



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

Joris Hemelaar, Ramyadarsini 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

Lancet HIV 2020; 7: e772–81

See Comment page e773

*Members are listed in the appendix

Nuffield Department of Population Health (J Hemelaar DPhil) and Nuffield Department of Women's & Reproductive Health, Women's Centre, John Radcliffe Hospital (J Hemelaar, R Elangovan BM BCh, J Yun BM BCh, L Dickson-Tetteh BM BCh), University of Oxford, Oxford, UK; Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Oxford, UK (S Kirtley MSc); and Fast-Track Implementation Department (E Gouws-Williams PhD) and Strategic Information Department (P D Ghys PhD), UNAIDS, Geneva, Switzerland

Correspondence to: Dr Joris Hemelaar, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK
jorishemelaar@gmail.com

See Online for appendix

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 Unpublished original HIV-1 subtyping data were collected through a survey among experts in the field who were members of the WHO–UNAIDS Network for HIV Isolation and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% (7978 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 immunodeficiency 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.^{5–7} The second half of the 20th century was characterised by the global spread of HIV and ongoing diversification.^{8,9} 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

Research in context

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), 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.

when an individual is co-infected with two or more different virus strains.^{10,11} Recombinants between subtypes are designated as either unique recombinant forms (URFs) or circulating recombinant forms (CRFs).⁶ 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.^{6,12,13} CRFs can undergo further recombination with other pure subtypes or recombinants, resulting in secondary recombinants, leading to an increasingly complex array of recombinants. Some geographical areas, especially in Africa, Asia, and South America, have been noted to generate high numbers of recombinants.¹⁴

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 diagnostic and viral load assays.^{2–4,15} Moreover, recombinant viruses are composed of different combinations of subtype or CRF segments, which potentially confer the viruses with an evolutionary advantage due to enhanced transmission, pathogenesis, or drug resistance.^{16,17}

Although distinct CRFs continue 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 these databases are not representative of populations and therefore not suitable for epidemiological studies.¹³ Published studies often report representative samples, but are limited by geography, publication bias, and time delay between sampling and publication, leading

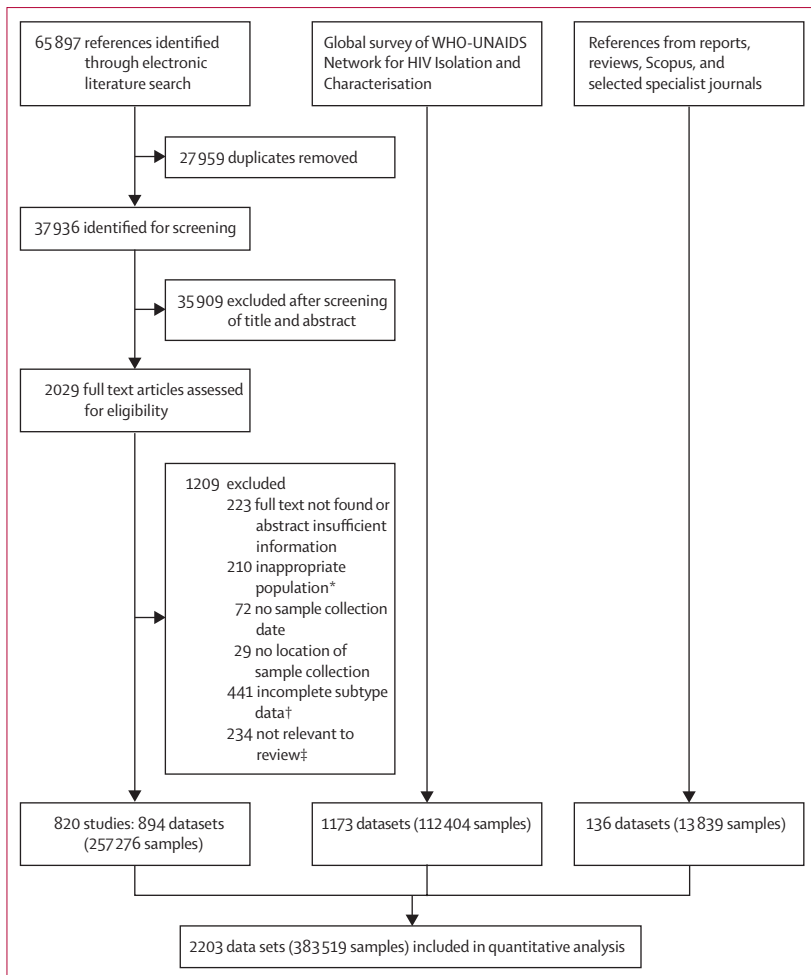


Figure 1: Study flow diagram

*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.

to incomplete coverage.¹⁷ 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.¹⁸ 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 Emtree 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).¹⁸

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 dataset.

Data analysis

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

	Caribbean	Latin America	Western and central Europe & North America	Eastern Europe & central Asia	South Asia (India)	Southeast Asia	East Asia	Oceania	Middle East & north Africa	West Africa	East Africa	Ethiopia	Central Africa	Southern Africa	Global
Number of people living with HIV															
1990-99	270 647	777 108	1 213 002	1 853 25	917 464	773 999	208 047	25 114	24 066	2 991 282	3 933 298	720 140	755 554	5 550 537	18 345 584
2000-04	422 471	1 106 688	1 625 178	5 621 73	2 185 896	1 329 216	5 41 451	47 847	79 018	4 920 236	4 412 631	1 038 564	1 119 699	11 081 253	30 472 321
2005-09	366 821	1 327 862	1 901 896	8 41 816	2 226 780	1 609 187	6 90 389	56 252	125 841	5 140 275	4 269 141	794 908	1 173 716	11 690 044	32 214 928
2010-15	295 748	1 586 605	2 236 833	1 256 028	2 136 545	1 893 512	810 004	65 354	171 944	5 419 010	4 704 986	707 751	1 143 531	12 493 788	34 921 639
Proportion of people infected with recombinants, n (%)															
1990-99	39 973 (14.8%)	24 575 (3.2%)	39 979 (3.3%)	97 89 (5.3%)	3 197 (0.3%)	5 850 18 (75.6%)	56 838 (27.3%)	1 998 (8.0%)	21 49 (8.9%)	418 266 (14.0%)	394 807 (10.0%)	0	124 571 (16.5%)	5 455 (0.1%)	17 066 614 (9.3%)
2000-04	2 363 (0.6%)	78 942 (7.1%)	78 801 (4.8%)	17 572 (3.1%)	14 792 (0.7%)	1 112 576 (83.7%)	334 592 (61.8%)	997 (2.1%)	6 061 (7.7%)	2 910 591 (59.2%)	908 038 (20.6%)	0	339 418 (30.3%)	210 402 (1.9%)	6 015 145 (19.7%)
2005-09	2 697 (0.7%)	168 641 (12.7%)	123 065 (6.5%)	19 081 (2.3%)	20 211 (0.9%)	1 449 933 (90.1%)	408 073 (59.1%)	4 455 (7.9%)	6 381 5 (50.7%)	3 222 035 (62.7%)	613 868 (14.4%)	902 4 (1.1%)	556 940 (47.5%)	40 916 (0.4%)	6 702 755 (20.8%)
2010-15	19 935 (6.7%)	193 025 (12.2%)	171 454 (7.7%)	298 199 (23.7%)	66 497 (3.1%)	1 516 285 (80.1%)	652 181 (80.5%)	9 412 (14.4%)	116 491 (67.7%)	370 624 6 (68.4%)	654 760 (13.9%)	0	534 997 (46.8%)	39 035 (0.3%)	7 978 517 (22.8%)
Proportion of people infected with URFs, n (%)															
1990-99	34 976 (12.9%)	11 878 (1.5%)	8 020 (0.7%)	15 88 (0.9%)	3 197 (0.3%)	23 889 (3.1%)	2 928 (1.4%)	0	21 20 (8.8%)	103 098 (3.4%)	391 946 (10.0%)	0	66 750 (8.8%)	5 455 (0.1%)	65 584 6 (3.6%)
2000-04	408 (0.1%)	62 282 (5.6%)	17 425 (1.1%)	27 60 (0.5%)	10 900 (0.5%)	77 479 (5.8%)	36 065 (6.7%)	43 (0.1%)	903 (1.1%)	664 485 (13.5%)	884 570 (20.0%)	0	173 383 (15.5%)	102 095 (0.9%)	2 032 799 (6.7%)
2005-09	533 (0.1%)	140 881 (10.6%)	32 591 (1.7%)	0	17 484 (0.8%)	132 576 (8.2%)	16 981 (2.5%)	566 (1.0%)	5 494 (4.4%)	373 463 (7.3%)	549 134 (12.9%)	4 512 (0.6%)	298 730 (25.5%)	25 382 (0.2%)	1 598 327 (5.0%)
2010-15	8 498 (2.9%)	153 069 (9.6%)	43 229 (1.9%)	14 896 (1.2%)	66 497 (3.1%)	111 004 (5.9%)	40 661 (5.0%)	11 066 (1.7%)	6 46 (0.4%)	838 476 (15.5%)	591 140 (12.6%)	0	243 041 (21.3%)	22 142 (0.2%)	2 134 405 (6.1%)
Proportion of people infected with CRFs, n (%)															
1990-99	4 997 (1.8%)	6 882 (0.9%)	31 959 (2.6%)	8 200 (4.4%)	0	5 611 29 (72.5%)	53 909 (25.9%)	1 998 (8.0%)	28 (0.1%)	315 168 (10.5%)	1 126 (0.0%)	0	57 821 (7.7%)	0	1 043 216 (5.7%)
2000-04	1 955 (0.5%)	17 508 (1.6%)	61 390 (3.8%)	14 812 (2.6%)	3 893 (0.2%)	1 037 247 (78.0%)	298 527 (55.1%)	953 (2.0%)	5 159 (6.5%)	2 246 105 (45.7%)	23 468 (0.5%)	0	166 035 (14.8%)	108 307 (1.0%)	3 985 359 (13.1%)
2005-09	2 164 (0.6%)	28 716 (2.2%)	90 475 (4.8%)	19 081 (2.3%)	2 727 (0.1%)	1 318 512 (81.9%)	391 146 (56.7%)	3 889 (6.9%)	58 320 (46.3%)	2 858 574 (55.6%)	74 494 (1.7%)	4 512 (0.6%)	258 210 (22.0%)	15 534 (0.1%)	5 126 354 (15.9%)
2010-15	11 438 (3.9%)	39 956 (2.5%)	128 225 (5.7%)	283 304 (22.6%)	0	1 405 281 (74.2%)	611 520 (75.5%)	8 306 (12.7%)	115 845 (67.4%)	2 867 770 (52.9%)	63 620 (1.4%)	0	291 955 (25.5%)	16 894 (0.1%)	5 844 113 (16.7%)
Number of distinct CRFs															
1990-99	2	5	25	3	0	10	5	1	1	9	1	0	9	0	29
2000-04	6	12	38	4	2	11	11	3	4	12	6	0	12	9	43
2005-09	6	12	40	3	2	18	17	7	8	16	9	1	18	4	54
2010-15	6	14	44	4	0	8	10	11	10	19	5	0	23	5	57
1990-2015	8	19	45	7	2	18	18	12	12	21	13	1	27	14	58

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

Table: Global and regional proportions and numbers of recombinants in 1990-2015

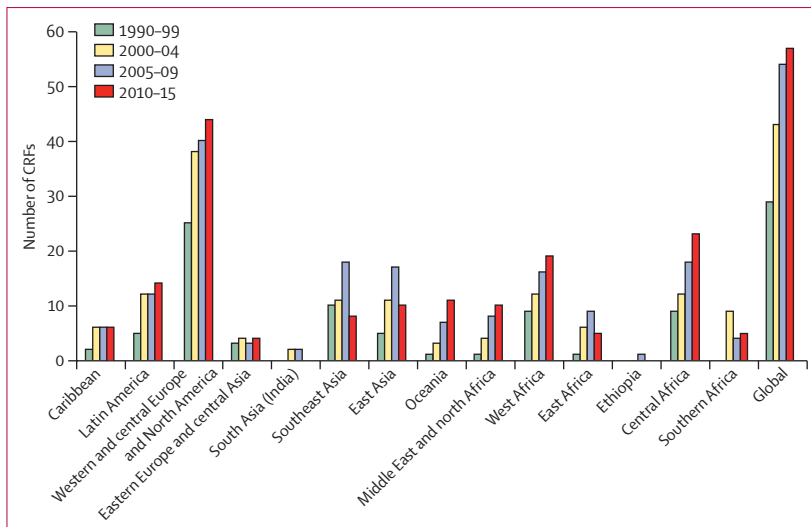


Figure 2: Regional and global numbers of CRFs in 1990–2015
CRF=circulating recombinant form.

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, 1173 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 2203 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 (ten), Middle East and north Africa (ten), 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% (1838 433 of 7978 517) of global recombinants, CRF02_AG contributed 33.9% (2701 364), and URFs contributed 26.7% (2131 450). The proportion of other CRFs increased over time and was 16.4% (1307 270) in 2010–15. Other major CRFs in 2010–15 were CRF63_02A1 (3.3%; 264 805 of 7978 517 global recombinants), CRF06_cpx (2.8%; 222 562), CRF07_BC

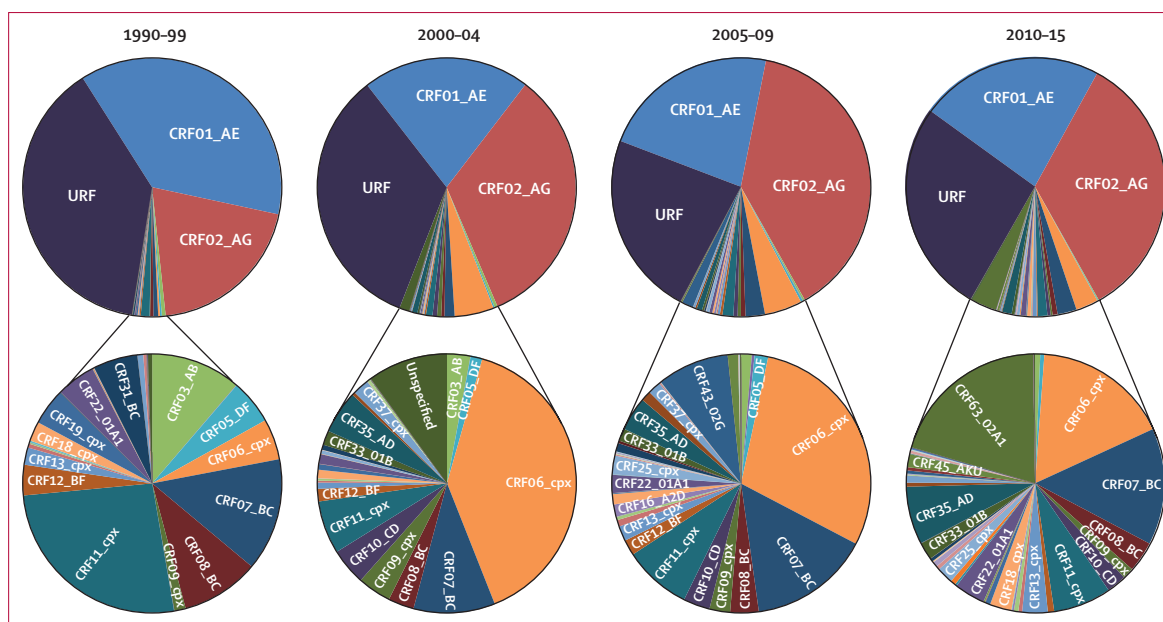


Figure 3: Global distribution of recombinants in 1990–2015

The top row shows CRFs and URFs as proportions of all recombinants globally. The bottom row shows proportions of CRFs other than CRF01_AE and CRF02_AG (appendix pp 11–15). CRF=circulating recombinant form. URF=unique recombinant form.

(2.4%; 189882), CRF11_cpx (1.2%; 94729), and CRF35_AD (1.2%; 94245).

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%; 66497 of 66497 recombinants), east Africa (90.3%; 591140 of 654760), Latin America (79.3%; 153069 of 193025), southern Africa (56.7%; 22142 of 39035), central Africa (45.4%; 243041 of 534997), and the Caribbean (42.6%; 8498 of 19935). CRF01_AE contributed the largest proportion of recombinants in southeast Asia (90.3%; 1369161 of 1516285), Oceania (74.5%; 7010 of 9412), and east Asia (58.6%; 381996 of 652181). CRF02_AG contributed the largest proportion in west Africa (67.6%; 2504438 of 3706246) and western and central Europe and North America (39.4%; 67546 of 171454).

In the Caribbean, the main CRFs in 2010–15 were CRF19_cpx (24.8%; 4935 of 19935 recombinants), CRF23_BG (17.2%; 3424) and CRF18_cpx (10.5%; 2089; figure 4B, appendix pp 16–23). In Latin America, CRFs included CRF07_BC (6.1%; 11857 of 193025) and CRF12_BF (2.9%; 5630). In Western and central Europe and North America, the other main CRFs were CRF06_cpx (5.7%; 9729 of 171454), CRF12_BF (2.8%; 4788), CRF14_BG (3.6%; 6144), and CRF28_BF (2.7%; 4684). CRF63_02A1 contributed 88.9% (265172 of 298199) of recombinants in eastern Europe and central Asia and CRF35_AD contributed 79.1% (92187 of 116491) of recombinants in the Middle East and north Africa. In east Asia, CRF07_BC (27.2%; 177541 of 652181) and CRF08_BC (7.1%; 46090) were prominent. CRF06_cpx (5.5%; 201947 of 3706246) was the other

most common CRF in west Africa, whereas it was CRF10_CD in east Africa (3.8%; 24995 of 654760) and southern Africa (18.7%; 7285 of 39035). In central Africa, the other main CRFs were CRF11_cpx (8.7%; 46304 of 534997), CRF13_cpx (6.0%; 32017), CRF18_cpx (4.7%; 25311), CRF25_cpx (4.5%; 24249), and CRF45_AKU (3.9%; 20672).

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 globally (<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.

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.²⁰ 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.²¹ 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.²² 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.¹ 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%).¹ 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.²³ 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.²⁴

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 383 519 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.¹⁸ 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.¹⁸ 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.¹⁸

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.¹³ 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.^{25,26} 58 distinct CRFs were identified in the epidemiological studies that contributed to our study, although 106 CRFs have been described to date.¹² 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.¹² Finally, we did not have actual viral sequence data and consequently we were unable to do phylogenetic, phylodynamic, or phylogeographic analyses.²⁵

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.^{1,27} Ongoing recombination could lead to the generation of new variants that viral assays do not detect or detect less efficiently.²⁷ 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.^{28,29}

Lastly, with the continued recombination of HIV variants, a risk of the generation of multidrug-resistant HIV viruses exists.¹⁶ Diagnostic, resistance, and viral load assays clearly need to be continuously adapted to the evolving HIV epidemic.

For UNAIDS AIDS info see
<http://aidsinfo.unaids.org>

The diversification of the HIV pandemic is a major challenge to the development of an HIV vaccine.⁴ 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.¹⁵ With the continuing generation of new recombinants, including second-generation recombinants (ie, recombinants of recombinants), this challenge becomes ever 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.³⁰ Unfortunately, the trial of this vaccine was halted due to lack of efficacy.³¹ 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,^{32,33} mosaic and polyvalent vaccines,³⁴ and focusing on conserved or structurally important regions of HIV.^{35,36} 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 recombinant 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

- UNAIDS. Global AIDS update 2020. Geneva: Joint United Nations Programme on HIV/AIDS, 2020.
- Hemelaar J. The origin and diversity of the HIV-1 pandemic. *Trends Mol Med* 2012; **18**: 182–92.
- Hemelaar J. Implications of HIV diversity for the HIV-1 pandemic. *J Infect* 2013; **66**: 391–400.
- Barouch DH, Korber B. HIV-1 vaccine development after STEP. *Annu Rev Med* 2010; **61**: 153–67.
- Worobey M, Gemmel M, Teuwen DE, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* 2008; **455**: 661–64.
- Robertson DL, Anderson JP, Bradac JA, et al. HIV-1 nomenclature proposal. *Science* 2000; **288**: 55–56.
- 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**: 319–22.
- 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**: 56–61.
- Tebit DM, Arts EJ. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infect Dis* 2011; **11**: 45–56.
- Vuilleumier S, Bonhoeffer S. Contribution of recombination to the evolutionary history of HIV. *Curr Opin HIV AIDS* 2015; **10**: 84–89.
- Robertson DL, Sharp PM, McCutchan FE, Hahn BH. Recombination in HIV-1. *Nature* 1995; **374**: 124–26.
- Zhou J, Li M, Min C, 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; published online Aug 19. <https://doi.org/10.1089/AID.2020.0101>.
- Los Alamos National Laboratory. HIV circulating recombinant forms (CRFs). March, 2020. <http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html> (accessed Jan 15, 2020).
- Lau KA, Wong, JLL. Current trends of HIV recombination worldwide. *Infect Dis Rep* 2013; **5** (suppl 1): e4.
- 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.
- Spira S, Wainberg MA, Loemba H, Turner D, Brenner BG. Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. *J Antimicrob Chemother* 2003; **51**: 229–40.
- Ariën KK, Vanham G, Arts EJ. Is HIV-1 evolving to a less virulent form in humans? *Nat Rev Microbiol* 2007; **5**: 141–51.
- 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 Infect Dis* 2019; **19**: 143–55.
- 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**: 40.
- 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. *J Virol* 2008; **82**: 9206–15.
- Memarnejadian A, Menbari S, Mansouri SA, et al. Transmitted drug resistance mutations in antiretroviral-naïve injection drug users with chronic HIV-1 infection in Iran. *PLoS One* 2015; **10**: e0126955.
- 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**: 415–20.

- 23 Beloukas A, Psarris A, Giannelou P, Kostaki E, Hatzakis A, Paraskevis D. Molecular epidemiology of HIV-1 infection in Europe: an overview. *Infect Genet Evol* 2016; **46**: 180–89.
- 24 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. *Infect Genet Evol* 2016; **46**: 200–08.
- 25 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.
- 26 Aldrich C, Hemelaar J. Global HIV-1 diversity surveillance. *Trends Mol Med* 2012; **18**: 691–94.
- 27 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. *J Clin Microbiol* 2018; **56**: e02045–17.
- 28 Gregson J, Tang M, Ndemi N, et al. Global epidemiology of drug resistance after failure of WHO recommended first-line regimens for adult HIV-1 infection: a multicentre retrospective cohort study. *Lancet Infect Dis* 2016; **16**: 565–75.
- 29 Stanford University. HIV drug resistance database. 2020. <https://hivdb.stanford.edu> (accessed Jan 15, 2020).
- 30 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; **5**: e366–78.
- 31 Adepoju P. Moving on from the failed HIV vaccine clinical trial. *Lancet HIV* 2020; **7**: e161.
- 32 Gaschen B, Taylor J, Yusim K, et al. Diversity considerations in HIV-1 vaccine selection. *Science* 2002; **296**: 2354–60.
- 33 Nickle DC, Jensen MA, Gottlieb GS, et al. Consensus and ancestral state HIV vaccines. *Science* 2003; **299**: 1515–18.
- 34 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.
- 35 Létourneau S, Im EJ, Mashishi T, et al. Design and pre-clinical evaluation of a universal HIV-1 vaccine. *PLoS One* 2007; **2**: e984.
- 36 Gaiha GD, Rossin EJ, Urbach J, et al. Structural topology defines protective CD8⁺ T cell epitopes in the HIV proteome. *Science* 2019; **364**: 480–84.