



Published in final edited form as:

AIDS. 2009 September 10; 23(14): F9–F17. doi:10.1097/QAD.0b013e32832faf61.

Spatial phylodynamics of HIV-1 epidemic emergence in east Africa

Rebecca R. Gray^a, Andrew J. Tatem^{b,c}, Susanna Lamers^d, Wei Hou^e, Oliver Laeyendecker^{f,g}, David Serwadda^{h,i}, Nelson Sewankambo^{h,j}, Ronald H. Gray^k, Maria Wawer^k, Thomas C. Quinn^{f,g}, Maureen M. Goodenow^a, and Marco Salemi^a

^aDepartment of Pathology, Immunology, and Laboratory Medicine, University of Florida, Gainesville, Florida, USA ^bDepartment of Geography and Emerging Pathogens Institute, University of Florida, Gainesville, Florida, USA ^cMalaria Public Health & Epidemiology Group, Centre for Geographic Medicine, KEMRI/Wellcome Trust Research Laboratories, Nairobi, Kenya ^dBiInfoExperts Inc. ^eDivision of Biostatistics, Department of Epidemiology and Health Policy Research, University of Florida, Gainesville, Florida ^fDepartment of Medicine, Johns Hopkins Medical Institutions, Baltimore ^gNational Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA ^hRakai Health Sciences Program, Kalisizo, Rakai District ⁱMakerere University School of Public Health, College of Health Sciences, Kampala, Uganda ^jMakerere University, Faculty of Medicine, College of Health Sciences, Kampala, Uganda ^kDepartment of Population, Family and Reproductive Health, Johns Hopkins University, Baltimore, Maryland, USA

Abstract

Design—We sought to investigate the evolutionary and historical reasons for the different epidemiological patterns of HIV-1 in the early epidemic. In order to characterize the demographic history of HIV-1 subtypes A and D in east Africa, we examined molecular epidemiology, geographical and historical data.

Methodology—We employed high-resolution phylodynamics to investigate the introduction of HIV-1A and D into east Africa, the geographic trends of viral spread, and the demographic growth of each subtype. We also used geographic information system data to investigate human migration trends, population growth, and human mobility.

Results—HIV-1A and D were introduced into east Africa after 1950 and spread exponentially during the 1970s, concurrent with eastward expansion. Spatiotemporal data failed to explain the establishment and spread of HIV based on urban population growth and migration. The low prevalence of the virus in the Democratic Republic of Congo before and after the emergence of the pandemic was, however, consistent with regional accessibility data, highlighting the difficulty in travel between major population centers in central Africa. In contrast, the strong interconnectivity between population centers across the east African region since colonial times has likely fostered the rapid growth of the epidemic in this locale.

Conclusion—This study illustrates how phylodynamic analysis of pathogens informed by geospatial data can provide a more holistic and evidence-based interpretation of past epidemics. We advocate that this ‘landscape phylodynamics’ approach has the potential to provide a framework both to understand epidemics’ spread and to design optimal intervention strategies.

Correspondence to Marco Salemi, Department of Pathology, Immunology, and Laboratory Medicine, University of Florida College of Medicine, 1376 Mowry Road, Gainesville, FL 32610, USA. Tel: +1 352 273 8164; fax: +1 352 273 8284; salemi@pathology.ufl.edu.

Author contributions: R.R.G. and M.S. conceived of the project, implemented the phylogenetic analyses, and wrote the manuscript. A.S. performed all GIS analyses. S.L. performed genetic analyses. W.H. performed statistical analyses. R.H.G., M.W., D.S., and N.S., T.C.Q., and M.M.G. assisted in reviewing results and provided historical information. T.C.Q. and M.M.G. assisted in writing the manuscript.

Keywords

Africa; epidemic; evolution; geographic information system; HIV-1; phylodynamics

Introduction

HIV-1 has been one of the most successful pathogens in human history, having infected thus far more than 60 million people [1]. HIV-1 comprises three groups (M, N, and O), each introduced through a separate zoonotic transmission of SIV infecting *Pan troglodytes troglodytes* [2], possibly from a band of chimpanzees in Cameroon [3]. Group M accounts for more than 95% of all HIV-1 infections and comprises nine subtypes and multiple circulating recombinant forms [4]. The origin of group M has been dated using phylogenetic techniques to 1921 [95% confidence interval (CI) 1908–1933] [5–7]. For the next 50 years, HIV-1 remained confined in west/central Africa, where the circulating virus accumulated considerable diversity by 1960 [5], and where the highest diversity now exists [8]. The virus was only detected in east Africa in the 1970s [9] and in Europe and North America by the 1980s, although the actual introduction probably occurred decades earlier [10]. The global spread of HIV-1 during the latter half of the 20th century is commonly attributed to increased migration between countries as a consequence of globalization; increasing urbanization; warfare and ethnic conflict surrounding postcolonization independence of many African countries; and changing sexual practices [11]. Although the relative contribution of these factors, in addition to founder effects [12], has certainly played an important role in the diffusion of the pandemic, the exact causes for the uneven spread of HIV-1 subtypes and the wide range of prevalence rates are still largely unknown.

East Africa was one of the first localities affected by the virus after the origin and diversification in west/central Africa [9], but epidemiological evidence indicates a markedly different epidemic history between the two regions. During the 1990s and 2000s, prevalence levels in the Democratic Republic of Congo (DRC) remained at only approximately 5%, yet reached well above 15% in Kenya and peaked above 30% in sentinel populations in Uganda [13–16]. Moreover, in contrast to the high viral diversity in west-central Africa, subtypes A and D account for the majority of the circulating viruses in east Africa, although they differ in frequency between regions [17,18]. We characterized the evolutionary history of these two subtypes and investigated the role that migration, urbanization, and infrastructure played in shaping the epidemic. Our approach was based on phylodynamics, an analytical method incorporating coalescent theory to infer pathogen genealogies [19], as well as geographic information system (GIS) data. This approach allowed us to examine the unresolved questions of the causes of the first spread of HIV-1 out of central Africa through a more comprehensive analysis than previous work, by incorporating evolutionary, epidemiological, as well as geopolitical factors.

Materials and methods

Dataset selection

We mined the Los Alamos HIV databases (<http://www.hiv.lanl.gov/content/index> and <http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html>) to compile datasets for subtypes A and D. To infer a reliable demographic history of HIV-1, we needed to select genetic regions with sufficient phylogenetic information. As such, we used the following criteria based on analysis of both real and simulated datasets [20,21]: first, overall mean distance more than 5%; second, phylogenetic noise less than 20% using likelihood mapping [20]; and third, length greater than 250nt [21]. The only regions that met our criteria were p24 and gp41. In particular, sequences of the C2V3 domain of gp120, although more numerous, failed to meet the second

criterion by displaying significantly stronger phylogenetic noise (21%) than p24 (9%) or gp41 (6%) datasets, and were on average too short (third criterion). Additionally, the available sequences had to satisfy the following criteria: sequences had already been published in peer-reviewed journals (except for the new Uganda cohort described below); there was no uncertainty about the subtype assignment of each sequence and sequences were classified as non-recombinant; sequences were not epidemiologically linked; only one sequence per individual could be randomly selected; and country of origin and sampling date were known and clearly established in the original publication. We also included sequences sampled in 1994–2002 from epidemiological unlinked participants living in the Rakai District Uganda (see Supplementary materials and methods for details), which have been deposited in GenBank. (Accession numbers GQ332766-GQ334183, GQ253666- GQ253897, GQ252692-GQ252793).

Bayesian inference of demographic history

Inference of phylogenetic trees, nonparametric estimates of effective population size (N_e) over time for the major east African clades, and logistic growth rates and nucleotide substitution rates were obtained using the Bayesian framework implemented in BEAST [22] (see Supplementary materials and methods for details). For each genealogy, ancestral states (i.e., the putative geographic origin of early lineages in the genealogy) were inferred with the maximum parsimony algorithm implemented in MacClade [23] (see Supplementary materials and methods for details).

Geographic information system data analysis

The urban population change data were calculated from the HYDE database (<http://www.mnp.nl/en/themasites/hyde/index.html>) version 3. The 1950, 1960, 1970, 1980, and 1990 urban population data were acquired, and for each dataset, the percentage changes in populations between decades were calculated and mapped for each approximately 10 km spatial resolution urban grid cell. Data on net migration rates were obtained from the United Nations World Population Prospects (United Nations Population Division, 2006) and are derived from various sources, including national figures on the number of immigrants and emigrants, estimates of net migration, estimates of international labor migration, and refugee stock data. Additional studies were aggregated from a combination of sources including the Africover Initiative (FAO-UN) and publicly available data provided by ESRI (<http://www.esri.com>). These data were then layered in ARCGIS software and used to form the geo-spatial foundation of the model presented (see Supplementary materials and methods for details).

Results

We inferred Bayesian genealogies for HIV-1 A and D p24 and gp41 genes using publicly available sequence data sampled from east, west, and central Africa (selected according to specific criteria discussed in the Supplementary methods) from the Los Alamos HIV databases (<http://www.hiv.lanl.gov>), as well as new sequence data from the Rakai district of Uganda (Table S1). We also included reference sequences from subtypes C, D, G, H, J, and K as outgroups. All sequences were subtyped using the REGA subtyping tool [24] and manually checked by obtaining neighbor-joining phylogenies including additional reference sequences. Bayesian genealogies were generated according to constant, exponential, logistic, and Bayesian skyline plot (BSP) demographic models under either a strict or relaxed molecular clock. The relaxed molecular clock always performed significantly better than the strict clock (Bayes factors >20), and the models allowing population growth were consistently better supported than the constant size model (data not shown). The HIV-1A genealogies for both genes and all demographic models showed, as expected, two well supported monophyletic

clades ($P = 1.0$) within subtype A clustering A1 and A2 sequences, respectively. For both genes, all of the known sequences from the DRC were part of the A2 clade. In the p24 relaxed clock tree with BSP prior (Fig. 1a, fully annotated tree in S1A), east African sequences were also monophyletic, although the support was only moderate ($P = 0.62$) and the clade was not present under all demographic models (data not shown). A few sequences from west Africa appeared to be intermixed in the east Africa clade in the gp41 tree relaxed clock with BSP prior (Fig. 1b, S1B). In the genealogies for subtype D (Fig. S1c and d), all of the sequences from east Africa clustered together as did the sequences from west Africa. Although these clades were weakly supported ($P < 0.5$), they were present under most demographic models. Overall, the genealogies were consistent with a model of limited introduction of each subtype into east Africa, followed by subsequent expansion. The putative geographic origin for the internal nodes of each tree was inferred by a weighted maximum parsimony reconstruction of ancestral states. In both p24 and gp41 trees, the earliest internal nodes in the A1 clade were assigned to Uganda (branches in red, Fig. 1a and b) and arose at least one decade prior to the Kenya strains. Even if a sampling bias cannot be excluded, a similar number of strains were sampled from Uganda and Kenya for both genes, strengthening our confidence in the result. In all genealogies, more than 90% of the lineages leading to the east African sequences were already present by 1980 (Fig. 1c and d).

The median date of the time of the most recent common ancestor (TMRCA) for the root of each tree estimated using the relaxed clock assumption and the BSP prior was between 1915 and 1922 (Table S2), which is consistent with the origin of the group M HIV-1 epidemic as previously reported [5–7]. The origin of the HIV-1 A1 clade was estimated as 1948 for both genes. The TMRCA for the east Africa subtype D clade, which contained more than 90% of the east African sequences, was estimated between 1958 (gp41) and 1967 (p24). Large confidence intervals accompany these estimates; however, the median dates were similar among the different demographic models tested (data not shown).

In order to investigate the demographic history of the subtypes, we compared a constant model of growth with the exponential model and a nonparametric BSP model for either the A1 clade (HIV-1A) or the east African subtype D clade (HIV-1D). In all cases, the Bayesian model allowing population growth was a better fit to the data than the constant model (Table S3). For all genes, the high posterior densities (HPDs) for the exponential growth rate were greater than one, suggesting further evidence for a model of growth best fitting the data. We then considered the estimates of effective number of infectious individuals over time obtained from the BSP model. For each subtype, the BSP from p24 (Fig. 2a–b) and gp41 (Fig. 2c–d) appeared similar: all datasets were characterized by a low number of effective infections during the initial epidemic, consistent with the limited introduction model discussed in the previous sections, and an exponential growth phase during the 1970s, followed by a leveling off period during the mid-1980s/early 1990s. These trend and timing were consistent among several different root priors and for both the full A1 clade as well as just the east African sequences for subtype A (data not shown).

Net migration rates for east and central African countries between 1955 and 2000 were plotted (Fig. 3; raw data are given in Table S4). In general, we did not observe any specific trend that correlated with the uneven spread of HIV-1 subtypes and wide range of viral prevalence rates in central-east Africa. No country showed a net migration influx higher than five of 1000 people during the time of the initial spread and expansion of the epidemic (1960–1985). Kenya experienced net migration around zero, whereas the DRC, Sudan, and Tanzania did not exceed ± 5 migrants/1000 until 1990. During the 1970s, a modest negative migratory outflow ($-5/1000$) was observed in Uganda, occurring during the coup by Idi Amin and subsequent civil war. Rwanda exhibited a sharp negative outflow ($-10/1000$) during the 1960s, coinciding with ethnic conflict and the war of independence, but a dramatic decrease in migration during the

first half of the 1990s ($>-50/1000$), coinciding with the ethnic genocide that occurred in 1994, followed again by an increase in the second half of the decade.

Changes in population density in the region were examined using a census-based spatially referenced population database [25]. Difference maps were created to investigate how population distribution changed between decades. Although these maps are derived from different resolutions of input census data for each country and thus should be interpreted with caution, the overall trends in population distribution changes are clear. Figure 4 highlights how strong population growth occurred in regions across northern Tanzania, Uganda, and the most densely populated areas of Kenya in the 1960s. At the same time, northern DRC and the more rural areas of Kenya exhibited overall declines in population. The 1970s were characterized by a mixed pattern, with large areas of Uganda and Tanzania actually experiencing declines in population numbers, whereas other countries showed strong growth. Kenya consistently showed high population growth in the most densely populated areas.

Accessibility maps estimate the travel time to the nearest major city (defined as those with a population greater than 500 000 in 2000), using road/track-based travel (Fig. 5a) [26]. Although based principally on 1990s data due to the sparseness of earlier comparable datasets, the east African road network architecture has changed little since colonial times [27]. Accessibility was computed using a cost–distance algorithm that calculates the ‘cost’ of traveling between two locations on a regular raster grid [28]. The DRC is characterized by several isolated networks of relatively high accessibility, centered on the three major urban areas of Kinshasa in the west, Kisangani in the north, and Kananga, Likasi, and Lubumbashi in the south. Distinct routes of travel are evident in each network, and accessibility rapidly declines with distance from the hub cities. Separating these major networks are hundreds of square miles with poor access to major population centers. The western city of Kinshasa has connectivity with Luanda; however, the southern end of this network is also completely isolated. The northern city of Kisangani shows very minor connections with east Africa, where the major population hubs are Kampala, Nairobi, Mombasa, and Dar Es Salaam. In contrast to the pattern in the DRC, transport network isolation in east Africa is not so evident, with the four principal cities and the majority of Uganda and Kenya all relatively well connected and accessible. Moreover, this well connected east Africa network is relatively isolated from both the DRC and the principal population centers of Sudan, Ethiopia, and Somalia to the north and east. The southern DRC city of Lubumbashi has a small connection with a separate network that extends further south. Major roadways are superimposed on the map and reinforce the connectivity patterns.

We then plotted the distribution of HIV-1 subtypes A, C, D, and other subtypes as reported by numerous sources (Fig. 5b) [29–37]. Three major cities in the DRC have very different distributions of subtypes. Kinshasa has very high proportion of ‘other’, similar to the situation in Luanda. Subtype A is present at nearly 40% in Kisangani, which accounts for well over 50% in Kigali, Kampala, and Nairobi. Subtype D has varying proportions in the east African cities, but is much higher there than in western or southern Africa. In Dar Es Salaam, part of the east Africa network, the proportion of subtype A decreases to nearly 25%, whereas in Moshi, which is not connected in this network, subtype C is the most frequent. Subtype C is at a proportion of greater than 50% in Lubumbashi in the southern DRC, and is nearly 100% in Lusaka. Thus, the distribution of HIV-1 subtypes appears to closely mirror the accessibility networks as well.

Discussion

The temporal and geographic origins of HIV-1 group M have been well characterized [5–7, 38], although the reasons for the uneven distribution of subtypes and prevalence rates remain unclear. We characterized the evolutionary history of subtypes A and D in east Africa and investigated the role migration, population growth, and infrastructure played in shaping the

epidemic in this region as compared with the DRC. The HIV-1 epidemic has a divergent history in these two regions. The prevalence of HIV-1 in the DRC, the source of the pandemic, reached only approximately 5% during the 1980s and 1990s, even in the capital city of Kinshasa [14, 39]. In contrast, prevalence rates reached 20 and 30% in sentinel populations of Kenya [15] and Uganda [16], respectively, and had already reached 61% in prostitutes in Nairobi, Kenya, by 1985 [40]. The results from our analysis indicate that subtype A likely entered east Africa after 1950 (time of origin of the A1 clade, which includes noneast African sequences) and subtype D entered after 1960. Both subtypes grew exponentially during the 1970s, which is consistent with the high rate of infections observed by the early 1980s.

In order to determine why east Africa experienced much higher prevalence than the DRC, we employed GIS to investigate migration rates and population growth in central/east Africa. We found that net migration rates were stable from 1955 to 1990 in both the DRC and Kenya, whereas a slight decline was observed in Uganda. High values of net out-migration were found for Rwanda, which coincided with periods of ethnic conflict. However, no differences were observed that could account for the low prevalence in the DRC in contrast to Uganda and Kenya. We then investigated the population growth in this region between 1960–1970 and 1970–1980. Here again, we found no consistent trends among the countries that could account for the discrepancy in prevalence rates. In both decades, all countries experienced growth and decline in population size in different regions, although the area around Lake Victoria did experience high growth during both decades.

The investigation of accessibility to and among urban centers in central-east Africa revealed instead an interesting pattern. We found that the three major urban centers of the DRC (Kinshasa, Lubumbashi, and Kisangani) are isolated from one another, but each do share a small corridor of connectivity with another region of Africa (Angola, Zambia, and east Africa, respectively). The distribution of subtypes in west, south, and east Africa is similar to the distribution of the DRC city with which they share a network: high diversity is found in Kinshasa and western Africa, subtype C is highest in Lubumbashi and southern Africa, and subtype A is highest in Kisangani and east Africa. Thus, the unequal spread of subtypes in Africa does appear to reflect founder effect as well as the distribution in the DRC urban center from which a particular wave of infection originated.

The relative isolation of the DRC population centers, which were among the first in which the epidemic was detected initially [11], likely served as a protection against a more widespread epidemic for many decades and explains the low prevalence levels in this country. However, once the infection entered the east African region, the presence of well connected cities fostered the rapid spread of the virus throughout the region. The major highways likely served as a transit route with groups such as mobile prostitutes and their clients [40], soldiers and truck drivers, introducing the virus into new networks and villages [11]. Furthermore, the relative ease of mobility within east Africa was fostered by the East African Community, which was established following independence and provided freedom of movement between Uganda, Kenya, and Tanzania until its dissolution in 1977 [41]. The virtual absence of both subtypes A and D in Ethiopia, where HIV-1C accounts for almost 99% of the infections, is consistent with the relative inaccessibility between the principal population centers of east Africa and Ethiopia.

Warfare is commonly cited as a reason for the spread of the epidemic (as reviewed in [42]). However, this correlation is not seen in this region, as both Angola and Uganda suffered ethnic war during this time, and yet prevalence rates in Angola are still less than 5% [32]. A recent meta-analysis of published HIV-1 prevalence statistics in regions of conflict suggests that warfare does not appear to correlate with an increase in HIV-1 prevalence [42]. Although this result is perhaps counterintuitive, the authors of that study note that because the circumstances

surrounding various conflict situations are often very different, each should be considered independently. Our own analysis also suggests a more nuanced correlation between dissemination of HIV-1 and war, in which conflict may be insufficient for facilitating the spread of a new epidemic, but can provide favorable circumstances in an area of high connectivity.

In conclusion, we have described how molecular epidemiology informed by geographical and demographic data can shed light on the evolution of infectious diseases. Our data highlight that accessibility (and the infrastructure providing the ease of travel) is likely to be a critical factor in the formation and spread of new epidemics, whereas factors such as migration, population growth, and warfare seem to contribute marginally. Additional factors, such as ethnicity, tribal affiliations, behavior, and host genetics may also contribute to the spread of subtypes worldwide, and should be incorporated in future studies. Moreover, because the number of sequences in our datasets was limited, it will be important to investigate further our model and conclusions by including new sequences from additional countries as they become available. The interdisciplinary framework we employed demonstrates the power of phylodynamics informed by spatiotemporal demographic and GIS data to analyze the ways in which humans map themselves to the territory and the resulting effect on infectious disease epidemics. Further development of this 'landscape phylodynamics' approach, including the implementation of statistical tests correlating phylodynamic and GIS data, has the potential to provide a better understanding of epidemic spread and to inform the design of optimal intervention strategies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported in part by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, NIH. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract N01-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. M.S. was supported by NIH award number 00069244 and the Experimental Pathogen Innovative Grant (2008) awarded from UF.

References

- UNAIDS. Report on the global AIDS epidemic: 2008. UNAIDS; 2008
http://data.unaids.org/pub/GlobalReport/2008/jc1510_2008_global_report_pp1_10_en.pdf
- Hahn BH, Shaw GM, De Cock KM, Sharp PM. AIDS as a zoonosis: scientific and public health implications. *Science* 2000;287:607–614. [PubMed: 10649986]
- Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, et al. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* 2006;313:523–526. [PubMed: 16728595]
- Leitner, T.; Korber, B. The Human Retroviruses and AIDS 2005 Compendium. Los Alamos: Theoretical Biology, Biophysics; 2005. HIV-1 subtype and circulating recombinant form (CRF) reference sequences, 2005.
- Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* 2008;455:661–664. [PubMed: 18833279]
- Salemi M, Strimmer K, Hall WW, Duffy M, Delaporte E, Mboup S, et al. Dating the common ancestor of SIVcpz and HIV-1 group M and the origin of HIV-1 subtypes using a new method to uncover clock-like molecular evolution. *FASEB J* 2001;15:276–278. [PubMed: 11156935]
- Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, et al. Timing the ancestor of the HIV-1 pandemic strains. *Science* 2000;288:1789–1796. [PubMed: 10846155]
- Vidal N, Peeters M, Mulanga-Kabeya C, Nzilambi N, Robertson D, Ilunga W, et al. Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic

- Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. *J Virol* 2000;74:10498–10507. [PubMed: 11044094]
9. Saxinger WC, Levine PH, Dean AG, de The G, Lange-Wantzin G, Moghissi J, et al. Evidence for exposure to HTLV-III in Uganda before 1973. *Science* 1985;227:1036–1038. [PubMed: 2983417]
 10. Gilbert MT, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M. The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci U S A* 2007;104:18566–18570. [PubMed: 17978186]
 11. Iliffe, J. *The African AIDS epidemic: a history*. Athens: Ohio University Press; 2006.
 12. Rambaut A, Posada D, Crandall K, Holmes E. The causes and consequences of HIV evolution. *Nat Rev Genet* 2004;5:52–61. [PubMed: 14708016]
 13. UNAIDS. UNAIDS/WHO/UNICEF Epidemiological fact sheets on HIV and AIDS. Geneva: WHO; 2004.
 14. UNAIDS. Geneva: WHO; 2008. Epidemiological Fact Sheet on HIV and AIDS: Democratic Republic of Congo. http://www.who.int/globalatlas/predefinedReports/EFS2008/full/EFS2008_CD.pdf
 15. UNAIDS. Geneva: WHO; 2008. Epidemiological Fact Sheet on HIV and AIDS: Kenya. http://www.who.int/globalatlas/predefinedReports/EFS2008/full/EFS2008_KE.pdf
 16. UNAIDS. Geneva: WHO; 2008. Epidemiological Fact Sheet on HIV and AIDS: Uganda. http://www.who.int/globalatlas/predefinedReports/EFS2008/full/EFS2008_UG.pdf
 17. Lutalo T, Gray RH, Wawer M, Sewankambo N, Serwadda D, Laeyendecker O, et al. Survival of HIV-infected treatment-naïve individuals with documented dates of seroconversion in Rakai, Uganda. *AIDS* 2007;21(Suppl 6):S15–S19. [PubMed: 18032934]
 18. Vasan A, Renjifo B, Hertzmark E, Chaplin B, Msamanga G, Essex M, et al. Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. *Clin Infect Dis* 2006;42:843–852. [PubMed: 16477563]
 19. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol* 2006;4:e88. [PubMed: 16683862]
 20. Strimmer K, von Haeseler A. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc Natl Acad Sci U S A* 1997;94:6815–6819. [PubMed: 9192648]
 21. Felsenstein J. Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? *Mol Biol Evol* 2006;23:691–700. [PubMed: 16364968]
 22. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007;7:214. [PubMed: 17996036]
 23. Maddison, WP.; Maddison, D. *MacClade version 4: analysis of phylogeny and character evolution*. Sunderland, MA: Sinauer Associates; 2003.
 24. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, Seebregts C, et al. An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* 2005;21:3797–3800. [PubMed: 16076886]
 25. Klein Goldewijk, K.; van Drecht, G. HYDE 3: current and historical population and land cover. In: Bouwman, AF.; Kram, R.; Klein Goldewijk, K., editors. *Integrated modelling of global environmental change. An overview of IMAGE 2.4*. Bilthoven: Netherlands Environmental Assessment Agency; 2006.
 26. Nelson, A. Estimated travel time to the nearest city of 50,000 or more people in year 2000. Ispra, Italy: Global Environment Monitoring Unit–Joint Research Centre of the European Commission; 2008.
 27. Porter G. Living in a walking world: rural mobility and social equity issues in sub-Saharan Africa. *Geogr Rev* 2002;30:285–300.
 28. Uchida, H.; Nelson, A. Agglomeration Index: Towards a New Measure of Urban Concentration: Background paper for the World Bank's World Development Report. World Bank; 2009.
 29. Rinke de Wit TF, Tsegaye A, Wolday D, Hailu B, Aklilu M, Sanders E, et al. Primary HIV-1 subtype C infection in Ethiopia. *J Acquir Immune Defic Syndr* 2002;30:463–470. [PubMed: 12154336]
 30. Eshleman SH, Guay LA, Mwatha A, Brown E, Musoke P, Mmiro F, et al. Comparison of mother-to-child transmission rates in Ugandan women with subtype A versus D HIV-1 who received single-

- dose nevirapine prophylaxis: HIV Network For Prevention Trials 012. *J Acquir Immune Defic Syndr* 2005;39:593–597. [PubMed: 16044013]
31. Servais J, Lambert C, Karita E, Vanhove D, Fischer A, Baurith T, et al. HIV type 1 pol gene diversity and archived nevirapine resistance mutation in pregnant women in Rwanda. *AIDS Res Hum Retroviruses* 2004;20:279–283. [PubMed: 15117451]
 32. Yang C, Li M, Mokili JL, Winter J, Lubaki NM, Mwandagalirwa KM, et al. Genetic diversification and recombination of HIV type 1 group M in Kinshasa, Democratic Republic of Congo. *AIDS Res Hum Retroviruses* 2005;21:661–666. [PubMed: 16060838]
 33. Vidal N, Mulanga C, Bazepeo SE, Mwamba JK, Tshimpaka JW, Kashi M, et al. Distribution of HIV-1 variants in the Democratic Republic of Congo suggests increase of subtype C in Kinshasa between 1997 and 2002. *J Acquir Immune Defic Syndr* 2005;40:456–462. [PubMed: 16280702]
 34. Bártolo I, Rocha C, Bartolomeu J, Gama A, Marcelino R, Fonseca M, et al. Highly divergent subtypes and new recombinant forms prevail in the HIV/AIDS epidemic in Angola: new insights into the origins of the AIDS pandemic. *Infect Genet Evol* 2009;9:672–682. [PubMed: 18562253]
 35. Handema R, Terunuma H, Kasolo F, Kasai H, Sichone M, Mulundu G, et al. Emergence of new HIV-1 subtypes other than Subtype C among antenatal women in Lusaka, Zambia. *AIDS Res Hum Retroviruses* 2001;17:759–763. [PubMed: 11429116]
 36. Neilson JR, John GC, Carr JK, Lewis P, Kreiss JK, Jackson S, et al. Subtypes of human immunodeficiency virus type 1 and disease stage among women in Nairobi, Kenya. *J Virol* 1999;73:4393–4403. [PubMed: 10196337]
 37. Harris M, Serwadda D, Sewankambo N, Kim B, Kigozi G, Kiwanuka N, et al. Among 46 near full length HIV type 1 genome sequences from Rakai District, Uganda, subtype D and AD recombinants predominate. *AIDS Res Hum Retroviruses* 2002;18:1281–1290. [PubMed: 12487816]
 38. Sharp PM, Bailes E, Chaudhuri RR, Rodenburg CM, Santiago MO, Hahn BH. The origins of acquired immune deficiency syndrome viruses: where and when? *Philos Trans R Soc Lond B Biol Sci* 2001;356:867–876. [PubMed: 11405934]
 39. Mulanga-Kabeya C, Nzilambi N, Edidi B, Minlangu M, Tshimpaka T, Kambembo L, et al. Evidence of stable HIV seroprevalences in selected populations in the Democratic Republic of the Congo. *AIDS* 1998;12:905–910. [PubMed: 9631144]
 40. Piot P, Plummer FA, Rey MA, Ngugi EN, Rouzioux C, Ndinya-Achola JO, et al. Retrospective seroepidemiology of AIDS virus infection in Nairobi populations. *J Infect Dis* 1987;155:1108–1112. [PubMed: 3471817]
 41. Bakewell, O.; de Haas, H. African migrations: continuities, discontinuities and recent transformations. In: Chabal, P.; Engel, U.; de Haan, L., editors. *African alternatives*. Leiden: Brill; 2007. p. 95-118.
 42. Spiegel PB, Bennedsen AR, Claass J, Bruns L, Patterson N, Yiweza D, et al. Prevalence of HIV infection in conflict-affected and displaced people in seven sub-Saharan African countries: a systematic review. *Lancet* 2007;369:2187–2195. [PubMed: 17604801]

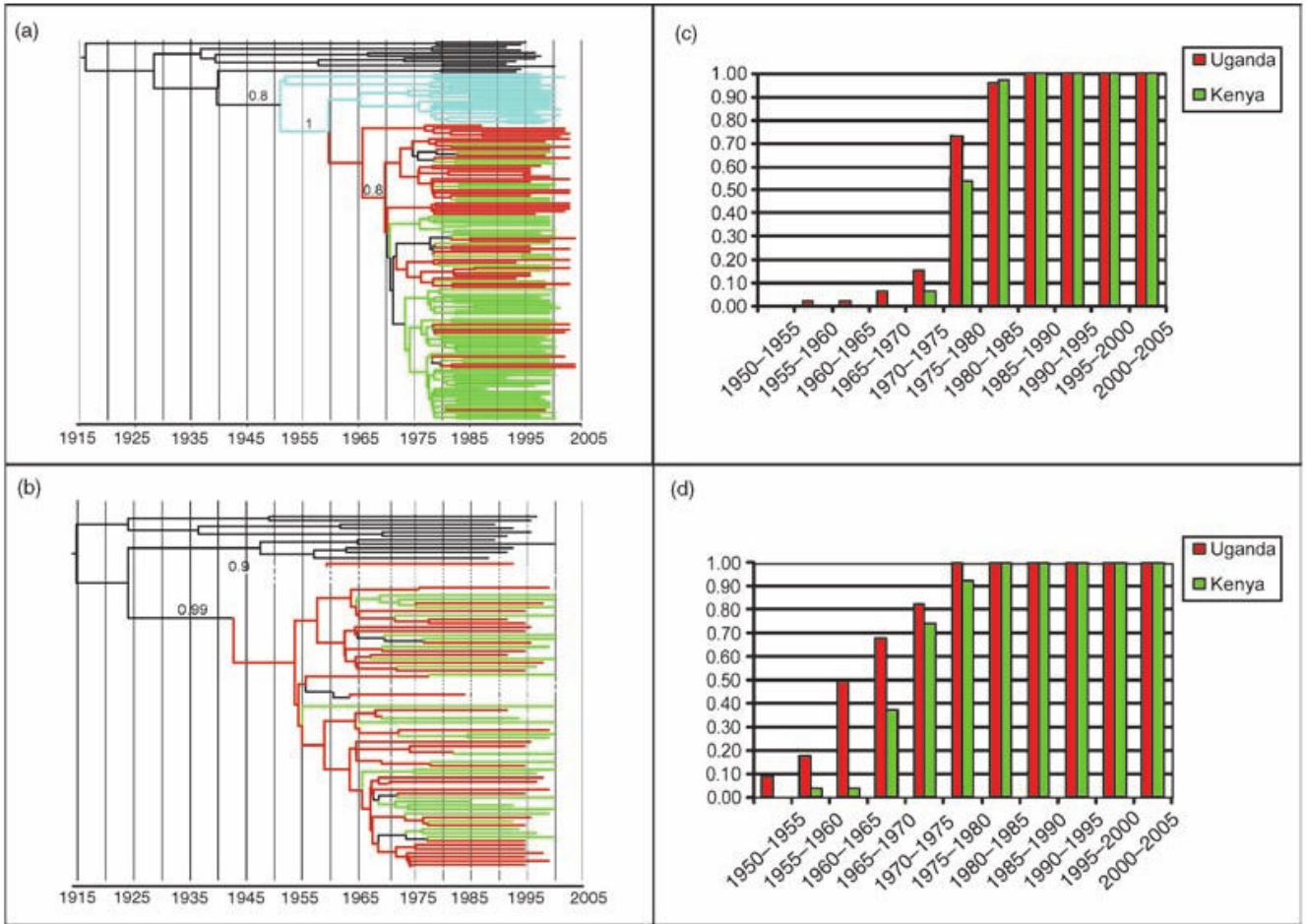


Fig. 1. Genealogy and phylogeography of HIV-1A in Africa

Bayesian maximum clade credibility trees assuming relaxed molecular clock and Bayesian skyline plot coalescent prior for p24 (a) and gp41 (b), generated from the posterior distribution of trees less a 50% burn-in. Branch lengths are shown in years according to the scale bar at the bottom of each panel. Posterior probabilities for major nodes are indicated. Tip branches are colored to represent the country of sampling: red = Uganda, green = Kenya, yellow = Tanzania or Rwanda, blue = west/central Africa, black = outgroups. Internal branches are colored according to the maximum parsimony reconstruction of the ancestral geographic origin. The complete phylogenetic trees are depicted in Fig. S1. (c and d) The bar charts represent the proportion of the total number of lineages (y-axis) present in the HIV-1 genealogies for each 5-year interval (x-axis) beginning with 1955–1960. Genetic lineages that originated in Uganda and Kenya (as inferred from the genealogies in panel a and b) are shown in red and green, respectively. (c) HIV-1A p24 gene. (d) HIV-1A gp41 gene.

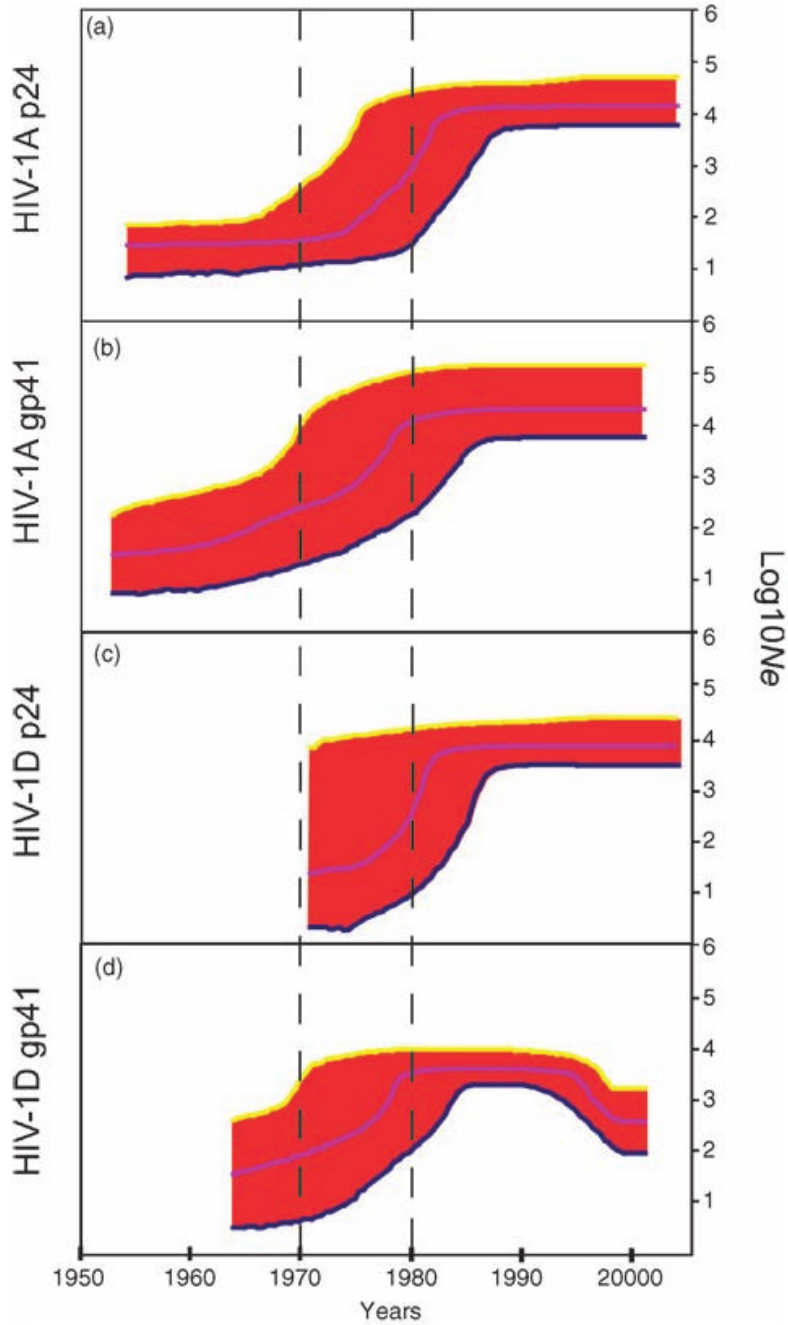


Fig. 2. HIV-1 subtype A and D epidemic history and net migration trends for central and east Africa Bayesian skyline plots of strains included in the A1 clade for HIV-1 A p24 (a) and gp41 (b), and in the east African clade for HIV-1D p24 (c) and gp41 (d). Nonparametric estimates of HIV-1 effective population size (effective number of infections, N_e) over time were inferred for the four datasets. The blue line represents the median estimate of N_e with the shaded area showing the 95% high posterior density (HPD) estimates. Broken vertical lines indicate the period of exponential growth for all four analyses.

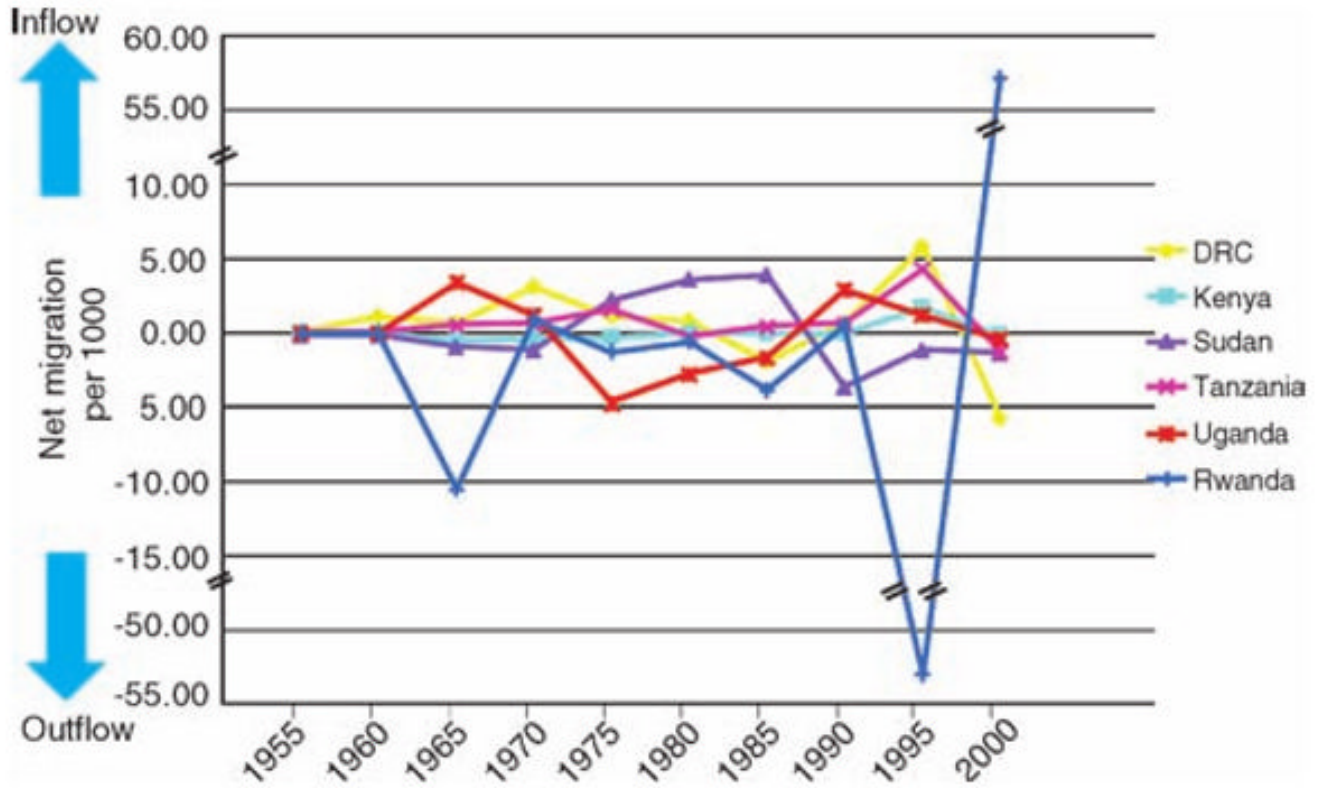


Fig. 3. Net migration rates over 5-year intervals for east Africa

The symbols represent the net migration (number of people in or out-migrating per 1000 individuals on the y-axis) for 5-year intervals (x-axis) from 1955 to 2000. Countries are represented by different colors (DRC, yellow; Kenya, light blue; Sudan, purple; Tanzania, pink; Uganda, red; Rwanda, green). Data were obtained from the United Nations World Population Prospects database <http://esa.un.org/unpp/>.

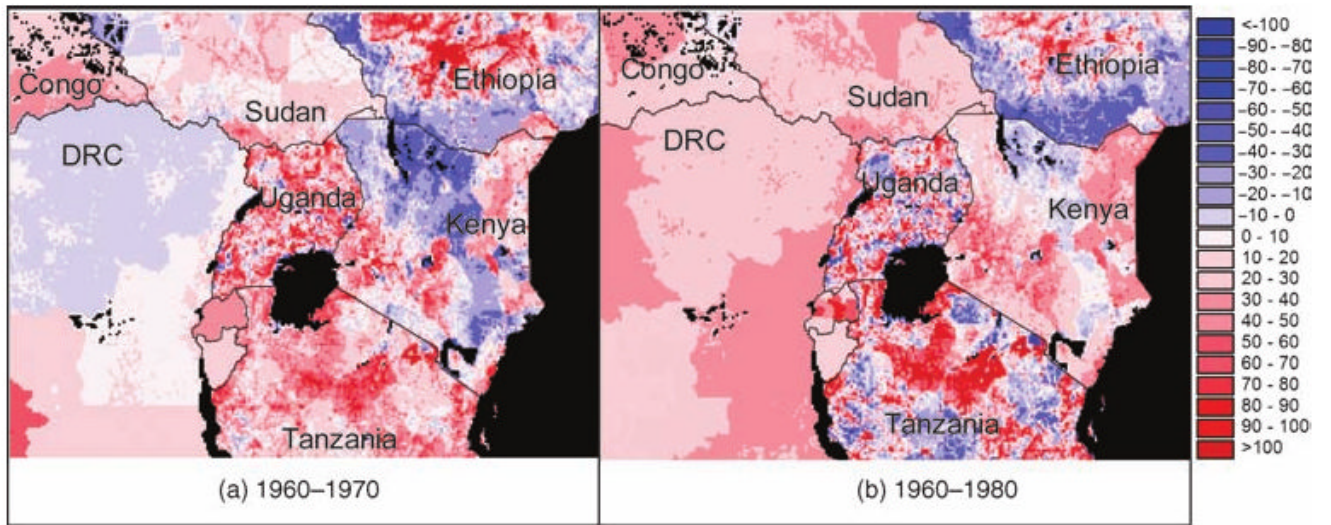


Fig. 4. Population distribution changes in east Africa
 Percentage change in population per grid square is mapped for (a) 1960–1970 and (b) 1970–1980.

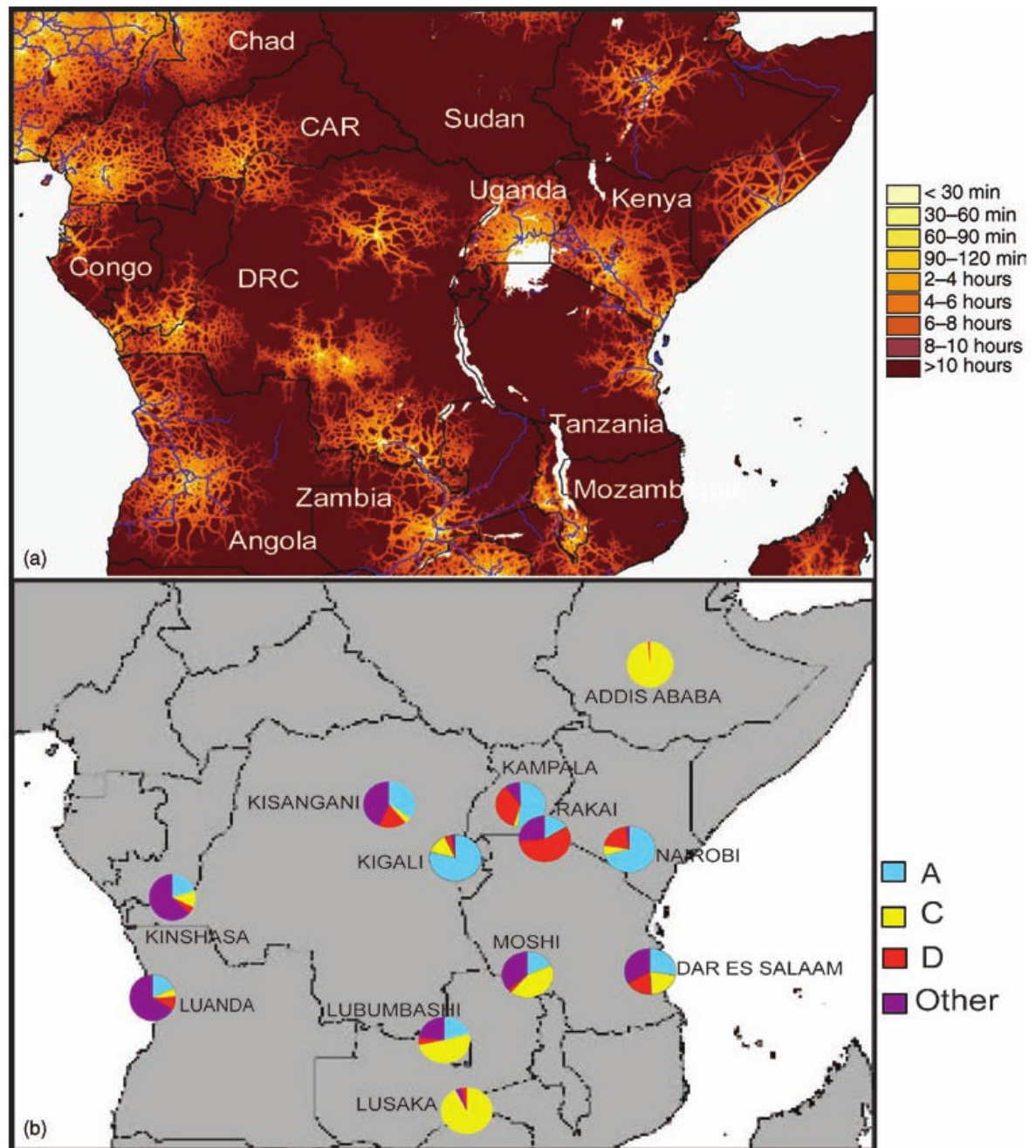


Fig. 5. Accessibility and subtype distribution in central and east Africa

(a) The gradient of colors indicate the estimated travel time to the nearest city of population more than 500 000 people, with yellow at one extreme indicating low travel times (<30 min) and red at the other extreme indicating long travel times (>10 h). Major roadways are indicated in blue. The data are available for download from <http://gem.jrc.ec.europa.eu/gam/download.htm> [26]. (b) HIV-1 subtype distribution in major cities from sources as follows Addis Ababa [27], Dar es Salaam [18], Kampala [28], Kigali [29], Kinshasa [30], Kisangani [31], Luanda [32], Lubumbashi [31], Lusaka [33], Moshi [18], Nairobi [34], and Rakai [35].