed examination showed that these trends could not be attributed to the subtyping methods used or the type and number of genome segments analysed in the different periods (data not shown). However, given the relatively recent establishment of rules governing CRF nomenclature [4], samples with discordant subtypes in different genome segments may have been classified as recombinants (or URFs), before the relevant CRFs were characterized, especially in the earlier period. It is probable that the overall proportion of recombinants is underestimated in both periods due to the limited number of full-length sequences available (0.6–5%, data not shown).

Independent studies in Uganda, Tanzania and Kenya suggest that subtype D infection is associated with faster disease progression than subtype A in populations in which they cocirculate, despite similar plasma viral loads [10–12]. In addition, a higher rate of heterosexual

transmission of subtype A than subtype D was reported [13]. A long-term study examining the subtype distribution in Kenya during the expanding epidemic found a significant decrease in the proportion of subtype D and a slight increase in subtype A [26]. These reports are consistent with our findings of an increase in the proportion of subtype A and a decrease in subtype D in east Africa and globally.

The explanations for the current global subtype distribution and the recent changes observed are probably multifactorial and include founder effects, population growth and urbanization and improved transport links and migration [3,27]. It is uncertain at present whether biological properties of different subtypes and recombinants play a role in their differential spread.

HIV diversity, in populations and in individuals, is one of the major challenges in HIV vaccine development. It seems clear that vaccine immunogen sequence should match as closely as possible to the viral sequences circulating in the target population. Up-to-date and accurate information on HIV subtype distribution is, therefore, essential, not least to ensure that efforts and funding are allocated according to the regional and global impact of the various HIV subtypes and recombinants. Our study highlights that the distribution of subtypes and recombinants globally and regionally is extremely complex. This diversity may be addressed by the use of consensus, ancestral, centre-of-tree or mosaic sequences in vaccines [20,28].

The challenges posed by the genetic diversity of the HIV-1 pandemic to prevention and treatment efforts demand that global molecular epidemiology surveillance is continued and improved. Particularly, full length sequencing of samples obtained from nationally representative surveys, taking account of geographical differences and differences in transmission routes/risk groups, are urgently needed.

## Acknowledgements

J.H. is an Academic Clinical Fellow supported by the National Institute of Health Research (NIHR), UK.

J.H. conceived and designed the project, collected HIV subtype data from contributors, conducted the literature search, analysed the data, prepared the figures and tables, interpreted the data and wrote the manuscript. E.G. provided the HIV epidemiology data, performed statistical analysis and interpreted the data. P.D.G. provided supervision and interpreted the data. S.O. conceived and designed the project, provided supervision and interpreted the data. All authors participated in the

writing of the manuscript. Members of the WHO-UNAIDS Network for HIV Isolation and Characterisation contributed HIV subtyping data to the study.

The authors alone are responsible for the views expressed in this publication which does not necessarily reflect the views of the WHO and the Joint United Nations Programme on HIV/AIDS (UNAIDS). The authors declare that they have no conflicts of interest.

The boundaries and names shown and the designations used in Fig. 2 do not imply the expression of any opinion whatsoever on the part of the WHO or the Joint United Nations Programme on HIV/AIDS (UNAIDS) concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted lines on maps represent approximate borderlines for which there may not yet be full agreement.

Contributing members of the WHO/UNAIDS Network for HIV Isolation and Characterisation [name of contributor (country where contributor is based)]: S. Agwale (Nigeria); J.-P. Allain and L. Fischetti (UK); C. Archibald, J. Brooks and M. Ofner (Canada); E. Belabbes (Algeria); J. Brandful (Ghana); M. Bruckova and M. Linka (Czech Republic); F. Buonaguro and L. Buonaguro (Italy); J. Carr (USA); D. Cooper, T. Kelleher, A. Carrera, P. Cunningham, D. Dwyer and F. Raikanikoda (Australia); B. Ensoli and S. Butto (Italy); M. Essex and V. Novitsky (USA); H. Fleury (France); F. Gao (USA); G.-M. Gershy-Damet (Zimbabwe); Z. Grossman and S. Maayan (Israel); X. He (China); D. Ho and L. Zhang (USA); M. Hoelscher (Germany); M. Hosseinipour and J. van Oosterhout (Malawi); P. Kaleebu and R. Goodall (Uganda); M. Kalish and P. Kanki (USA); E. Karamov (Russian Federation); D. Kombate-Noudjo and A. Dagnra (Togo); T. Leitner (USA); I. Lorenzana de Rivera (Honduras); F. McCutchan (USA); F. Mhalu, W. Urassa, F. Mosha and D. Mloka (Tanzania); M. Morgado (Brazil); J. Mullins, M. Campbell, C. Rousseau, J. Herbeck and M. Rolland (USA); J. Najera and M. Thomson (Spain); P. Nyambi (USA); A. Papa (Greece); J. Pape and C. Nolte (Haiti); M. Peeters (France); J.-M. Reynes (Cambodia); M. Salminen (Finland); H. Salomon and M. Carillo (Argentina); B. Schroeder (New Zealand); M. Segondy and B. Montes (France); J. Servais, A. Pelletier, K. Kayitenkore and J.-C. Karasi (Rwanda); R. Shankarappa (USA); Y. Shao, X. He and J. Xu (China); T. Smolskaya (Russian Federation); M. Soares and A. Tanuri (Brazil); E. Songok (Kenya); R. Sutthent (Thailand); Y. Takebe (Japan); H. Ushijima and T. Quang (Japan and Vietnam); P. Van de Perre and Méda (Burkina Faso); A. van Sighem (the Netherlands); A.-M. Vandamme and J. Vercauteren (Belgium); C. Williamson, H. Bredell and D. Stewart (South Africa); D. Wolday (Ethiopia); J. Xu (China); C. Yang (USA); D. Yirell (UK); L. Zhang, R. Zhang and Z. Chen (China).

## References

1. UNAIDS. *AIDS epidemic update 2009*. Geneva: UNAIDS; 2009.
2. Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, *et al.* **Chimpanzee reservoirs of pandemic and nonpandemic HIV-1.** *Science* 2006; **313**:523–526.
3. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, *et al.* **Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960.** *Nature* 2008; **455**:661–664.
4. Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK, *et al.* **HIV-1 nomenclature proposal.** *Science* 2000; **288**:55–56.
5. Hemelaar J, Gouws E, Ghys PD, Osmanov S. **Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004.** *AIDS* 2006; **20**:W13–W23.
6. Plantier JC, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemee V, *et al.* **A new human immunodeficiency virus derived from gorillas.** *Nat Med* 2009; **15**:871–872.
7. Kim JE, Beckthold B, Chen Z, Mihowich J, Malloch L, Gill MJ. **Short communication: identification of a novel HIV type 1 subtype H/J recombinant in Canada with discordant HIV viral load (RNA) values in three different commercial assays.** *AIDS Res Hum Retroviruses* 2007; **23**:1309–1313.
8. Geretti AM, Harrison L, Green H, Sabin C, Hill T, Fearnhill E, *et al.* **Effect of HIV-1 subtype on virologic and immunologic response to starting highly active antiretroviral therapy.** *Clin Infect Dis* 2009; **48**:1296–1305.
9. Martinez-Cajas JL, Pai NP, Klein MB, Wainberg MA. **Differences in resistance mutations among HIV-1 nonsubtype B infections: a systematic review of evidence (1996–2008).** *J Int AIDS Soc* 2009; **12**:11.
10. Kiwanuka N, Laeyendecker O, Robb M, Kigozi G, Arroyo M, McCutchan F, *et al.* **Effect of human immunodeficiency virus type 1 (HIV-1) subtype on disease progression in persons from Rakai, Uganda, with incident HIV-1 infection.** *J Infect Dis* 2008; **197**:707–713.
11. Baeten JM, Chohan B, Lavreys L, Chohan V, McClelland RS, Certain L, *et al.* **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* 2007; **195**:1177–1180.
12. Vasan A, Renjifo B, Hertzmark E, Chaplin B, Msamanga G, Essex M, *et al.* **Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype.** *Clin Infect Dis* 2006; **42**:843–852.
13. Kiwanuka N, Laeyendecker O, Quinn TC, Wawer MJ, Shepherd J, Robb M, *et al.* **HIV-1 subtypes and differences in heterosexual HIV transmission among HIV-discordant couples in Rakai, Uganda.** *AIDS* 2009; **23**:2479–2484.
14. Renjifo B, Gilbert P, Chaplin B, Msamanga G, Mwakagile D, Fawzi W, *et al.* **Preferential in-utero transmission of HIV-1 subtype C as compared to HIV-1 subtype A or D.** *AIDS* 2004; **18**:1629–1636.
15. Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, *et al.* **Antibody neutralization and escape by HIV-1.** *Nature* 2003; **422**:307–312.
16. Goonetilleke N, Liu MK, Salazar-Gonzalez JF, Ferrari G, Giorgi E, Ganusov VV, *et al.* **The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection.** *J Exp Med* 2009; **206**:1253–1272.
17. Barouch DH. **Challenges in the development of an HIV-1 vaccine.** *Nature* 2008; **455**:613–619.
18. Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, Detours V. **Evolutionary and immunological implications of contemporary HIV-1 variation.** *Br Med Bull* 2001; **58**:19–42.
19. Lee JK, Stewart-Jones G, Dong T, Harlos K, Di Gleria K, Dorrell L, *et al.* **T cell cross-reactivity and conformational changes during TCR engagement.** *J Exp Med* 2004; **200**:1455–1466.
20. Gaschen B, Taylor J, Yusim K, Foley B, Gao F, Lang D, *et al.* **Diversity considerations in HIV-1 vaccine selection.** *Science* 2002; **296**:2354–2360.
21. Pitisuttithum P, Gilbert P, Gurwith M, Heyward W, Martin M, van Griensven F, *et al.* **Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand.** *J Infect Dis* 2006; **194**:1661–1671.
22. Buchbinder SP, Mehrotra DV, Duerr A, Fitzgerald DW, Mogg R, Li D, *et al.* **Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial.** *Lancet* 2008; **372**:1881–1893.
23. Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, Kaewkungwal J, Chiu J, Paris R, *et al.* **Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand.** *N Engl J Med* 2009; **361**:2209–2220.
24. Arien KK, Vanham G, Arts EJ. **Is HIV-1 evolving to a less virulent form in humans?** *Nat Rev Microbiol* 2007; **5**:141–151.
25. UNAIDS. *Report on the global AIDS epidemic*. Geneva: UNAIDS; 2008.
26. Rainwater S, DeVange S, Sagar M, Ndinya-Achola J, Mandaliya K, Kreiss JK, *et al.* **No evidence for rapid subtype C spread within an epidemic in which multiple subtypes and intersubtype recombinants circulate.** *AIDS Res Hum Retroviruses* 2005; **21**:1060–1065.
27. Gray RR, Tatem AJ, Lamers S, Hou W, Laeyendecker O, Serwadda D, *et al.* **Spatial phylodynamics of HIV-1 epidemic emergence in east Africa.** *AIDS* 2009; **23**:F9–F17.
28. Corey L, McElrath MJ. **HIV vaccines: mosaic approach to virus diversity.** *Nat Med* 2010; **16**:268–270.