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

Joris Hemelaar, Ramyidarsini Elangovan, Jason Yun, Leslie Dickson-Tetteh, Shona Kirtley, Eleanor Gouws-Williams, Peter D Ghys, on behalf of the WHO-UNAIDS Network for HIV Isolation and Characterisation*

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 the PubMed, Embase (Ovid), CINAHL (Ebscohost), and Global Health (Ovid) databases for HIV-1 subtyping studies published from Jan 1, 1990, to Dec 31, 2015. Unpublished original HIV-1 subtyping data were collected through a survey among experts in the field who were members of the WHO–UNAIDS Network for HIV Isolation and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 were done 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, CRD42017067164.

Findings Our global data collection yielded an HIV-1 molecular epidemiology database of 383,519 samples from 116 countries in 1990–2015. We found that the proportion of recombinants increased over time, both globally and in most regions, reaching 22.8% (7,978,517 of 34,921,639) of global HIV-1 infections in 2010–15. Both the proportion and the number of distinct CRFs detected increased over time to 16.7% and 57 CRFs in 2010–15. The global and regional distribution of HIV-1 recombinants was diverse and evolved over time, and we found large 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% (2,701,364 of 7,978,517) of all recombinant infections in 2010–15. URFs accounted for 26.7% (2,131,450 of 7,978,517), CRF01_AE for 23.0% (1,838,433), and other CRFs for 16.4% (1,307,270) of all recombinant infections in 2010–15. Although other CRFs accounted for small proportions of infections globally (<1% each), they were prominent in regional epidemics, including in east and southeast Asia, west and central Africa, Middle East and north Africa, and eastern Europe and central Asia. In addition, in 2010–15, central Africa (21.3% [243,041 of 1,143,531]), west Africa (15.5% [838,476 of 5,419,010]), east Africa (12.6% [591,140 of 4,704,986]), and Latin America (9.6% [153,069 of 1,586,605]) had high proportions of URFs.

Interpretation HIV-1 recombinants are increasingly prominent in global and regional HIV epidemics, which has important implications for the development of an HIV vaccine and the design of diagnostic, resistance, and viral load assays. Continued and improved surveillance of the global molecular epidemiology of HIV is crucial.

Funding None.

Copyright © 2020 Elsevier Ltd. All rights reserved.

Introduction
In 2019, 38 million people were estimated to be living with HIV worldwide and the HIV pandemic continues to be a major global health problem. Despite the increasing availability of antiretroviral therapy worldwide, 690,000 deaths and 1.7 million new infections occurred in 2019.1 A key characteristic of the HIV pandemic is its large global genetic diversity, which affects diagnosis, treatment and drug resistance, viral load measurement, transmission, pathogenesis, immune response, and vaccine development.2,4

After zoonotic transmission of simian immuno-deficiency virus from chimpanzees to humans in the beginning of the 20th century, HIV-1 group M diversified in central Africa in the first half of the century, leading to distinct subtypes, designated by the letters A, B, C, D, F, G, H, J, K, and L.3 The second half of the 20th century was characterised by the global spread of HIV and ongoing diversification.6–8 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
Recombination has important implications for the HIV epidemic. Recombinants between subtypes are designated as either circulating recombinant forms (CRFs), defined as recombinant HIV-1 genomes that have infected three or more epidemiologically unrelated individuals, or unique recombinant forms (URFs), defined as recombinants with no evidence of onward transmission. Although an increasing number of distinct CRFs are being identified (106 CRFs have been described to date), the contribution of specific CRFs to the global and regional HIV epidemics is unknown and no systematic review on this topic has previously been done. 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. Search terms included medical subject headings and Emtree terms, as well as free text words and synonyms, including “HIV”, “recombinant”, “CRF”, “URF”, and “epidemiology”. We found no studies on the contribution of specific CRFs, or of URFs, to the global and regional HIV epidemics. 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 therefore not suitable for epidemiological studies. We 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**

To our knowledge, 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 383 519 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. Both the proportion and the number of distinct CRFs detected increased over time, with 57 CRFs identified globally in 2010–15. The global and regional distribution of HIV-1 recombinants was diverse and evolved over time, and we found large regional variation in the numbers (0–44 CRFs), types (58 distinct CRFs), and proportions (0–80·5%) of HIV-1 recombinants. Globally, the most common recombinants in 2010–15 were CRF02_AG, URFs, and CRF01_AE. Other CRFs accounted for small proportions of infections globally but were prominent in regional epidemics. Central, west, and east Africa and Latin America had high proportions of URFs.

**Implications of all the available evidence**

Numbers and proportions of HIV-1 recombinants are increasing in regional and global HIV epidemics, with wide regional variation and evolving global distribution patterns. This increasing genetic diversity of the HIV pandemic has important implications for HIV transmission and pathogenesis, diagnosis, treatment and drug resistance, and viral load measurement, as well as the immune response to HIV. Ongoing recombination necessitates continued and improved surveillance of the global molecular epidemiology of HIV-1 to inform the design of a globally effective HIV vaccine and the development of viral assays, which are crucial to achieving the UNAIDS 90:90:90 treatment targets.
to incomplete coverage. 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 did a systematic review and a global survey. We searched the PubMed, Embase (Ovid), CINAHL (Ebscohost) and Global Health (Ovid) databases to identify HIV subtyping studies published between Jan 1, 1990, and Dec 31, 2015. Search terms included medical subject headings and Entree terms, as well as free text words and synonyms, including “HIV”, “recombinant”, “CRF”, “URF”, and “epidemiology” (appendix pp 3–6). 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). Authors JH, RE, JY, and LD-T screened titles and abstracts, retrieved relevant full-text articles, and assessed articles against the eligibility criteria.

Further published data were obtained from reviewing issues of four specialist journals (AIDS, Journal of AIDS, Journal of Virology, AIDS Research and Human Retroviruses) published between Jan 1, 1990, and Feb 29, 2016, the WHO HIV Drug Resistance Report 2012, reviews on HIV diversity, and papers indexed on the Scopus citation database that referenced previous publications on global HIV-1 molecular epidemiology (appendix pp 7–9).

Unpublished original HIV-1 subtyping data were 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 (pp 45–47).

**Eligibility criteria and data extraction**

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

From each dataset, authors JH, RE, JY, and LD-T extracted the following information: country, city or region, year(s) when samples were collected, study type, population, subtyping method(s), genome segment(s) analysed, and the subtyping data (ie, the number of HIV-1 subtypes, CRFs, and URFs) in each dataset. The country designation of a dataset 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 (eg, gag, pol, or 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 majority of data was acquired by sequencing (99.8% in 2010–15), mostly of partial genome sequences, mainly pol (94.2% in 2010–15). The patient-identifiable information was retrieved at any stage and consent was presumed to have been obtained by the researchers who submitted or published each dataset.

**Data analysis**

Countries were grouped into 14 regions (appendix p 10) and data analysis was stratified into four time periods:

---

Figure 1: Study flow diagram

*Eg, HIV-positive immigrants only. †Eg, data given for subtype B or non-B samples only. ‡Eg, subtypes referring to disease states, not HIV subtypes.
### Global and regional proportions and numbers of recombinants in 1990–2015

<table>
<thead>
<tr>
<th>Region</th>
<th>Number of people living with HIV</th>
<th>Proportion of people infected with recombinants, n (%)</th>
<th>Number of distinct CRFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990–99</td>
<td>270 467</td>
<td>777 108</td>
<td>123 002</td>
</tr>
<tr>
<td>2005–09</td>
<td>366 821</td>
<td>1 327 862</td>
<td>1 901 896</td>
</tr>
<tr>
<td>2010–15</td>
<td>295 748</td>
<td>1 586 605</td>
<td>2 236 833</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region</th>
<th>Proportion of people infected with recombinants, n (%)</th>
<th>Number of distinct CRFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990–99</td>
<td>39 973</td>
<td>24 575</td>
</tr>
<tr>
<td>2000–04</td>
<td>23 363</td>
<td>78 942</td>
</tr>
<tr>
<td>2005–09</td>
<td>26 671</td>
<td>162 841</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region</th>
<th>Proportion of people infected with URFs, n (%)</th>
<th>Number of distinct CRFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990–99</td>
<td>34 975</td>
<td>118 787</td>
</tr>
<tr>
<td>2000–04</td>
<td>40 018</td>
<td>62 282</td>
</tr>
<tr>
<td>2005–09</td>
<td>53 018</td>
<td>14 881</td>
</tr>
<tr>
<td>2010–15</td>
<td>84 985</td>
<td>153 069</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region</th>
<th>Proportion of people infected with CRFs, n (%)</th>
<th>Number of distinct CRFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990–99</td>
<td>18 775</td>
<td>39 979</td>
</tr>
<tr>
<td>2000–04</td>
<td>19 955</td>
<td>17 958</td>
</tr>
<tr>
<td>2005–09</td>
<td>21 506</td>
<td>28 716</td>
</tr>
<tr>
<td>2010–15</td>
<td>11 438</td>
<td>39 956</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region</th>
<th>Number of distinct CRFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990–99</td>
<td>2</td>
</tr>
<tr>
<td>2000–04</td>
<td>6</td>
</tr>
<tr>
<td>2005–09</td>
<td>6</td>
</tr>
<tr>
<td>2010–15</td>
<td>6</td>
</tr>
<tr>
<td>1990–15</td>
<td>8</td>
</tr>
</tbody>
</table>

CRFs = circulating recombinant forms. URFs = unique recombinant forms.

Table: Global and regional proportions and numbers of recombinants in 1990–2015.
Regional and global numbers of CRFs in 1990–2015

1990–99, 2000–04, 2005–09, and 2010–15. Country-specific estimates of the number of people living with HIV in each year were obtained from UNAIDS and the mean number of people living with HIV in each country was calculated for each time period. In each time period, the proportions of individual CRFs and URFs in each country were multiplied by the number of people living with HIV in the respective countries to produce regional and global estimates of the distribution of CRFs and URFs. We then used regional estimates of the absolute numbers of infections caused by each CRF or URF to determine the global spread of each CRF or URF over the regions. We also determined the number of distinct CRFs reported in each time period both regionally and globally.

All calculations were done in Microsoft Excel. This systematic review is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis guidelines, as applicable. This study is registered with PROSPERO, CRD42017067164.

Role of the funding source

There was no funding source for this study. 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

The systematic literature review for the period 1990–2015 yielded 894 datasets with 257,276 samples (figure 1). In addition, 1,173 datasets (112,404 samples) were obtained from the global survey of the WHO-UNAIDS Network for HIV Isolation and Characterisation and 136 datasets (13,839 samples) were identified from the references of reports, reviews, and journals, yielding a total of 2,203 datasets and 383,519 samples from 116 countries. 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.

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

Most regions saw increases in the proportion of recombinants over the study period, although the proportion of people infected with recombinants varied greatly by region (table). In 2010–15, high proportions of people were infected with recombinants in east Asia (80.5%), southeast Asia (80.1%), west Africa (68.4%), and the Middle East and north Africa (67.7%), largely due to CRFs. In central Africa, the proportion of people infected with recombinants was 46.8%, which comprised both CRFs (25.5%) and the largest proportion of URFs (21.3%) of any region. Other regions with high proportions of URFs were west Africa (15.5%), east Africa (12.6%), and Latin America (9.6%). Ethiopia, southern Africa, and south Asia (India) had very low proportions of recombinants.

The number of distinct CRFs globally increased consistently throughout the four time periods (table, figure 2). In 1990–99, 29 CRFs were identified across all datasets studied, which increased to 43 in 2000–04, 54 in 2005–09, and 57 in 2010–15 (note that time periods are not equal in duration). Compared with 2005–09, for the period 2010–15, CRF44_BF, 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 had the highest number of CRFs across all time periods, which increased consistently to 44 CRFs in 2010–15. The number of CRFs was also high in central Africa (23 CRFs), west Africa (19), and Latin America (14) in 2010–15. Oceania (11), east Asia (10), Middle East and north Africa (10), and southeast Asia (eight) also had considerable numbers of CRFs. By 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.

The global proportions of recombinants changed across the four time periods (figure 3, appendix pp 11–15). Throughout all time periods, CRF01_AE, CRF02_AG, and URFs accounted for the majority of global recombinants; by 2010–15 CRF01_AE contributed 23.0% (1,838,433 of 7,978,517) of global recombinants, CRF02_AG contributed 33.9% (2,701,364), and URFs contributed 26.7% (2,131,450).

The proportion of other CRFs increased over time and was 16.4% (1,307,270) in 2010–15. Other major CRFs in 2010–15 were CRF63_02A1 (3.3%; 264,805 of 7,978,517 global recombinants), CRF06_cpx (2.8%; 222,562), CRF07_BC

Figure 2: Regional and global numbers of CRFs in 1990–2015

CRF=circulating recombinant form.
In the regional distribution of recombinants in 2010–15 (figure 4A, appendix pp 16–23), URFs accounted for the largest proportion of recombinants in south Asia (100%; 66 497 of 66 497 recombinants), east Africa (90·3%; 591 140 of 654 760), Latin America (79·3%; 151 069 of 193 025), southern Africa (56·7%; 22 142 of 39 035), central Africa (45·4%; 243 041 of 534 997), and the Caribbean (42·6%; 8498 of 19 935). CRF01_AE contributed the largest proportion of recombinants in southeast Asia (90·3%; 1 369 161 of 1 516 285), Oceania (74.5%; 7010 of 9412), and east Asia (58·6%; 381 996 of 652 181). CRF02_AG contributed the largest proportion in west Africa (67·6%; 2 504 438 of 3 706 246) and western and central Europe and North America (39·4%; 67 546 of 171 454). CRF63_02A1 contributed 88.9% (265 172 of 298 199) of recombinants in eastern Europe and central Asia and CRF35_AD contributed 79·1% (92 187 of 116 491) of recombinants in the Middle East and north Africa. In east Asia, CRF07_BC (27·2%; 177 541 of 652 181) and CRF08_BC (7·1%; 46 090) were prominent. CRF06_cpx (5·5%; 201 947 of 3 706 246) was the other most common CRF in west Africa, whereas it was CRF10_CD in east Africa (3·8%; 24 995 of 654 760) and southern Africa (18·7%; 7285 of 39 035). In central Africa, the other main CRFs were CRF11_cpx (8·7%; 46 304 of 534 997), CRF13_cpx (6·0%; 32 017), CRF18_cpx (4·7%; 25 311), CRF25_cpx (4·5%; 24 249), and CRF45_AKU (3·9%; 20 672).

Discussion

To our knowledge, 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. Increases in both the proportion and the number of distinct CRFs were detected, with 57 CRFs identified globally in 2010–15. The global and regional distribution of HIV-1 recombinants was diverse and evolved over time, and we found large regional variation in the numbers (0–44 CRFs), types (58 distinct CRFs), and proportions (0–80·5%) of HIV-1 recombinants. Globally, the most prevalent recombinants were CRF02_AG (33·9% of recombinants), URFs (26·7%), and CRF01_AE (23·0%) in 2010–15. Although other CRFs played smaller roles (<1% of global infections each), they were increasingly prominent in regional epidemics, including in east and southeast Asia, west and central Africa, Middle East and north Africa, and eastern Europe and central Asia. In addition, high proportions of URFs were found in central Africa (21·3%), west Africa (15·5%), east Africa (12·6%), and Latin America (9·6%) in 2010–15.
The diverse distribution patterns of HIV variants are determined by complex factors, including social transmission networks, urbanisation, transportation networks, migration, founder effects, and population growth.\textsuperscript{8,9} The increases in both the number and proportion of recombinants, especially CRFs, over time suggest that recombinants could have an evolutionary advantage, in terms of transmissibility and pathogenesis, compared with established HIV strains.\textsuperscript{17,19} However, evaluating this hypothesis in epidemiological studies is difficult, due to confounding by clinical parameters such as duration of infection and viral load, host genetic, behavioural, and environmental factors, and the differential availability of antiretroviral therapy.

Cocirculation of diverse HIV variants is a prerequisite for the formation of new recombinants (ie, 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, suggesting that these regions are recombination hotspots. Central and

Figure 4: Regional distribution of recombinants in 2010–2015
Countries were grouped into 14 regions (appendix p 10) and regions are shaded differentially on the world map. (A) All recombinants (URFs and CRFs) in 2010–15. (B) CRFs, other than CRF01_AE and CRF02_AG, in 2010–15 (appendix pp 11–15). CRF=circulating recombinant form. URF=unique recombinant form.
west Africa also had high numbers and proportions of CRFs, including many complex CRFs (composed of >2 subtypes), such as CRF06_cpx, CRF11_cpx, and CRF13_cpx. By contrast, in east Africa, a relatively small number of CRFs were found, which made a small contribution to the regional epidemic. The CRFs in east Africa are composed of the locally cocirculating subtypes A, C, and D, such as CRF10_CD, CRF21_A2D, and CRF35_AD, and CRF50_A1D. 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 might have formed more recently or could transmit less easily.

Several CRFs make important contributions in specific regions. For instance, CRF07_BC and CRF08_BC are prominent in China, CRF35_AD is prominent in the Middle East and north Africa, and CRF63_02A1 has increased considerably since 2010 in eastern Europe and central Asia. Intravenous drug use has been implicated in the origin and spread of CRF07_BC and CRF08_BC in China. The dominant CRF in the Middle East and north Africa, CRF35_AD, was originally identified in Afghanistan among people who inject drugs. Migration of people from Afghanistan to Pakistan, Iran, and other countries in the region facilitated its spread, especially among people who inject drugs. CRF63_02A1 established an epidemic among people who inject drugs in Uzbekistan and Kazakhstan, from where it disseminated to people who inject drugs in Russia. The HIV epidemics in the Middle East and north Africa, and eastern Europe and central Asia have seen the largest increases in new infections globally (increases of 10% and 29%, respectively, between 2010 and 2018), of which people who inject drugs account for 37% of new infections in the Middle East and north Africa and 41% in eastern Europe and central Asia. Of note, these surges in new infections are enabled by low antiretroviral therapy coverage in the Middle East and north Africa (32%) and eastern Europe and central Asia (38%). Western and central Europe and North America had the highest number of distinct CRFs throughout the study period, which was predominantly driven by immigration of women from outside Europe and subsequent heterosexual transmission within Europe. By contrast, in southern Africa, Ethiopia, and south Asia (India), where subtype C is dominant, the proportion of recombinant infections was very low throughout the study period. This pattern is probably the result of a founder effect, with ongoing transmissions occurring within these regions rather than migration into the regions.

This study has several strengths. To our knowledge, it 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 383519 samples covering 1990–2015. Our study was based on published and unpublished epidemiology studies, thereby increasing the coverage and representativeness compared with existing HIV sequence databases. In addition to its unprecedented large size, a strength of our study is that the majority of data was acquired by sequencing, mostly of partial genome sequences.

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 in absolute numbers of samples and depth of coverage in relation to the size of epidemics in each country. Other limitations include heterogeneity among datasets in study design, population or risk groups, geographical sites, number and type of genome segments, subtyping methods, and publication bias.

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. However, given that most samples were characterised in only one genome segment, we are likely to have underestimated recombination. Indeed, over time, the methods used and the number and type of genome segments analysed decreased and became dominated by pol sequencing for the 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.

58 distinct CRFs were identified in the epidemiological studies that contributed to our study, although 106 CRFs have been described to date. This discrepancy is in part due to the fact that only 74 CRFs were described at the time of our data collection (ie, 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. Finally, we did not have actual viral sequence data and consequently we were unable to do phylogenetic, phylodynamic, or phylogeographic analyses.

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. Ongoing recombination could lead to the generation of new variants that viral assays do not detect or detect less efficiently. This is of particular concern in regions with high proportions of URFs, such as Latin America and east, west, and central Africa. Regions with many different CRFs, such as western and central Europe and North America, 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.
Lastly, with the continued recombination of HIV variants, a risk of the generation of multidrug-resistant HIV viruses exists. Diagnostic, resistance, and viral load assays clearly need to be continuously adapted to the evolving HIV epidemic.

The diversification of the HIV pandemic is a major challenge to the development of an HIV vaccine. A globally effective HIV vaccine will need to protect against divergent HIV subtypes and recombinants. Variation between HIV subtypes is around 17–35% at the amino acid level, depending on the subtypes and genome regions considered. With the continuing generation of new recombinants, including second-generation recombinants (ie, recombinants of recombinants), this challenge becomes even more difficult. Given the large divergence in HIV sequences, matching immunogen sequences to circulating strains is likely to be important. The HIV epidemic in South Africa is largely driven by subtype C and, as a consequence, an HIV vaccine that was recently evaluated in South Africa was based on subtype C isolate sequences. Unfortunately, the trial of this vaccine was halted due to lack of efficacy. If a subtype-specific vaccine proves effective in the future, it will also 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, and focusing on conserved or structurally important regions of HIV.

Moreover, with ongoing evolution and recombination, an HIV vaccine might 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 recombinants forms. We found 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.

Contributors 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 the 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 PDG provided data on the number of people living with HIV in each country. All authors read and approved the final version of the manuscript.

Declaration of interests 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 might 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 on the UNAIDS AIDS info website.

Editorial note: the Lancet Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

References


