

## **Title**

Global and regional epidemiology of HIV-1 recombinants in 1990-2015: a systematic review and global survey.

## **Authors**

Joris Hemelaar, DPhil

Ramyiadarsini Elangovan, BM BCh

Jason Yun, BA

Leslie Dickson-Tetteh, BM BCh

Shona Kirtley, MSc

Eleanor Gouws-Williams, PhD

Peter D Ghys, PhD

WHO-UNAIDS Network for HIV Isolation and Characterisation \*

## **Affiliations**

Nuffield Department of Women's & Reproductive Health, Women's Centre, John Radcliffe Hospital, University of Oxford, Oxford, UK (J Hemelaar DPhil, R Elangovan BM BCh, J Yun BA, L Dickson-Tetteh BM BCh).

Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Oxford, UK (S Kirtley MSc).

Fast-Track Implementation Department, UNAIDS, Geneva, Switzerland (E Gouws-Williams PhD)

Strategic Information Department, UNAIDS, Geneva, Switzerland (P D Ghys PhD)

**Corresponding author**

Dr. Joris Hemelaar

Nuffield Department of Women's & Reproductive Health, University of Oxford, Women's Centre, John Radcliffe Hospital, Oxford, OX3 9DU, UK.

Email: [jorishemelaar@gmail.com](mailto:jorishemelaar@gmail.com)

\* Members are listed in the appendix.

## **Summary**

### **Background**

Global HIV-1 genetic diversity and evolution form a major challenge to treatment and prevention efforts. An increasing number of distinct HIV-1 recombinants have been identified worldwide, but their contribution to the global epidemic is unknown. We aimed to estimate the global and regional distribution of HIV-1 recombinant forms during 1990-2015.

### **Methods**

We assembled a global HIV-1 molecular epidemiology database through a systematic literature review and a global survey. We searched PubMed, EMBASE (Ovid), CINAHL (Ebscohost), and Global Health (Ovid) for HIV-1 subtyping studies published from Jan 1, 1990, to Dec 31, 2015. Unpublished original HIV-1 subtyping data was collected through a survey among experts in the field who were members of the WHO–UNAIDS Network for HIV Isolation and Characterisation. We included prevalence studies with HIV-1 subtyping data collected during 1990-2015. Countries were grouped into 14 regions and analyses conducted for four time periods (1990-99, 2000-04, 2005-09 and 2010-15). The distribution of circulating recombinant forms (CRFs), and unique recombinant forms (URFs) in individual countries was weighted according to the UNAIDS estimates of the number of people living with HIV in each country to generate regional and global estimates of numbers and proportions of HIV-1 recombinants in each time period. The systematic review is registered with PROSPERO, number CRD42017067164.

### **Findings**

Our global data collection yielded an HIV-1 molecular epidemiology database of 383,519 samples from 116 countries over 1990-2015. We found that the proportion of recombinants increased over time, both globally and in most regions. This was due to increases in both the proportion and the number of distinct CRFs detected over time, with 57 CRFs identified globally in 2010-2015. The global and regional distribution of HIV-1 recombinants was highly diverse and evolved over time, and we found extraordinary regional variation in the numbers (0-44 CRFs), types (58 distinct CRFs) and proportions (0-80.5%) of HIV-1 recombinants. Globally, CRF02\_AG was the most prevalent recombinant, accounting for 33.9% (2705110/7978517) of all recombinant infections in 2010-2015. URFs accounted for 26.7% (2134405/7978517), CRF01\_AE for 23.0% (1840982/7978517), and other CRFs for 16.4% (1309082/7978517). Although other CRFs played smaller roles globally (<1% (<349216/34921639) of global infections each) they played increasingly important roles in regional epidemics, including in East and South-East Asia, West and Central Africa, Middle East & North Africa, and Eastern Europe & Central Asia. In addition, Central (21.3%; 243041/1143531), West (15.5%; 838476/5419010) and East (12.6%; 591140/4704986) Africa, as well as Latin America (9.6%; 153069/1586605), were found to have high proportions of URFs.

## **Interpretation**

Recombinants play an increasing role in global and regional HIV epidemics, which has important implications for the development of an HIV vaccine, as well as design of diagnostic, resistance and viral load assays. Continued and improved surveillance of the global molecular epidemiology of HIV is crucial.

**Funding**

None.

**Keywords**

Human Immunodeficiency Virus, HIV, recombinant, circulating recombinant form, CRF, unique recombinant form, URF, molecular epidemiology, systematic review.

## INTRODUCTION

In 2018 an estimated 37.9 million people were living with HIV worldwide and the HIV pandemic continues to be a major global health problem. Despite the increasing availability of anti-retroviral therapy (ART) worldwide, there were 770,000 deaths and 1.7 million new infections in 2018.<sup>1</sup> One of the key characteristics of the HIV pandemic is its extraordinary global genetic diversity, which impacts upon diagnosis, treatment and drug resistance, viral load measurement, transmission, pathogenesis, immune response and vaccine development.<sup>2-4</sup>

After zoonotic transmission of SIV from chimpanzees to humans in the beginning of the 20<sup>th</sup> century, HIV-1 group M diversified in central Africa in the first half of the 20<sup>th</sup> century, leading to distinct subtypes, designated by the letters A, B, C, D, F, G, H, J, K and L.<sup>5-7</sup> The second half of the 20<sup>th</sup> century was characterised by the global spread of HIV and ongoing diversification.<sup>8,9</sup> HIV-1 genetic variability arises largely due to the error-prone reverse transcriptase enzyme, which leads to high rates of mutation and recombination. Recombination occurs when an individual is co-infected with two or more different virus strains.<sup>10</sup> Recombinants between subtypes are designated as either Unique Recombinant Forms (URFs) or Circulating Recombinant forms (CRFs).<sup>6</sup> URFs refer to unique recombinant sequences without evidence of onward transmission. CRFs are defined as recombinant HIV-1 genomes that have infected three or more epidemiologically unrelated individuals. CRFs are consecutively named, in accordance with an internationally defined nomenclature; 106 distinct CRFs have been described thus far.<sup>6,11,12</sup> CRFs can undergo further recombination with other 'pure' subtypes or recombinants, resulting in secondary recombinants, leading to an increasingly complex array of recombinants. Some areas in the world, especially in Africa, Asia, and South America, have been noted to generate high numbers of recombinants.<sup>13</sup>

Recombination has important implications for the HIV pandemic. Each recombinant form has a distinct global and regional spread, thereby complicating the development of a globally effective HIV vaccine as well as HIV diagnostic and viral load assays.<sup>2-4,14</sup> Moreover, recombinant viruses are composed of different combinations of segments of subtypes/CRFs and this potentially confers these viruses with an evolutionary advantage due to enhanced transmission, pathogenesis or drug resistance.<sup>15,16</sup>

Although an increasing number of distinct CRFs continues to be identified, a comprehensive assessment of the contribution of specific CRFs, as well as URFs, to the global and regional HIV epidemics is currently lacking. Sequence databases contain many HIV sequences, often generated to answer specific research questions or define unusual sequences. However, samples in the databases are not representative of populations and hence not suitable for epidemiological studies.<sup>12</sup> Published studies often report representative samples, but are limited by geography, publication bias, and time delay between sampling and publication, leading to incomplete coverage.<sup>16</sup> In this study we provide a detailed analysis of the global and regional distribution of HIV-1 recombinants during 1990-2015 using the largest global HIV-1 molecular epidemiology database assembled to date.

## METHODS

### Data collection

To build a global HIV-1 molecular epidemiology database, we conducted a systematic review and a global survey.<sup>17</sup> We searched the Pubmed, EMBASE (Ovid), CINAHL (Ebscohost) and Global Health (Ovid) databases to identify HIV subtyping studies published between January 1, 1990, and December 31, 2015. Search terms included Mesh headings and Emtree terms, as well as free text words and synonyms, including “HIV”, “recombinant”, “CRF”, and “URF” and “epidemiology” (appendix pp3-6). No methodological or language filters were used. All references obtained by the searches were combined to form a central database of citations in Endnote reference manager (Endnote X7; Clarivate Analytics, Philadelphia, PA, USA). Reviewers JH, RE, JY, and LD-T screened titles and abstracts, retrieved relevant full text articles and assessed articles against the eligibility criteria (appendix p2).

Further published data was obtained from reviewing issues of four specialist journals (*AIDS*, *Journal of AIDS*, *Journal of Virology*, *AIDS Research and Human Retroviruses*) published between January 1990 and February 2016, the WHO HIV Drug Resistance Report 2012, reviews on HIV diversity, as well as papers indexed on the Scopus citation database which referenced previous publications on global HIV-1 molecular epidemiology (appendix pp7-9).

Unpublished original HIV-1 subtyping data was collected through a survey among experts in the field who were members of the WHO–UNAIDS Network for HIV Isolation and Characterisation. We contacted, by email or fax, researchers who were known to be working on HIV-1 molecular epidemiology based on previous publications, conference abstracts, or informal networking. We asked them to contribute unpublished primary HIV-1 subtyping data that had been collected as part of independent studies by completing a pre-formulated data collection template. The subtyping data as provided in each submitted dataset were taken as correct. We excluded untyped samples. A full list of contributors is included in the appendix.



## **Eligibility criteria and data extraction**

Published and unpublished studies were eligible for inclusion if they were prevalence studies of people living with HIV (PLHIV) with 20 or more samples with known country and year of sample collection (between 1990-2015) and with original HIV-1 subtyping data.

From each dataset, reviewers JH, RE, JY, and LD-T extracted the following information: country, city/region, year(s) when samples were collected, study type, population, subtyping method(s), genome segment(s) analysed, and the subtyping data, i.e. the number of HIV-1 subtypes, Circulating Recombinant Forms (CRFs), and Unique Recombinant Forms (URFs) in each data set. The country designation of a data set was determined by the country where the samples were taken and not by the country of origin of the participants. Subtyping methods included sequencing, heteroduplex mobility assay, and serotyping. Any genome segment (e.g. *gag*, *pol*, *env*) or the full-length genome could be used for subtyping. No minimum sequence length was specified and all online subtyping tools were accepted. The vast majority of data was acquired by sequencing (99.8% in 2010-2015), mostly of partial genome sequences, mainly *pol* (94.2% in 2010-2015).<sup>17</sup>

No patient identifiable information was retrieved at any stage and consent was presumed to have been obtained by the researchers who submitted or published each data set.

## **Data analysis**

Countries were grouped into 14 regions (appendix p10) and data analysis was stratified into 4 time periods: 1990-99, 2000-04, 2005-09 and 2010-15. Country-specific estimates of the number of PLHIV in each year were obtained from UNAIDS and the mean number of PLHIV in each country was calculated for each time period (Table 1). In each time period, the proportions of individual CRFs and URFs in each country were multiplied by the number of

PLHIV in the respective countries to produce regional and global estimates of the distribution of CRFs and URFs (appendix pp11-15). We then used regional estimates of the absolute numbers of infections caused by each CRF/URF to determine the global spread of each CRF/URF over the regions (appendix pp24-28). We also determined the number of distinct CRFs reported in each time period both regionally and globally (Table 1).

All calculations were conducted in Windows Excel. This systematic review is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines, as applicable. This review is registered online with PROSPERO, number CRD42017067164.

#### **Role of the funding source**

This study received no funding. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

## **RESULTS**

### **Data collection**

The systematic literature review for the period 1990-2015 yielded 894 data sets with 257,276 samples (Figure 1). This was supplemented with 1173 data sets (112,404 samples) from the global survey of the WHO-UNAIDS Network for HIV Isolation and Characterisation and a further 136 data sets (13,839 samples) from references from reports, reviews and journals, yielding a total of 2,203 data sets and 383,519 samples from 116 countries over 1990-2015 (Figure 1).

Countries were grouped into 14 regions (appendix p10) and data analysis was stratified into 4 time periods: 1990-99, 2000-04, 2005-09 and 2010-15. We achieved excellent coverage, with 52,319 samples from 77 countries in 1990-99, 107,513 samples from 93 countries in 2000-04, 146,728 samples from 97 countries in 2005-09, and 76,959 samples from 84 countries in 2010-15.

### **Global and regional proportions of HIV-1 recombinants**

The proportion of global HIV-1 infections attributable to recombinants increased consistently throughout 1990-2015 and reached 22.8% (7978517/34921639) in 2010-15 (Table 1). The global proportion of CRFs consistently increased during the study period and contributed 16.7% (5844113/34921639) in 2010-15, with URFs accounting for a further 6.1% (2134405/34921639) of global infections.

Most regions saw increases in the proportion of recombinants, although the proportion of people infected with recombinants varied greatly by region (Table 1). In 2010-15, high proportions of recombinants were found in East Asia (80.5%; 652181/810004), South East Asia (80.1%; 1516285/1893512), West Africa (68.4%; 3706246/5419010) and Middle East

and North Africa (MENA; 67.7%; 116491/171944), largely due to CRFs. Central Africa had 46.8% (534997/1143531) recombinants, which was due to both CRFs (25.5%; 291955/1143531) and the largest proportion of URFs (21.3%; 243041/1143531) of any region. Other regions with large proportions of URFs were West Africa (15.5%; 838476/5419010), East Africa (12.6%; 591140/4704986) and Latin America (9.6%; 153069/1586605). Ethiopia, Southern Africa and South Asia (India) had extremely low proportions of recombinants.

### **Global and regional numbers of CRFs**

The number of distinct CRFs globally also increased consistently throughout the four time periods (Table 1, Figure 2). In 1990-99, 29 CRFs were identified across all datasets studied, which increased to 43 in 2000-2004, 54 in 2005-2009, and 57 in 2010-2015 (note that time periods are not equal in length). Compared to 2005-2009, for the period 2010-2015 CRF46\_BF1, CRF60\_BC, CRF63\_02A1 and CRF73\_BG were reported, but CRF38\_BF1 was not reported. Although the number of CRFs increased over time in most regions, there was large regional variation in the number of CRFs. Western and Central Europe and North America (WCENA) had the highest number of CRFs across all time periods, which increased consistently to 44 CRFs in 2010-15. There were also high numbers of CRFs in Central and West Africa (23 and 19, respectively) and Latin America (14) in 2010-15. South East Asia (8), East Asia (10), MENA (10) and Oceania (11) also had significant numbers of CRFs. In contrast, Ethiopia and South Asia (India) had very low numbers of CRFs across 1990-2015, with no CRFs recorded in either region in 2010-15.

### **Global distribution of HIV-1 recombinants**

The global proportions of recombinants changed across the four time periods (Figure 3, appendix pp11-15). Throughout all time periods, CRF01\_AE, CRF02\_AG and URFs

accounted for the majority of global recombinants and by 2010-15 they contributed 23.0% (1840982/7978517), 33.9% (2705110/7978517) and 26.7% (2134405/7978517) of global recombinants, respectively. The proportion of other CRFs increased over time and was 16.4% (1309082/7978517) in 2010-15. Major other CRFs in 2010-15 were CRF06\_cpx (2.8% (222870/7978517) of global recombinants), CRF07\_BC (2.4%; 190145/7978517), CRF11\_cpx (1.2%; 94860/7978517), CRF35\_AD (1.2%; 94375/7978517) and CRF63\_02A1 (3.3%; 265172/7978517).

### **Regional distribution of HIV-1 recombinants**

The regional distribution of recombinants in 2010-2015 is displayed in Figure 4 (see appendix pp16-23 for all time periods). In the Caribbean, 6.7% (19935/295748) of infections were recombinants in 2010-2015, an increase from 0.7% (2697/366821) in 2005-2009 (Table 1). The main recombinants in 2010-15 were URFs (42.6% (8498/19935) of recombinants), CRF19\_cpx (24.8%; 4935/19935), CRF23\_BG (17.2%; 3424/19935) and CRF18\_cpx (10.5%; 2089/19935) (Figure 4, appendix pp11-15). In Latin America, recombinants accounted for 12.2% (193025/1586605) of infections in 2010-15, including URFs (79.3%; 153069/193025), CRF07\_BC (6.1%; 11857/193025) and CRF12\_BF (2.9%; 5630/193025). In WCENA, a consistent increase in the proportion of recombinants was seen over time, with 7.7% (171454/2236833) recombinants in 2010-15. The main contributors were CRF02\_AG (39.4%; 67784/171454), URFs (25.1%; 43229/171454), CRF01\_AE (11.3%; 19526/171454) and CRF06\_cpx (5.7%; 9764/171454). In Eastern Europe and Central Asia (EECA), 23.7% (298199/1256028) of infections were recombinants in 2010-15, which was a large increase compared to preceding periods. The main recombinants here were CRF63\_02A1 (88.9%; 265172/298199), CRF02\_AG (6.0%; 17836/298199) and URFs (5.0%; 14896/298199).

In South Asia (India), only 3.1% (66497/2136545) of infections were recombinants in 2010-15, which were all URFs. In South-East Asia, recombinants accounted for 80.1% (1516285/1893512) of infections in 2010-15, mainly CRF01\_AE (90.3%; 1378602/1516285), URFs (7.3%; 111004/1516285) and CRF33\_01B (1.8%; 28147/1516285). East Asia had the highest proportion of recombinants (80.5%; 652181/810004) in 2010-15, which included CRF01\_AE (58.6%; 381996/652181), CRF07\_BC (27.2%; 177541/652181), CRF08\_BC (7.1%; 46090/652181) and URFs (6.2%; 40661/652181). In Oceania, 14.4% (9412/65354) of infections were recombinants in 2010-15, an increase from 7.9% (4455/56252) in 2005-2009, mainly CRF01\_AE (74.5%; 7010/9412), URFs (11.7%; 1106/9412) and CRF02\_AG (11.2%; 1053/9412).

In MENA, recombinants accounted for 67.7% (116491/171944) of infections in 2010-2015, an increase from 50.7% (63815/125841) in 2005-2009. The main recombinants in 2010-2015 were CRF35\_AD (79.1%; 92187/116491), CRF02\_AG (9.9%; 11530/116491) and CRF06\_cpx (8.5%; 9928/116491). In West Africa, 68.4% (3706246/5419010) of infections were recombinants in 2010-15, including CRF02\_AG (67.6%; 2504438/3706246), URFs (22.6%; 838476/3706246) and CRF06\_cpx (5.4%; 201947/3706246). In East Africa, recombinants accounted for 13.9% (654760/4704986) of infections in 2010-15, mainly URFs (90.3%; 591140/654760), CRF01\_AE (4.7%; 30637/654760) and CRF10\_CD (3.8%; 24995/654760). In Ethiopia, no recombinants were detected in 2010-15. In Central Africa, recombinants accounted for 46.8% (534997/1143531) of infections in 2010-15, which was very similar to 2005-2009 (47.5%; 556940/1173716), but an increase compared to 2000-04 and 1990-99 (30.3% (339418/1119699) and 16.5% (124571/755554, respectively). The main recombinants in 2010-15 were URFs (45.4%; 243041/534997), CRF02\_AG (16.5%; 88510/534997), CRF11\_cpx (8.7%; 46304/534997), CRF13\_cpx (6.0%; 32017/534997), CRF18\_cpx (4.7%; 25311/534997), CRF25\_cpx (4.5%; 24249/534997), and CRF45\_AKU

(3.9%; 20672/534997). In Southern Africa, only 0.3% (39035/12493788) of infections were recombinants in 2010-15.

### **Global spread of HIV-1 recombinants**

Finally, we examined the global spread over the regions of those recombinants which contributed most to the global HIV epidemic in 2010-15 (see appendix p2, and pp24-44 for all recombinants in 2010-15). URFs accounted for 26.7% (2134405/7978517) of all global recombinant infections in 2010-15 and played particularly important roles in Latin America and Central, West and East Africa, where they contributed 79.3% (153069/193025), 45.4% (243041/534997), 22.6% (838476/3706246) and 90.3% (591140/654760) of recombinant infections, respectively (appendix pp11-15). CRF01\_AE accounted for 23.0% (1840982/7978517) of global recombinant infections. This variant played an especially prominent role in both South-East Asia and East Asia, causing 90.3% (1378602/1516285) and 58.6% (381996/652181) of recombinant infections, respectively. CRF02\_AG was the largest recombinant globally, accounting for 33.9% (2705110/7978517) of all global recombinant infections, and played a particularly important role in West Africa, where it gave rise to 67.6% (2504438/3706246) of recombinant infections. CRF06\_cpx accounted for 2.8% (222870/7978517) of recombinant HIV-1 infections worldwide, playing a role in MENA, WCENA and West Africa, contributing 8.5% (9928/116491), 5.7% (9764/171454) and 5.4% (201947/3706246), respectively, of recombinants. CRF07\_BC and CRF08\_BC account for 2.4% (190145/7978517) and 0.6% (49998/7978517), respectively, of global recombinants and are concentrated in East Asia, where they contributed 27.2% (177541/652181) and 7.1% (46090/652181) of recombinants. CRF11\_cpx caused 1.2% (94860/7978517) of global recombinant HIV-1 infections and contributed 8.7% (46304/534997) of recombinants seen in Central Africa. Similarly, CRF13\_cpx, CRF18\_cpx, CRF25\_cpx and CRF45\_AKU

contributed 6.0% (32017/534997), 4.7% (25311/534997), 4.5% (24249/534997) and 3.9% (20672/534997), respectively, of recombinants in Central Africa. CRF33\_01B comprised 0.4% (28616/7978517) of recombinant infections globally and contributed 1.8% (28147/1516285) of recombinants in South-East Asia. CRF35\_AD played an important role in MENA, accounting for 79.1% (92187/116491) of recombinants in this region, thereby contributing 1.2% (94375/7978517) of all recombinant infections globally. Finally, CRF63\_02A1 accounted for 3.3% (265172/7978517) of global recombinant infections and played a prominent role in EECA, contributing 88.9% (265172/298199) of recombinants in this region.



## DISCUSSION

This is the first study to comprehensively analyse the global and regional distribution of HIV-1 recombinant forms, using the largest global HIV-1 molecular epidemiology database assembled to date. We found that the proportion of recombinants increased over time, both globally and in most regions. This was due to increases in both the proportion and the number of distinct CRFs detected, with 57 CRFs identified globally in 2010-2015. The global and regional distribution of HIV-1 recombinants was highly diverse and evolved over time, and we found extraordinary regional variation in the numbers (0-44 CRFs), types (58 distinct CRFs) and proportions (0-80.5%) of HIV-1 recombinants. Globally, CRF02\_AG was the most prevalent recombinant, accounting for 33.9% (2705110/7978517) of recombinants in 2010-2015. URFs accounted for 26.7% (2134405/7978517) of recombinants, CRF01\_AE for 23.0% (1840982/7978517), and other CRFs for 16.4% (1309082/7978517). Although other CRFs played smaller roles globally (<1% (<349216/34921639) of global infections each) they played increasingly important roles in regional epidemics, including in East and South-East Asia, West and Central Africa, Middle East & North Africa, and Eastern Europe & Central Asia. In addition, Central (21.3%; 243041/1143531), West (15.5%; 838476/5419010) and East (12.6%; 591140/4704986) Africa, as well as Latin America (9.6%; 153069/1586605), were found to have high proportions of URFs.

The diverse distribution patterns of HIV variants are determined by complex factors, including social transmission networks, urbanization, transportation networks, migration, founder effects, and population growth.<sup>8,9</sup> The increases in both the number and proportion of recombinants, especially CRFs, over time suggest that recombinants may have an evolutionary advantage, in terms of transmissibility and pathogenesis, compared to established HIV

strains.<sup>16,18</sup> However, it remains difficult to ascertain this in epidemiological studies, due to confounding by clinical parameters such as duration of infection and viral load, and host genetic, behavioural and environmental/geographic factors, and the differential availability of antiretroviral therapy.

Co-circulation of diverse HIV variants is a prerequisite for the formation of new recombinants, i.e. URFs, which become CRFs if successfully propagated in the population. Latin America and East, West and Central Africa were found to have high proportions of URFs, marking these regions out as recombination hotspots. In Central and West Africa this correlated with high numbers and proportions of CRFs. In contrast, in East Africa a relatively small number of CRFs were found which make a small contribution to the regional epidemic. The CRFs in East Africa are composed of the locally co-circulating subtypes A, C and D, such as CRF10\_CD, CRF21\_A2D, and CRF35\_AD. In Latin America the formation of URFs has led to the establishment of a large number of distinct B/F and B/C CRFs, although these CRFs do not currently form a large proportion of HIV infections. These less abundant CRFs may have formed more recently or may transmit less easily.

Several CRFs make important contributions in specific regions. For instance, CRF07\_BC and CRF08\_BC are critical players in China, CRF35\_AD plays a crucial role in MENA, and CRF63\_02A1 has risen sharply in EECA. Intravenous drug use has been implicated in the origin and spread of CRF07\_BC and CRF08\_BC in China.<sup>19</sup> The dominant CRF in MENA, CRF35\_AD, was originally identified in Afghanistan amongst PWID. Migration of Afghans to Pakistan, Iran and other countries in the region facilitated its spread, especially amongst the intravenous drug user population.<sup>20</sup> CRF63\_02A1 established an epidemic among PWID in Uzbekistan and Kazakhstan, from where it disseminated to PWID in Russia.<sup>21</sup> The HIV epidemics in both MENA and EECA have seen the largest increases in new infections in the world (10% and 29%, respectively, between 2010-2018) which are driven by CRF infections

in PWID, the key populations which account for 37% and 41% of new infections in these respective regions.<sup>1</sup> Of note, these surges in new infections are enabled by low ART coverage in MENA (32%) and EECA (38%).<sup>1</sup> WCENA had the highest number of distinct CRFs throughout the study period, driven first by immigration of predominantly heterosexual women from outside Europe and subsequent spread within Europe.<sup>22</sup> In contrast, in Southern Africa, Ethiopia, and South Asia (India), which are dominated by subtype C, the role played by recombinants remained negligible throughout the entire study period. This is likely the result of a founder effect, with ongoing transmissions occurring within these regions rather than migration into these regions.<sup>23</sup>

This study has several strengths. This is the first study to quantify the contribution of individual CRFs and URFs to the global and regional HIV epidemics. We assembled the largest global HIV-1 molecular epidemiology database, including 383,519 samples covering 1990-2015. Our study was based on published and unpublished epidemiology studies, instead of HIV sequence databases, thereby increasing the coverage and representativeness.<sup>17</sup> In addition to its unprecedented large size, a strength of our study is that the vast majority of data is acquired by sequencing (99.8% in 2010-2015), mostly of partial genome sequences, mainly *pol* (94.2% in 2010-2015).

Our study has some limitations. The accuracy of our estimates depends on the quantity and quality of the underlying data. Although we assembled a very large database there was inevitably variation in the spatial and temporal coverage as well as absolute numbers of samples and depth of coverage in relation to the size of epidemics in each country.<sup>17</sup> Other limitations of our study included heterogeneity among data sets in study design, population/risk

groups, geographical sites, number and type of genome segments, subtyping methods, and publication bias.<sup>17</sup>

Over time new CRFs have been discovered and described, which could contribute to the increase in number and proportions of CRFs over the course of our study.<sup>12</sup> However, given that most samples were characterised in only one genome segment it is likely that we underestimated recombination. Indeed, over time the methods used and number and type of genome segments analysed decreased and became dominated by *pol* sequencing for the vast majority of samples. To improve future HIV-1 molecular epidemiology studies, samples should be sequenced in multiple genome segments and preferably the full length genome, which is increasingly feasible with next generation sequencing techniques.<sup>24,25</sup>

58 distinct CRFs were identified in the epidemiological studies which contributed to our study, although 106 CRFs have been described to date.<sup>11</sup> This discrepancy is in part due to the fact that only 74 CRFs were described at the time of our data collection (i.e. up until 2015). In addition, new CRFs are often discovered due to targeted analysis of unusual sequences and unfortunately these samples have not always been described as part of epidemiological studies suitable for inclusion in our analysis.<sup>11</sup>

Finally, we did not have actual viral sequence data and consequently we were unable to perform phylogenetic, phylodynamic or phylogeographic analyses.<sup>24</sup>

The increasing numbers and contributions of URFs and CRFs to the global and regional HIV epidemics have important implications for HIV diagnostic, resistance and viral load assays, which are crucial to achieving the UNAIDS 90:90:90 treatment targets.<sup>1,26</sup> On-going recombination will lead to the generation of new variants, which viral assays may not detect or detect less efficiently.<sup>26</sup> This is of particular concern in regions with high levels of URFs, such as Latin America and East, West and Central Africa. Regions with many different CRFs, such

as WCENA, face a similar challenge. Moreover, resistance to antiretroviral drugs is influenced by HIV subtypes, although drug resistance patterns are not well described in less common subtypes, CRFs and URFs.<sup>27,28</sup> Lastly, with the continued recombination of HIV variants, there is a risk of the generation of multi-drug-resistant HIV viruses.<sup>15</sup> It is clear that diagnostic, resistance and viral load assays need to be continuously adapted to the evolving HIV epidemic.

The diversification of the HIV pandemic is a major challenge to the development of a HIV vaccine.<sup>4</sup> A globally effective HIV vaccine will need to protect against divergent HIV subtypes and recombinants. Variation between HIV subtypes is around 17-35% at amino acid level, depending on the subtypes and genome regions considered.<sup>14</sup> With the continuing generation of new recombinants, including “second generation recombinants” (i.e. recombinants of recombinants), this challenge becomes ever more difficult with time. Given the large divergence in HIV sequences, matching immunogen sequences to circulating strains will likely be important. The HIV epidemic in South Africa is largely driven by subtype C and, as a consequence, an HIV vaccine that is currently being evaluated in efficacy trials in South Africa is based on subtype C isolate sequences.<sup>29</sup> If this vaccine proves effective, it will need to be evaluated in other regions where other subtypes and recombinants predominate. A variety of approaches are being taken to address HIV diversity, including the use of artificial centralised sequences, such as consensus, ancestral or centre-of tree sequences,<sup>30,31</sup> mosaic and polyvalent vaccines,<sup>32</sup> and focusing on conserved or structurally important regions of HIV.<sup>33,34</sup> Moreover, with ongoing evolution and recombination, a vaccine may need to be changed periodically, like influenza vaccines.

In summary, our study is the first to comprehensively analyse the global and regional distribution of HIV-1 recombinant forms. We found very high and increasing numbers and

proportions of HIV recombinants, with wide regional variation and distinct global distributions of recombinants. Ongoing recombination necessitates continued and improved surveillance of the global molecular epidemiology of HIV-1 in order to inform development of an HIV vaccine and viral assays, which are crucial to achieving the UNAIDS 90:90:90 treatment targets.

## **Research in Context panel**

### **Evidence before this study**

Global HIV-1 genetic diversity and evolution form a major challenge to treatment and prevention efforts. Recombinants between subtypes are designated as either Circulating Recombinant Forms (CRFs) or Unique Recombinant Forms (URFs). CRFs are defined as recombinant HIV-1 genomes that have infected three or more epidemiologically unrelated individuals and URFs refer to unique recombinant sequences without evidence of onward transmission. Although an increasing number of distinct CRFs are identified (106 CRFs have been described so far), the contribution of specific CRFs to the global and regional HIV epidemics is unknown and no systematic review on this topic has previously been conducted. We conducted a systematic literature review by searching PubMed, EMBASE (Ovid), CINAHL (Ebscohost), and Global Health (Ovid) for HIV-1 subtyping studies published from Jan 1, 1990, to Dec 31, 2015. Search terms included Mesh headings and Emtree terms, as well as free text words and synonyms, including “HIV”, “recombinant”, “CRF”, and “URF” and “epidemiology” (appendix pp3-6). We found that data on the contribution of specific CRFs, as well as URFs, to the global and regional HIV epidemics was lacking. Published studies often report representative samples, but are limited by geography, publication bias, and time delay between sampling and publication, leading to incomplete coverage. Sequence databases contain many HIV sequences, often generated to answer specific research questions or define unusual sequences, but samples in the database are not representative of populations and hence not suitable for epidemiological studies. We therefore aimed to estimate the global and regional distribution of HIV recombinant forms during 1990-2015 through a systematic literature review and a global survey.

### **Added value of this study**

This is the first study to comprehensively analyse the global and regional distribution of HIV-1 recombinant forms. A systematic literature review and a global survey of experts generated the largest global HIV-1 molecular epidemiology database assembled to date, including 383519 samples from epidemiological studies in 116 countries collected during 1990-2015. We found that the proportions of recombinants increased over time, both globally and in most regions. This was due to increases in both the proportion and the number of distinct CRFs detected over time, with 57 CRFs identified globally in 2010-2015. The global and regional distribution of HIV-1 recombinants was highly diverse and evolved over time, and we found extraordinary regional variation in the numbers (0-44 CRFs), types (58 distinct CRFs) and proportions (0-80.5%) of HIV-1 recombinants. Globally, CRF02\_AG was the most common recombinant, accounting for 33.9% (2705110/7978517) of recombinants in 2010-2015. URFs accounted for 26.7% (2134405/7978517), CRF01\_AE for 23.0% (1840982/7978517), and other CRFs for 16.4% (1309082/7978517). Although other CRFs played smaller roles globally (<1% of global infections each) they played increasingly important roles in regional epidemics, including in East and South-East Asia, West and Central Africa, Middle East & North Africa, and Eastern Europe & Central Asia. In addition, Central (21.3%; 243041/1143531), West (15.5%; 838476/5419010) and East (12.6%; 591140/4704986) Africa, as well as Latin America (9.6%; 153069/1586605), were found to have high proportions of URFs.

### **Implications of all the available evidence**

Recombinants play an increasing role in global and regional HIV epidemics, which has important implications for the development of an HIV vaccine, as well as design of diagnostic, resistance and viral load assays. Continued and improved surveillance of the global molecular epidemiology of HIV is crucial.



### **Author contributions**

JH conceived, designed and coordinated the study, wrote the systematic review protocol, assisted with the literature search, assessed eligibility of manuscripts, collected additional published data, conducted the global survey, performed data extraction, designed the analysis, figures and tables, interpreted the data and wrote the manuscript.

RE analysed and interpreted data, made figures, and wrote the first draft of the manuscript.

RE, JY, and LD-T screened the electronic literature search results for relevant manuscripts, assessed their eligibility, extracted data, and collected additional published data.

SK designed and did the electronic literature search.

EG-W and PG provided data on the number of people living with HIV in each country.

All authors read and approved the final version of the manuscript.

### **Conflicts of interest**

We declare no competing interests.

### **Data Sharing**

Country-level published HIV subtyping data used in this study will be made available upon request to the corresponding author. Unpublished HIV subtyping data may be made available upon request at the discretion of the relevant contributing member of the WHO-UNAIDS Network for HIV Isolation and Characterisation. Country-level HIV prevalence estimates are available from <http://aidsinfo.unaids.org/>.

## Figure legends

### Figure 1. Study flow diagram.

Sources of HIV subtyping data. See Methods for details of the electronic literature search, the global survey of the WHO-UNAIDS Network for HIV Isolation and Characterisation, and other published data sources.

\* For example, HIV-positive immigrants only.

† For example, data given for subtype B and non-B samples only.

‡ For example, subtypes referred to disease states, not HIV subtypes.

### Figure 2. Regional and global numbers of Circulating Recombinant Forms in 1990-2015.

Regional and global numbers of distinct Circulating Recombinant Forms (CRFs) in 1990-99, 2000-04, 2005-9, 2010-15. CRF = Circulating Recombinant Form. WCENA = Western and Central Europe & North America, EECA = Eastern Europe & Central Asia, MENA = Middle East & North Africa. For data see Table 1.

### Figure 3. Global distribution of recombinants in 1990-2015.

Global proportions of CRFs and URFs in 2010-2015 (A), 2005-2009 (B), 2000-2004 (C), and 1990-1999 (D).

Top row: CRFs and URFs as proportions of all recombinants globally.

Bottom row: proportions of CRFs other than CRF01\_AE and CRF02\_AG.

CRF = Circulating Recombinant Form; URF = Unique Recombinant Form.

For data see appendix pp11-15.

**Figure 4. Regional distribution of recombinants in 2010-2015.**

Pie charts imposed on regions show the proportions of CRFs and URFs in each region. Countries were grouped into 14 regions (see appendix p10) and regions are shaded differentially on the world map. (A) All recombinants (URFs and CRFs) in 2010-2015. (B) CRFs, other than CRF01\_AE and CRF02\_AG, in 2010-2015.

CRF = Circulating Recombinant Form; URF = Unique Recombinant Form.

For data see appendix pp11-15.

**Table 1. Global and regional proportions and numbers of recombinants in 1990-2015.**

Recombinants, URFs and CRFs as a proportion (percentage) of global and regional HIV infections in the periods 1990-99, 2000-04, 2005-09, and 2010-15.

CRFs = Circulating Recombinant Forms, URFs = Unique Recombinant Forms.

## References

1. UNAIDS. Global AIDS Update. 2019.
2. Hemelaar J. The origin and diversity of the HIV-1 pandemic. *Trends in Molecular Medicine* 2012; **18**: 182-92.
3. Hemelaar J. Implications of HIV diversity for the HIV-1 pandemic. *Journal of Infection* 2013; **66**: 391-400.
4. Barouch DH, Korber B. HIV-1 vaccine development after STEP. *Annu Rev Med* 2010; **61**:153-67.
5. Worobey M, Gemmel M, Teuwen DE, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* 2008; **455**(7213): 661-4.
6. Robertson DL, Anderson JP, Bradac JA, et al. HIV-1 nomenclature proposal. . *Science* 2000; **288**: 55-6.
7. Yamaguchi J, Vallari A, McArthur C, et al. Brief Report: Complete Genome Sequence of CG-0018a-01 Establishes HIV-1 Subtype L. *J Acquir Immune Defic Syndr* 2020; **83**(3): 319-22.
8. Faria NR, Rambaut A, Suchard MA, et al. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* 2014; **346**(6205): 56-61.
9. Tebit DM, Arts EJ. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infectious Diseases* 2011; **11**: 45-56.
10. Vuilleumier S, Bonhoeffer S. Contribution of recombination to the evolutionary history of HIV. . *Curr Opin HIV AIDS* 2015; **10**: 84-9.
11. Zhou J, Li M, Chen M, Ma Y, Shao Y, Xing H. Near Full-Length Genomic Characterization of a Novel HIV-1 Circulating Recombinant Form (CRF106\_cpx) Identified among Heterosexuals in China. *AIDS Res Hum Retroviruses* 2020. doi: 10.1089/AID.2020.0101
12. Los Alamos National Laboratory.  
<http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>.
13. Lau KA, & Wong, J. J. . Current Trends of HIV Recombination Worldwide. *Infectious disease reports*, *Infectious Disease Reports* 2013; **5 (Suppl 1)** (e4 ).
14. Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, V. D. Evolutionary and immunological implications of contemporary HIV-1 variation. *Bri Med Bull* 2001; **58**: 19-42.
15. Spira S, Wainberg, M. A., Loemba, H., Turner, D. & Brenner, B. G. . Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. *Journal of Antimicrobial Chemotherapy* 2003; **51**: 229-40.
16. Arien KK, Vanham G, Arts EJ. Is HIV-1 evolving to a less virulent form in humans? *Nat Rev Microbiol* 2007; **5**(2): 141-51.
17. Hemelaar J, Elangovan R, Yun J, et al. Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis. *Lancet Infectious Diseases* 2019; **19**(2): 143-55.
18. Njai HF, Gali Y, Vanham G, et al. The predominance of Human Immunodeficiency Virus type 1 (HIV-1) circulating recombinant form 02 (CRF02\_AG) in West Central Africa may be related to its replicative fitness. *Retrovirology* 2006; **3**(1): 40.
19. Tee KK, Pybus OG, Li X-J, et al. Temporal and spatial dynamics of human immunodeficiency virus type 1 circulating recombinant forms 08\_BC and 07\_BC in Asia. *Journal of virology* 2008; **82**(18): 9206-15.

20. Memarnejadian A, Menbari S, Mansouri SA, et al. Transmitted Drug Resistance Mutations in Antiretroviral-Naive Injection Drug Users with Chronic HIV-1 Infection in Iran. *PLoS One* 2015; **10**(5): e0126955.
21. Kostaki EG, Karamitros T, Bobkova M, et al. Spatiotemporal Characteristics of the HIV-1 CRF02\_AG/CRF63\_02A1 Epidemic in Russia and Central Asia. *AIDS Res Hum Retroviruses* 2018; **34**(5): 415-20.
22. Apostolos Beloukas AP, Polina Giannelou, Evangelia Kostaki, Angelos Hatzakis, Dimitrios Paraskevis,. Molecular epidemiology of HIV-1 infection in Europe: An overview. *Infection, Genetics and Evolution* 2016; **46**: 180-9.
23. Wilkinson E, Rasmussen D, Ratmann O, Stadler T, Engelbrecht S, de Oliveira T. Origin, imports and exports of HIV-1 subtype C in South Africa: A historical perspective. *Infection, Genetics and Evolution* 2016; **46**: 200-8.
24. Paraskevis D, Nikolopoulos GK, Magiorkinis G, Hodges-Mameletzis I, Hatzakis A. The application of HIV molecular epidemiology to public health. *Infect Genet Evol* 2016; **46**: 159-68.
25. Aldrich C, Hemelaar J. Global HIV-1 diversity surveillance. *Trends in Molecular Medicine* 2012; **18**(12): 691-4.
26. Stone M, Bainbridge J, Sanchez AM, et al. Comparison of Detection Limits of Fourth- and Fifth-Generation Combination HIV Antigen-Antibody, p24 Antigen, and Viral Load Assays on Diverse HIV Isolates. *Journal of Clinical Microbiology* 2018; **56**(8): e02045-17.
27. TenoRes Study Group. Global epidemiology of drug resistance after failure of WHO recommended first-line regimens for adult HIV-1 infection: a multicentre retrospective cohort study. . *Lancet Infectious Diseases* 2016; **16**: 565-75.
28. HIV Drug Resistance Database. <https://hivdb.stanford.edu>.
29. Bekker L-G, Moodie Z, Grunenberg N, et al. Subtype C ALVAC-HIV and bivalent subtype C gp120/MF59 HIV-1 vaccine in low-risk, HIV-uninfected, South African adults: a phase 1/2 trial. *Lancet HIV* 2018; **8**: 30071-77.
30. Gaschen B, Taylor J, Yusim K, et al. Diversity Considerations in HIV-1 Vaccine Selection. *Science* 2002; **296**(5577): 2354-60.
31. Nickle DC, Jensen MA, Gottlieb GS, et al. Consensus and ancestral state HIV vaccines. *Science* 2003; **299**(5612): 1515-8.
32. Barouch DH, Tomaka FL, Wegmann F, et al. Evaluation of a mosaic HIV-1 vaccine in a multicentre, randomised, double-blind, placebo-controlled, phase 1/2a clinical trial (APPROACH) and in rhesus monkeys (NHP 13-19). . *Lancet* 2018; **392**: 232-43.
33. Letourneau S, Im EJ, Mashishi T, et al. Design and pre-clinical evaluation of a universal HIV-1 vaccine. *PLoS One* 2007; **2**(10): e984.
34. Gaiha GD, Rossin EJ, Urbach J, et al. Structural topology defines protective CD8(+) T cell epitopes in the HIV proteome. *Science* 2019; **364**(6439): 480-4.