LUND UNIVERSITY

The HIV-1 epidemic in Ethiopia – transmission patterns, antiretroviral drug resistance and treatment outcomes

Assefa, Dawit

2022

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Assefa, D. (2022). *The HIV-1 epidemic in Ethiopia – transmission patterns, antiretroviral drug resistance and treatment outcomes*. [Doctoral Thesis (compilation), Department of Translational Medicine]. Lund University, Faculty of Medicine.

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

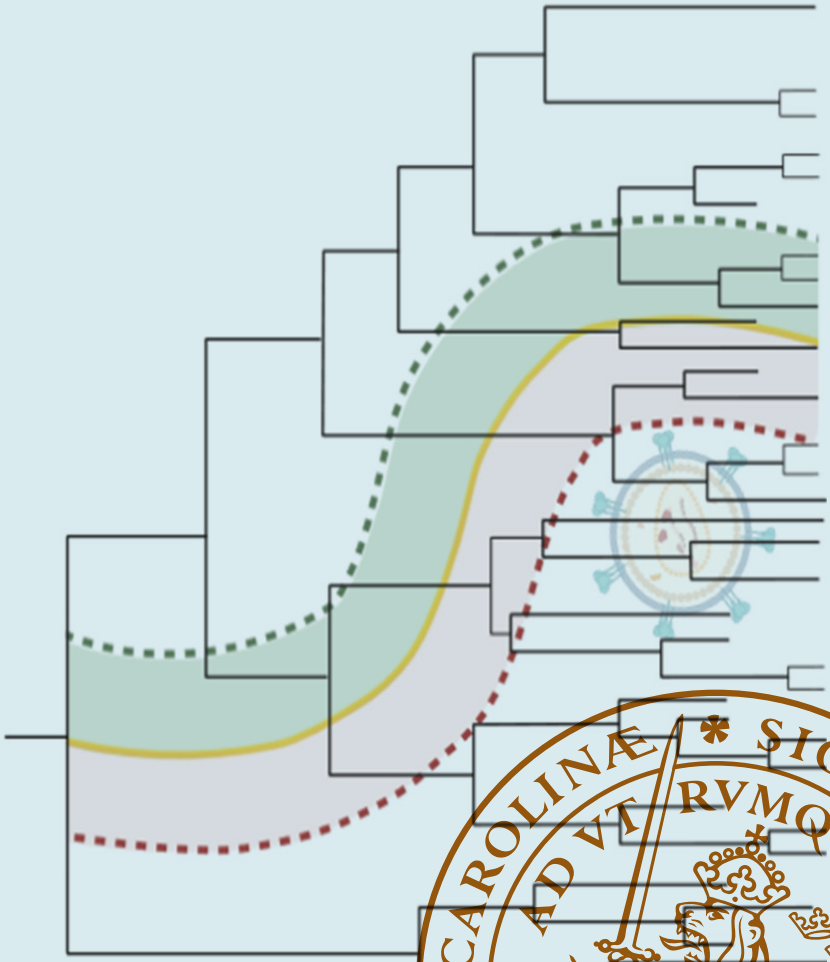
Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

The HIV-1 epidemic in Ethiopia – transmission patterns, antiretroviral drug resistance and treatment outcomes

DAWIT ASSEFA ARIMIDE
DEPARTMENT OF TRANSLATIONAL MEDICINE | LUND UNIVERSITY



The HIV-1 epidemic in Ethiopia – transmission patterns, antiretroviral drug resistance and treatment outcomes

Dawit Assefa Arimide

LUND UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty Medicine, Lund University, Sweden.
To be defended at Lund University, Belfragesalen BMC D15, Lund, Thursday, September 15th, 2022, at 13:00.

Faculty opponent
Associate Professor Ujjwal Neogi, Karolinska Institute, Sweden

Supervisor

Professor Patrik Medstrand, Lund University, Sweden
Co-supervisors

Professor Per Björkman, Lund University, Sweden

Associate Professor Joakim Esbjörnsson, Lund University, Sweden

Organization Lund University	**Document name** Doctoral Dissertation
Faculty of Medicine Department of Translational Medicine	**Date of issue** September 15, 2022
Author: Dawit Assefa Arimide	Lund University

Title: The HIV-1 epidemic in Ethiopia – transmission patterns, antiretroviral drug resistance and treatment outcomes

Abstract

A comprehensive understanding of local HIV epidemiology is essential for monitoring transmission, designing, implementing, and evaluating HIV intervention strategies. In **paper I,** we used a total of 1276 HIV-1 subtype C *pol* sequences and employed state-of-art phylogenetic and phylodynamic tools to describe the molecular epidemiology of HIV in Ethiopia. Our results showed that the HIV epidemic in Ethiopia resulted from two independent introductions of the founder virus from Eastern Africa and southern African countries in 1975 and 1983, respectively. Our phylodynamic analysis also revealed that the HIV-1 epidemic in Ethiopia manifested expanding growth from its introduction until mid-1990s, followed by a sharp decline in HIV-1 transmissions. The epidemic decline coincided with early behavioral, preventive, and public health awareness campaigns implemented in Ethiopia, a decade before the introduction of antiretroviral therapy (ART) in the country.

Global evidence suggests that the rapid expansion of ART is associated with increase in pretreatment drug resistance (PDR) and acquired drug resistance (ADR), posing threat to both individual outcomes and the prospect of elimination of HIV as a public health threat. We employed WHO-recommended threshold survey method in **paper II** to assess the transmitted drug resistance (TDR) in Gondar. Our result showed a moderate level of TDR in Gondar, all of which were associated to non-nucleoside reverse transcriptase inhibitor (NNRTI). Our findings also revealed a high rate of HIVDR transmission with the G190A mutation. In **paper III**, we investigated the emergence of ADR among adults receiving ART in health centers. Our results showed that among 621 individuals, 16.3% had virological failure (VL≥500 copies/mL) at six and/or twelve months, of which 65.3% had ADR. In **paper IV**, we assessed the prevalence of virological failure, ADR and PDR among female sex workers (FSWs) who participated in the 2014, Ethiopian nationwide biobehavioral survey. PDR was detected in 16.5 % (63/381) FSWs of which 14.4%, 10.5% and 9.2% were associated to NNRTI, nucleoside reverse transcriptase inhibitors (NRTIs), and dual-class, respectively. Among the 239 FSWs on-ART, 59 (24.7%) had virological failure, of which 74.4% had one or more major HIV drug resistance mutations (HIVDRMs). In **paper V**, we found no dolutegravir-associated HIVDRMs among 460 INSTI-niave (integrase strand transfer inhibitor), participants in the 2017 Ethiopian national HIVDR surveillance, regardless of previous exposure to ART (NNRTIs, NRTIs and/or protease inhibitors). Furthermore, 64.9% of HIV-1 subtype C integrase amino acid positions were conserved (<1.0% variability).

Key words: HIV-1, HIV drug resistance (HIVDR), Phylodynamics, Phylogenetic, Molecular epidemology, transmitted HIVDR, acquired HIVDR, pre-treatment HIVDR, Integrase strand transfer inhibitors (INSTI), Ethiopia, antiretroviral therapy

Classification system and/or index terms (if any)

Supplementary bibliographical information		**Language:** English
ISSN and key title: 1652-8220 Doctoral Dissertation Series 2022:105		**ISBN:** 978-91-8021-266-3
Recipient's notes	**Number of pages 130**	Price
	Security classification	

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature ዳዊት አሰፋ አርምዴ Date 2022-08-10

The HIV-1 epidemic in Ethiopia – transmission patterns, antiretroviral drug resistance and treatment outcomes

Dawit Assefa Arimide

LUND UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty Medicine, Lund University, Sweden.
To be defended at Lund University, Belfragesalen BMC D15, Lund, Thursday, September 15th, 2022, at 13:00.

Faculty Opponent
Associate Professor Ujjwal Neogi, Karolinska Institute, Sweden

Examination Board

Professor Lennart Svensson, Linköping University, Sweden
Associate Professor Claus Bohn Christiansen, Lund University, Sweden
Associate Professor Jannik Fonager, Statens Serum Institute, Denmark

Cover photo by Zsófia Ilona Szojka (PhD)

Copyright pp 1-130 Dawit Assefa Arimide

Paper 1 © by the Authors (Published and open access)

Paper 2 © by the Authors (Published and open access)

Paper 3 © by the Authors (Published and open access)

Paper 4 Copyright © 2022 Wolters Kluwer Health, Inc

Paper 5 © by the Authors (Published and open access)

This work was supported by: The Faculty of Medicine, Lund University, Sweden; Swedish research council (Vetenskapsrådet); a donation to Lund University and Ethiopian Public Health Institute (EPHI).

Faculty of Medicine, Lund University, Sweden
Department of Translational Medicine

ISBN 978-91-8021-266-3
ISSN 1652-8220
Lund University, Faculty of Medicine Doctoral Dissertation Series 2022:105

Printed in Sweden by Media-Tryck, Lund University, Lund 2022

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

Dedicated to my family.

"Inspiration exists, but it has to find you working"

Pablo Picasso

(1881-1973)

Table of Contents

Summary 8
List of papers 10
List of figures 11
List of tables 12
Abbreviations 13

Introduction 15

History of HIV/AIDS 15
The origin of HIV 16
HIV/AIDS in Ethiopia 21
HIV genome and structure 23
HIV virion 23
HIV-1 genome 24
HIV genetic diversity 27
HIV-1 recombination 27
HIV genetic variants and global distribution 29
HIV replication cycle 31
HIV transmission 34
Pathogenesis of HIV-1 infection 35

Antiretroviral therapy 38

Antiretroviral therapy in Ethiopia 43
HIV-1 drug resistance 45
Mechanisms of HIV-1 drug resistance 45
Impact of HIV-1 drug resistance 46
Pretreatment drug resistance 47
HIV-1 drug resistance testing 48
Genotypic assay 48
Phenotypic assay 49

Phylogenetic analysis 50

Phylogenetic tree 50

Phylogenetic methods 51
Distance-based methods 52
Character-based methods.....52
Critical steps in phylogenetic analysis 56
Phylodynamics 61
The coalescent model 62
Birth-death model.....64
Application of HIV-1 molecular epidemiology to public health.....65
Materials and methods.....71
Study populations and sequences dataset 71
Methodology in details 74
Statistical methods.....81
Ethical approval.....82
Data availability.....82
Main findings and discussions.....83
Paper I.....83
Paper II 85
Paper III.....87
Paper IV.....88
Paper V 90
Conclusions, limitations and future perspectives 93
Acknowledgements.....97
Reference.....100

Summary

A comprehensive understanding of local HIV-1 epidemiology is essential for monitoring transmission, designing, implementing, and evaluating HIV intervention strategies. Although Ethiopia is one of the majorly affected countries by the HIV epidemic in sub-Saharan Africa, no recent comprehensive study has investigated the molecular epidemiology of HIV in Ethiopia. In **paper I**, we used a total of 1276 Ethiopia HIV-1 subtype C *pol* sequences and employed state-of-art phylogenetic and phylodynamic tools to describe the dynamics of the HIV-1 epidemic in Ethiopia. Our results showed that the HIV-1 epidemic in Ethiopia resulted from two independent introductions of founder virus from Eastern Africa and southern African countries in the mid-1970s and mid-1980s, respectively. Our phylodynamic analysis also revealed that the HIV-1 epidemic in Ethiopia manifested expanding growth from its introduction until the mid-1990s, followed by a sharp decline in HIV-1 transmissions. The epidemic decline coincided with early behavioral, preventive, and public health awareness campaigns implemented in Ethiopia a decade before the introduction of antiretroviral therapy (ART) in the country.

Over the last decades, the rapid expansion of ART has significantly reduced the risk of transmission and improved the survival and quality of life of HIV-infected patients. However, global evidence indicates that the rapid expansion of ART is associated with increase in pretreatment drug resistance (PDR) and acquired drug resistance (ADR), posing threat to both individual outcomes and the prospect of elimination of HIV as a public health threat. Following the increase of PDR to non-nucleoside reverse transcriptase inhibitors (NNRTIs), many countries, including Ethiopia, have switched to the dolutegravir (DTG)-based regimen as first- and second-line therapies. However, differences in naturally occurring polymorphisms (NOPs) have been linked to the development of different mutational pathways, resulting in varying levels of drug resistance against integrase strand transfer inhibitor (INSTIs) among different HIV-1 subtypes.

In Ethiopia, there is limited information on HIVDR prevalence (both PDR and ADR) among the general population and risky groups. However, a few studies showed an increase in HIV drug resistance (HIVDR) prevalence with the scale-up of ART in the country. In **paper II**, we employed the WHO-recommended threshold survey method to assess the transmitted drug resistance (TDR) in Gondar. Our results showed a moderate level of TDR in Gondar, all of which were associated to NNRTI. Our findings also revealed a high rate of HIVDR transmission with the G190A mutation in Gondar. In **paper III**, we investigated the emergence of ADR among adult patients receiving ART in health centers. Our result showed that among 621 individuals included in the study, 83.7% (101/621) had a virological failure (VL≥500 copies/mL) at six and/or twelve months, of which 65.3% had ADR. In **paper IV**, we assessed the prevalence of virological failure, ADR and PDR among

female sex workers (FSWs) who participated in the 2014, Ethiopian biobehavioral survey. PDR was detected in 16.5 % (63/381) of the 381 specimens from ART-naive FSWs. NNRTI-associated PDR was detected in 14.4%, while nucleoside reverse transcriptase inhibitor (NRTI) and dual-class were detected in 10.5% and 9.2%, respectively. Among the 239 FSWs on-ART 59 (24.7%) had a virological failure. Of these, 39 specimens were successfully genotyped, and 29 (74.4%) had one or more major HIV drug resistance mutations (HIVDRMs). In **paper V**, we showed that no DTG-associated HIVDRMs were detected among 460 INSTI-naïve, participants in the 2017 Ethiopian national HIVDR surveillance, regardless of previous exposure to ART (NNRTIs, NRTIs and/or protease inhibitors). Furthermore, of the 288 subtype C integrase amino acid positions, 187/288 (64.9%) were conserved (<1.0% variability). Analysis of the genetic barrier showed that subtype B and C had similar genetic barriers to DTG resistance at selected amino acid positions, except that subtype C had a higher genetic barrier to G140C and G140S mutations than subtype B, indicating that the Q148H/K/R DTG resistance pathway is less selected in subtype C. Furthermore, dolutegravir docking analysis revealed that NRTI, NNRTI and protease inhibitor (PI)-associated drug resistance mutations did not affect the native structure of the HIV-1 integrase, supporting the implementation of the wide scale-up of DTG-based regimes in Ethiopia.

In general, our molecular epidemiology findings provide critical information on the dynamics of the HIV-1 epidemic in Ethiopia and the importance of behavioral interventions along with antiretroviral therapy expansion in preventing and controlling HIV transmission. The HIVDR data from the various study groups will be critical for improving Ethiopia's national ART programme and those of other countries in a similar situation.

List of papers

This thesis is based on the five papers listed below.

I. **Arimide, D.A**., Esquivel-Gómez, L. R., Kebede, Y., Sasinovich, S., Balcha, T., Björkman, P., Kühnert, D., & Medstrand, P **Molecular epidemiology and transmission dynamics of the HIV-1 epidemic in Ethiopia: epidemic decline coincided with behavioral interventions before ART scale-up.** *Frontiers in Microbiology*. 2022;13**.**

II. **Arimide DA**, Abebe A, Kebede Y, Adugna F, Tilahun T, Kassa D, Assefa Y, Balcha TT, Björkman P, Medstrand P. **HIV-genetic diversity and drug resistance transmission clusters in Gondar, Northern Ethiopia, 2003-2013**. *PLoS One*. 2018 Oct 10;13(10):e0205446.

III. Reepalu A, **Arimide DA**, Balcha TT, Yeba H, Zewdu A, Medstrand P, Björkman P. **Drug Resistance in HIV-Positive Adults During the Initial Year of Antiretroviral Treatment at Ethiopian Health Centers**. *Open Forum Infect Dis*. 2021 Mar 6;8(4):ofab106.

IV. **Arimide DA**, Amogne MD, Kebede Y, Balcha TT, Adugna F, Ramos A, DeVos J, Zeh C, Agardh A, Chih-Wei Chang J, Björkman P, Medstrand P. **High level of HIV drug resistance and virological non-suppression among female sex workers in Ethiopia: a nation-wide cross-sectional study**. *J Acquir Immune Defic Syndr*. 2021 Dec 24.

V. **Arimide, D.A**.; Szojka, Z.I.; Zealiyas, K.; Gebreegziabxier, A.; Adugna, F.; Sasinovich, S.; Björkman, P.; Medstrand, P. **Pre-Treatment Integrase Inhibitor Resistance and Natural Polymorphisms among HIV-1 Subtype C Infected Patients in Ethiopia.** *Viruses* 2022, *14*, 729.

List of figures

Figure 1. Schematic outline of the origin of HIV-1 and HIV-2. 18

Figure 2. Estimated number of adults and children living with HIV-1 in 2020..... 20

Figure 3. HIV-1 prevalence in Ethiopia by region in 2019..... 23

Figure 4. Schematic representation of the HIV-1 virion. 24

Figure 5. Schematic representation of HIV-1 genome organization. 25

Figure 6. Mechanism of HIV-1 recombination. 28

Figure 7. World map illustrating the prevalence of HIV-1 group M subtypes within each region..... 31

Figure 8. HIV-1 replication cycle and site of action of different antiretroviral inhibitors 33

Figure 9. The natural course of HIV-1 infection in the absence of treatment. 37

Figure 10. Diagrammatic representation of a phylogenetic tree. 51

Figure 11. Phylogenetic tree landscape. 55

Figure 12. Sequence alignment. 57

Figure 13. A comparison of nucleotide substitution models..... 59

Figure 14. Illustration of the coalescent model used to estimate viral population size over time. 63

Figure 15. Schematic tree representing the birth-death process..... 64

Figure 16. A map of Ethiopian cities and towns that were included in a 2014 study of HIV drug resistance among female sex workers. 73

Figure 17. Maximum likelihood phylogenetic tree of the reference data set, derived from Garli v2.0. 77

List of tables

Table 1. Functions of HIV-1 proteins. 26

Table 2. Primers for the amplification and sequencing of the HIV-1 *pol* region. 74

Abbreviations

ADR	Acquired drug resistance
AIDS	Acquired immunodeficiency syndrome
aLRT-SH	Approximate likelihood ratio test, Shimodaira Hasegawa-like
ART	Antiretroviral therapy
ARV	Antiretroviral
AZT	Zidovudine
BEAST	Bayesian evolutionary analysis by sampling trees
BF	Bayesian factor
cART	Combined antiretroviral therapy
CCR5	C-C chemokine receptor type 5
CD4	Cluster of differentiation 4
CD8	Cluster of differentiation 8
cDNA	Complementary deoxyribonucleic acid
CRF	Circulating recombinant form
CTMC	Continuous-time Markov chain
DNA	Deoxyribonucleic acid
DTG	Dolutegravir
DRM	Drug resistance mutation
Env	Envelope gene
ESS	Effective sample size
FSW	Female sex workers
gag	Group-specific antigen gene
gp	Glycoprotein
GTR	General time-reversible substitution model
HAART	Highly active antiretroviral therapy
HIV	Human immunodeficiency virus
HIV-1	Human immunodeficiency virus type 1
HIV-2	Human immunodeficiency virus type 2
HIVDR	HIV drug resistance
HIVDRM	HIV drug resistance mutation
INSTI	Integrase strand transfer inhibitor
LMIC	Low- and middle-income country
MSM	Men who have sex with men
NGS	Next generation sequencing
NRTI	Nucleoside reverse transcriptase inhibitor

NNRTI	Non-nucleoside reverse transcriptase inhibitor
NVP	Nevirapine
PDR	Pretreatment drug resistance
PEP	Post-exposure prophylaxis
PI	Protease inhibitor
PLHIV	People living with HIV
PR	Protease
PrEP	Pre-exposure prophylaxis
RT	Reverse transcriptase
MCC	Maximum clade credibility
MHC	Major histocompatibility complex
RNA	Ribonucleic acid
SIV	Simian immunodeficiency virus
STI	Sexually transmitted infection
tar	Transactivation-responsive RNA
tat	Trans-activator of transcription gene
SDRM	Surveillance drug resistance mutation
ML	Maximum-likelihood
ORF	Open reading frame
TAM	Thymidine analogue mutation
TDF	Tenofovir
TDR	Transmitted drug resistance
tMRCA	Time to the most recent common ancestor
UNAIDS	Joint United Nations Programme on HIV and AIDS
URF	Unique recombinant form
Vif	Virus infectivity factor gene
VL	Viral load
Vpr	Virus protein R gene
Vpu	Virus protein U gene
WHO	World Health Organization

Introduction

History of HIV/AIDS

The first clinical evidence of a new immunodeficiency disease was reported in 1981 with a cluster of unusual *Pneumocystis jiroveci* pneumonia (PCP) and Kaposi's sarcoma among previously healthy young homosexual men in Los Angeles, followed by New York and San Francisco[1-3]. Shortly after these reports, several similar cases were reported among male homosexuals from Southern California and other parts of the world. The syndrome was initially referred to as Gay-Related Immune Deficiency (GRID). Similar cases were later reported in intravenous drug users, haemophiliacs[4, 5], Haitian immigrants [6], recipients of blood transfusions[7, 8], prostitutes, the female partners of men who had the diseases[9] and infants born to mothers with AIDS[10, 11], indicating a blood-borne as well as a sexually transmitted pathogen[12]. In September 1982, the CDC used "Acquired Immune Deficiency Syndrome (AIDS)" to describe the immune disorder and accompanying illness[13]. Over the following years, cases started to appear in Europe, and among immigrants from Sub-Saharan Africa either visited or resided in Europe.

In 1983, the first major scientific breakthrough clue to the aetiology of AIDS came with the isolation of a retrovirus from the lymph node of an individual with generalized lymphadenopathies of unknown origin at the Pasteur Institute in Paris by researchers Françoise Barré-Sinoussi and Luc Montagnier, which they named lymphadenopathy-associated virus (LAV)[14]. A year later, a researcher from the National Cancer Institute, Robert Gallo and colleagues, reported the isolation of a retrovirus from an AIDS patient, which they named the third of the human T-lymphotropic viruses (HTLV-III)[10, 15]. Later, an independent team, Ratner and co-workers, confirmed the viruses the French and Americans isolated as variants of the same retrovirus, the etiologic agent of AIDS and also published the first fully sequenced genome of the virus[16]. In 1986, the International Committee on Taxonomy of Viruses officially named the retrovirus "human immunodeficiency virus, HIV"[17]. In the same year, a new virus related to HIV-1 but immunologically distinct, HIV-2, was isolated and characterized in patients living in France but native to West Africa[18, 19].

The origin of HIV

Since its discovery in 1983, the origins of HIV-1 have been thoroughly investigated. Current scientific evidence suggests that HIV originated from multiple cross-species transmission of the simian immunodeficiency virus (SIV) from nonhuman primates[20]. The first credible evidence for cross-species transmission emerged in 1983, as a result of the isolation of SIV from rhesus macaques (*Macca Mulata*) suffering from AIDS-like symptoms at the new England Regional Primate Centre, which later demonstrated that macaques are not the natural host of SIV but were infected by the cross-species transmission of SIV from Sooty mangabeys (SIVsmm) in captivity[21-23]. SIV infection in macaques often results in rapid progression to immunodeficiency. This was further supported by the detection of antibodies to SIV among residents in Senegal[19, 24]. Evidence of a simian origin is now clear, as similar lentiviruses have been found in more than 40 species of African primates, and a geographical correlation exists between SIVs hosted in different primate species and HIV[25-28].

Molecular phylogenetic studies have revealed that HIV-1 and HIV-2 are the result of at least 13 SIV-to-human cross-species transmission events from three primate species: chimps, gorillas, and sooty mangabeys[29]. Four independent transmissions of SIV from chimps and gorillas to humans gave rise to HIV-1 group M (main), O (outlier), N (non-M, non-O), and P[27]. While the nine independent transmissions of SIV from Sooty mangabeys resulted in nine HIV-2 groups (Group A-I)[30-32] (**Fig. 1**). HIV-1 group M and N originate from two independent transmissions of SIV found in geographically distinct wild chimpanzees, *Pan troglodytes troglodytes* (Ptt) (SIVcpzPtt), in the southeast and south-central Cameron, whereas group P and O were originated from SIV found Western lowland gorillas (*Gorilla gorilla gorilla*) (SIVgor) in Cameroon[19, 33-36]. Exposure of the new host to virus-contaminated body fluids or tissues of the original host is one requirement for cross-species transmission. Chimpanzees hunt various species of monkey cooperatively and have most likely been exposed to SIVs multiple times throughout their evolution[32, 37]. SIVcpzPtt resulted from recombination events involving three different SIV strains, SIVgsn from greater spot-nosed monkeys (*Cercopithecus nictitans*) and SIVrcm from red-capped mangabeys (*Cercocebus torquatus*) as well as an unknown SIV strain[32, 38] (**Fig. 1**). SIVcpzPtt has also been proposed as the source of SIVgor, even though they do not eat meat and avoid chimp interactions. However, their habitat ranges overlap, and aggressive interactions could have resulted in SIV transmission[39].

Chimpanzee-human transmission is thought to have occurred in south-eastern Cameroon as a result of exposure to infectious blood and body fluids while hunting and butchering these animals for food[29]. Infection of primate handlers with simian retrovirus and high seroreactivity to SIV antigen in central African villages where

bushmeat is hunted or consumed provide strong evidence for ongoing cross-species transmission of SIV to humans[30, 40, 41]. There have been opportunities for chimp-to-human host jumps for hundreds or thousands of years, so it is reasonable to assume that many such transmissions have occurred in the past. However, such viruses did not reach detectable levels in the human population until the twentieth century. Only HIV-1 group M had the ability or opportunity to spread in the human population on a pandemic scale.

Recent evidence suggests that individuals infected with the group M HIV-1 virus people most likely travelled to Leopoldville, renamed Kinshasa, via the Sangha River waterways, the primary communication route at the time, where the initial major HIV-1 transmission was greatly accelerated[42, 43]. Human-to-human transmission of HIV-1group M in Kinshasa was greatly accelerated by the destabilization of social structures by invading colonial powers[44, 45], the emergence and rapid growth of major conurbations[43], growing sex trade, high-risk behavior, the concomitant high frequency of genital ulcer diseases (such as syphilis)[46, 47] as well as the widespread use of unsterile injections[48-50] may have provided an unprecedented opportunity for the spread of the viruses in west-central Africa during the early 20th century. The expansion of train routes connecting Kinshasa to other populated cities, such as Mbuji-Mayi and Lubumbashi, as well as population migration, may have also provided the virus with an unprecedented opportunity to spread throughout Africa. In support of this, the oldest known human samples containing HIV-1 were isolated from frozen serum samples and preserved lymph node tissues of Kinshasa residents in 1959 (ZR59) and 1960 (DRC60), respectively[43, 51]. Furthermore, Central Africa has the highest genetic diversity in terms of the number of co-circulating subtypes and intra-subtype diversity, implying that this region was the epicentre of HIV-1 M[52].

The timing of the zoonotic events that resulted in the spread of HIV in human populations has been a source of debate. Phylogenetic and molecular clock analysis has estimated the date of the most recent common ancestor (tMRCA) of HIV-1 group M, O and N, 1920 (1909–1930)[42], 1920 (1890–1940)[53] and 1963 (1948–1977)[54], respectively. In contrast, the cross-species transmission of HIV-2 groups A and B to humans was estimated to be around 1932 (1906–55) and 1935 (1907–61), respectively[53-55].

HIV-2

HIV-2 was first isolated in West Africa in mid-1980 among individuals living with AIDS[24, 56]. An estimated 1–2 million people worldwide are infected with HIV-2; however, its prevalence is decreasing due to the lower risk of horizontal and vertical transmission associated with lower plasma viral load, which is often undetectable[57, 58]. According to data from West African cohorts, up to 37% of untreated HIV-2 patients have an undetectable viral load[57, 59]. HIV-2 remains restricted mainly in

West Africa. The prevalence of HIV-2 in Guinea-Bissau, Senegal, Gambia, Sierra Leone, and the Ivory Coast ranges from 1 to 5%, while in all other West African countries, including Cape Verde, it is less than 1%[57, 60, 61]. HIV-2 infection has been reported in many other countries, including Spain, Portugal, France, India, Brazil, Germany, the United Kingdom and the United States[57, 61-64]. Among the nine distinct HIV-2 groups (A-I), only groups A and B are endemic; in contrast, all the other groups have been identified in only one or two individuals and are considered 'dead-end' indicating the continuous transmission of SIVsmm to humans. The high frequency of zoonotic transmissions might be due to the high prevalence of SIVsmm in Sooty mangabeys, which are kept as household pets or hunted for bush meat. In contrast to HIV-1, only two recombinant forms have been described, labelled as CRF01_AB and the unique recombinant form obtained by sequencing the entire genome of the virus obtained from an infected Japanese individual and two Nigerian patients most likely infected in their country of origin[65].

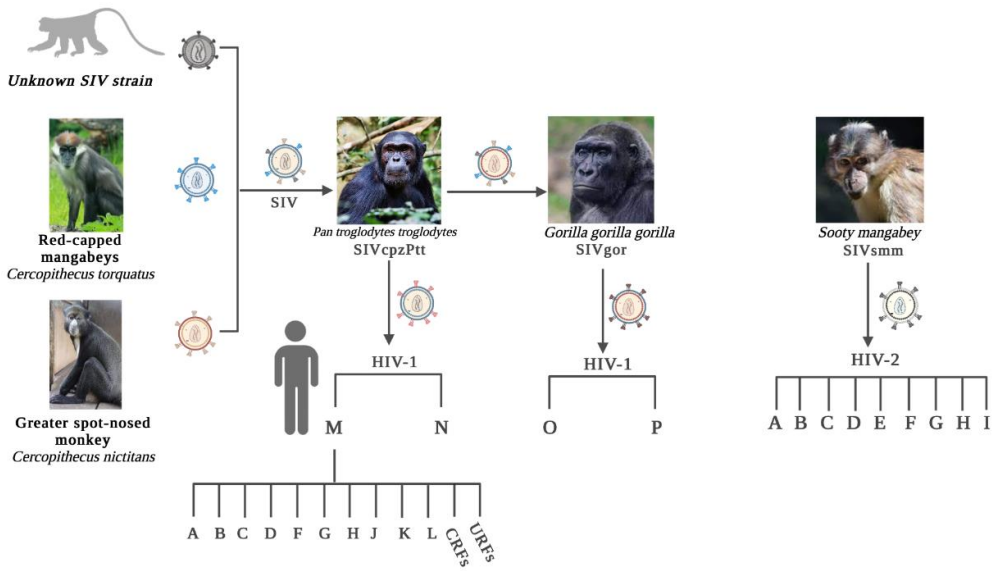


Figure 1. Schematic outline of the origin of HIV-1 and HIV-2.
African primates are naturally infected with more than 40 different lentiviruses, known as simian immunodeficiency viruses (SIVs), with a suffix denoting the primate species from which they are derived (SIVsmm from sooty mangabeys). The cross-species transmission event from monkeys resulted in a recombinant virion in Pan tronglodytes tronglodytes (SIVcpzPtt). The SIVcpzPtt, virus was then passed on to gorillas and humans, giving rise to SIVgor and HIV-1 groups M and N, respectively. HIV-1 groups O and P resulted from two zoonotic transmission events of SIVgor, whereas SIVsmm infecting sooty mangabeys was transmitted to humans at least nine times, resulting in the emergence of HIV-2 groups A through I. (Authors own artwork).

The global burden of HIV

HIV continues to be a major global health issue and is among the leading causes of death, in many low- and middle-income countries (LMICs). Since the beginning of the epidemic, an estimated 79.3 million people have been infected with HIV, and 36.3 million people have died from AIDS-related illnesses. At the end of 2020, an estimated 37.7 million HIV-1 infected people were living worldwide, with 6.1 million unaware that they were infected[66, 67].

The global expansion of antiretroviral therapy (ART) access, driven by the ambition to meet the 90-90-90 targets set for 2020, has resulted in remarkable progress in lowering the number of new HIV/AIDS infections and deaths. Indeed, in 2020, 84% of people living with HIV knew their status, 87% of those who knew their HIV status were receiving ART, and 90% of those receiving ART were virologically suppressed[67]. Millions of lives have been saved due to the global rollout of ART: an estimated 16.6 million AIDS-related death have been averted, with 47% decrease in AIDS-related mortality since 2010. This decline in HIV-1 incidence is more pronounced among children, consistent with increased ART coverage among pregnant women[67]. In 2020, 85% of HIV-positive pregnant women had access to antiretroviral medicines to prevent HIV-1 transmission to their children. However, in the same year, about 150,000 children were newly infected with HIV-1, which is a 53% decline compared to 2010. However, HIV/AIDS remains a global health crisis, with approximately 1.5 million new HIV-1 infections and 680,000 deaths from AIDS-related causes in 2020[66].

The prevalence and incidence rates of HIV-1 infection vary widely across the globe, reflecting the growth of local epidemics fuelled by distinct modes of transmission, socioeconomic environments, and behavioral factors. Sub-Saharan Africa has been the most heavily affected region throughout the history of the HIV-1 epidemic. In 2020, there were an estimated 20.6 million people living with HIV in eastern and southern Africa, 5.7 million in Asia and the Pacific, 4.7 million in western and central Africa, and 2.2 million in Western and Central Europe and North America[66] (**Fig. 2**). Within Africa, southern Africa is the most affected region, accounting for approximately 25% of new infections. Northern Africa, on the other hand, has significantly lower prevalence rates due to fewer high-risk cultural patterns that promote HIV transmission or spread[68, 69].

Depending on the local HIV-1 prevalence, the population-specific risks of infection differ. Women account for more than half (55%) of all adults (15-49 years) living with HIV worldwide, and HIV-1 (along with pregnancy-related complications) is the leading cause of death among women of reproductive age. In sub-Saharan Africa, women and girls accounted for 63% of all new HIV-1 infections in 2020[66]. Gender inequalities, disparities in access to services, and sexual violence make

women (particularly younger women) more vulnerable to HIV. Young people aged 15 to 24 account for roughly one-third of new HIV infections, and young women are disproportionately affected in some areas[66]. Despite accounting for only 10% of the population, adolescent girls and young women (aged 15 to 24 years) accounted for 25% of HIV infections in sub-Saharan Africa in 2020[66].

Other key populations (men who have sex with men (MSM), transgender people, intravenous drug users, and sex workers) are at the highest risk of HIV infection in lower-prevalence settings[70]. Transgender women are 34 times more likely to contract HIV than other adults; female sex workers are 26 times more likely to contract HIV than other adult women, and gay men and other men who have sex with men are 25 times more likely to contract HIV than heterosexual adult men[71]. Key populations and their sexual partners were responsible for 65 % of HIV infections worldwide in 2020 and 93 % of infections outside of Sub-Saharan Africa[67].

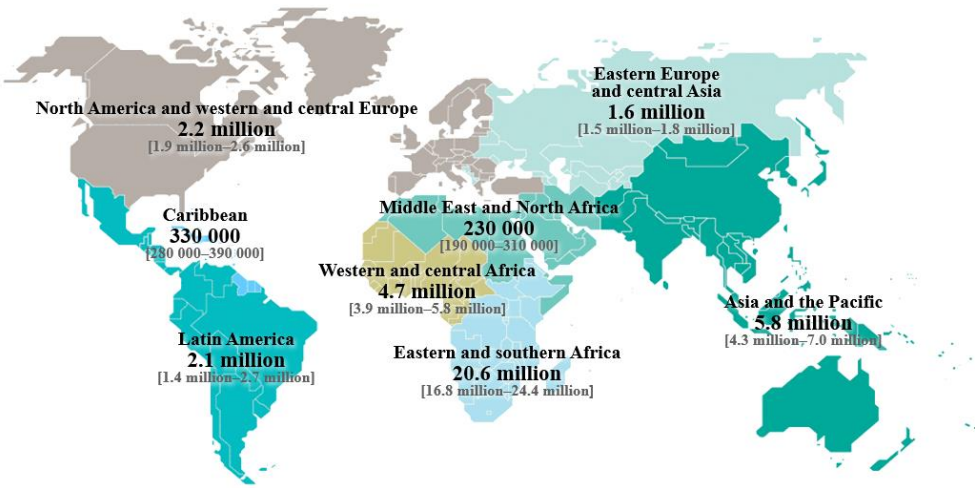


Figure 2. Estimated number of adults and children living with HIV-1 in 2020.

The data used in the map were from UNAIDS global HIV and AIDS factsheet, available at https://www.unaids.org/en/resources/fact-sheet , as of June 20, 2022)[66]

HIV/AIDS in Ethiopia

Ethiopia is one of the many sub-Saharan countries hit hard by the HIV pandemic. The first two HIV-1 positive cases were detected in 1984 among 167 samples collected, while the first hospitalized AIDS patient was reported in 1986 in Addis Ababa[72, 73]. In the following years, 13 of 528 samples tested for HIV from patients at Black Lion Hospital between 1985 and 1987 were positive for HIV-1. Since then, HIV/AIDS has claimed the lives of hundreds of thousands of people and has left behind hundreds of thousands of orphans. HIV transmission was primarily through heterosexual intercourse and, to a lesser extent, mother-to-child transmission.

Initially, the epidemic was concentrated among the high-risk groups (female sex workers, long-distance truck drivers, soldiers), urban areas, and major trade routes. A large-scale survey conducted in 1988 among 6234 female sex workers (FSW) working in 24 selected locations along the main trade route showed an HIV-1 prevalence of 5.3% - 38.1%. In some urban areas, a prevalence of 38% was reported among FSWs[74]. Among the sex workers, condom use was relatively low early in the epidemic, and HIV prevalence among sex workers rapidly reached high levels. In 1988 a study among FSWs in Addis Ababa showed a 24.7% HIV prevalence, while in clients at sexually transmitted infections (STI) clinics, 54.3% and 73.4% were reported in 1990 and 1998, respectively. Truck drivers and soldiers, who are highly mobile and have multi-sexual contact, were another high-risk group in Ethiopia. While among truck drivers, 17.3% HIV prevalence was reported in 1989[74]. HIV prevalence of 12% and 27% were reported among soldiers in 1990 and 1993, respectively[75-77]. Soldiers may have been infected through blood transfusions without screening[75]. Data on the HIV prevalence among the general population during the early period of infection (1980-1990) were limited due to a lack of diagnostic facilities and underreporting. However, based on data from the prenatal clinic from 28 urban and six rural test sites, the HIV prevalence in the sexually active population (15-49) was estimated to be 13.2% in urban, 2.3% in rural and 15.6% in Addis Ababa[75]. Similarly, among pregnant women attending antenatal care clinics (ANC) in Addis Ababa, HIV prevalence of 4% and 10.5% were reported in 1989 and 1990, respectively[77].

Several factors, including a lack of awareness, political instability and civil war, high mobility among FSWs, a high STI rate in high-risk groups and the general population, and high-risk sexual activities, all contributed to the epidemic's rapid spread during the early period. Other factors that may have contributed to the spread of HIV-1 in the country include a lack of prevention intervention, increased population movement due to urbanization, and demobilized soldiers[75, 78-81]. Non-sterile injections in medical settings, and the shared use of sharp instruments in traditional ceremonies, were common in Ethiopia during this time and may have contributed to the spread of HIV[73].

Since the first case report in 1984, an extensive campaign against HIV/AIDS resulted in increased awareness and significant reductions in sexual risk behaviour. Ethiopia established the first national HIV task force in 1985 and the first national AIDS prevention and control program in 1987. Ethiopia was the first African country to establish a task force to prevent and control HIV/AIDS and STI infections and a national plan for HIV epidemic response intervention[75, 77, 82]. Several biobehavioral interventions and awareness programs were implemented in the early 1990s through national media, schools, and public gatherings[83, 84]. These programs primarily focused on long-term health education, risk reduction, condom promotion, and the prevention and control of sexually transmitted infections (STI)[82, 84].

Although there is no adequate and reliable national survey data on behavioral risk factors in Ethiopia, the different program reviews conducted from 1989 to 1991, as well as two large-scale nationwide condom use surveys conducted from 1987 to 1993, revealed that the interventions had encouraging evidence of substantial change in sexual risk behavior, increased knowledge level about HIV/AIDS, increased condom use, and a substantial reduction in a non-regular partner and sexually transmitted disease[85].

Similarly, increased condom uses and lower numbers of non-regular partners were reported among high school students in Addis Ababa and Gondar during these times[77]. This has been reflected by the decline in the HIV infection rate since mid-1990 in Ethiopia. The HIV prevalence trend among young women (15-24 years) attending ANC, who are a proxy indicator of epidemic decline as they measure the frequency of relatively recent infection and less influenced by death in Addis Ababa between 1995-2003, has shown a significant decrease from 24.2% in 1995 to 12.9% in 2003[86, 87]. A similar decline in HIV prevalence among young blood donors in Addis Ababa and nine other towns was also documented[77].

The HIV epidemic in Ethiopia is characterized as mixed, with wide regional variations, gender variation, and concentrations in urban areas, including some distinct hotspot areas driven by key and priority populations[88]. Ethiopia has identified key and priority population groups based on local epidemiology. According to the Ethiopian National HIV Prevention Road Map (2018-2020), key population are FSWS, while priority populations include prisoners, widowed, distance drivers, separated or divorced women, mobile and resident workers in hotspot areas, PLHIV and their partners, adolescent girls and young women involved in transactional sex. This population group has significantly higher HIV prevalence rates than the general population and has limited access to services due to stigma and discrimination[88]. In 2020, the national adult HIV-1 prevalence was 0.96%, with a 2.9 % prevalence in cities, which is seven times higher than the 0.4% prevalence in rural areas[89, 90]. According to the National HIV Related Estimates and Projections, an estimated 622,326 people were living with HIV-1, including 44,138 children, in 2020[90]. In the same year, there were 11,715 new HIV-1 infections and

11,546 AIDS-related deaths[90]. The Gambela region had the highest HIV prevalence (4.45%), while Ethiopia's Somali region had the lowest (0.16%)[90](**Fig. 3**).

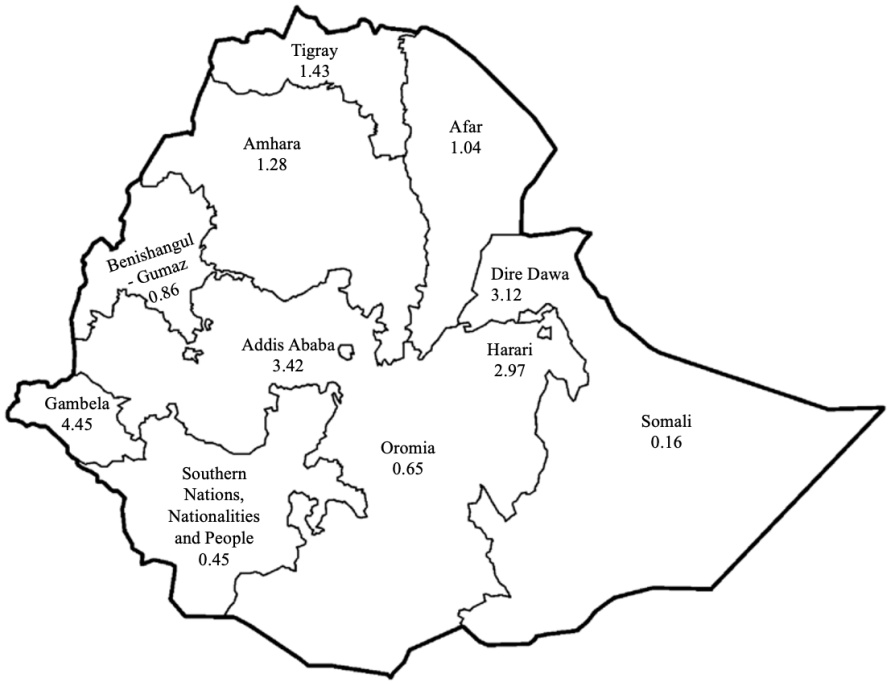


Figure 3. HIV-1 prevalence in Ethiopia by region in 2019.

The data used in the map were based on the HIV-related estimates and projections in Ethiopia for the Year-2019[91].

HIV genome and structure

HIV virion

HIV-1 belongs to the genus *Lentiviridae* within the family *Retroviridae*. Lenti reflects the slow rates of pathogenesis (long incubation period) associated with infection by these viruses. Mature HIV-1 virion is spherical with a diameter of approximately 100-150 nm[92]. It is enveloped by a lipid bilayer acquired as the virus buds from the infected cell. The virion lipid membrane contains 7-35 envelope trimers exposed evenly over the envelope, comprising viral surface glycoproteins (gpl20), transmembrane (gp41), and host cell-derived cellular membrane proteins. Lining the inside of the envelope is the matrix protein (MA, p17) that covers the cone-shaped viral core capsid (p24)[92]. The virus core contains two copies of single-stranded, positive-sense, genomic RNA, and viral enzymes reverse transcriptase

(RT), protease (PR), integrase (IN), as well as accessory regulatory proteins[92, 93]. A schematic representation of the HIV-1 structures is presented in figure 4.

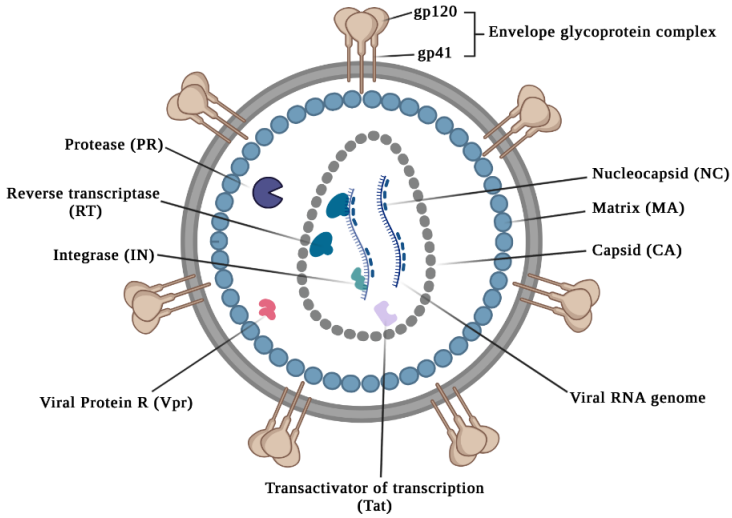


Figure 4. Schematic representation of the HIV-1 virion.
The mature HIV virion contains a viral core, which is surrounded by the lipid bilayer bearing glycoprotein (gp120 and gp41) spikes. The viral core consists of capsid protein (CA, p24), nucleocapsid protein (NC, p7), HIV-1 genome (two copies of positive-sense ssRNA), and enzymes including reverse transcriptase (RT), Integrase (IN) and trans-activator of transcription (Tat). Beneath the lipid membrane is a symmetrical layer of matrix proteins (MA, p17), protecting the capsid.

HIV-1 genome

The HIV-1 genome is approximately 10-kilo base pair (kb) long and contains nine genes in three reading frames flanked by viral long terminal repeats on both sides (5' and 3' LTRs). Both ends have transcriptional regulatory elements, RNA processing signals, packaging sites, and integration sites. The 5' LTR contains the enhancer/promoter sequences for viral transcription, and the 3' LTR contains the poly-adenylation signal[94]. Like all replication-competent retroviruses, the HIV-1 genome encodes three main types of protein, Gag (group-specific antigen), Env (Envelope) and Pol (polymerase). The *gag* gene encodes polyprotein p55, which is cleaved into structural proteins matrix (MA, p17), the capsid protein (CA, p24), the nucleocapsid (NC, p7) and nucleic acid-stabilising protein (p6). The *Env* gene encodes envelope glycoprotein gp160, which is cleaved to form the surface membrane (SU or gp120) and transmembrane glycoprotein (TM or gp41), which form exposed structures at the surface of the host cell. The *pol* (polymerase) gene encodes enzymatic proteins of the virus, protease (PR), reverse transcriptase (RT) and integrase (IN). In addition to the structural proteins, the HIV-1 genome encodes

two regulatory (*tat* and *rev*) genes necessary for the initiation of HIV replication and four accessory genes (*vif*, *vpr*, *nef*, and *vpu*) that have an impact on viral replication, virus budding and pathogenesis[16] (**Fig. 5**).

Tat (trans-activator of transcription) functions as an activating transcriptional protein, and rev (RNA splicing-regulator) facilitates the transport of unspliced mRNAs to the cytoplasm, both of which are necessary for the initiation of HIV replication. Nef (negative regulating factor) lowers the level of CD4 receptors on the cell surface and stimulates infected cells to divide. Vif (viral infectivity factor) promotes the production of infectious virions, vpr (virus protein r) promotes the transport of the pre-integration complex into the host nucleus after reverse transcription and vpu (virus protein U) responsible for the degradation of newly synthesised CD4 receptors and aids in the assembly and release of the virion. Moreover, vif, vpu and nef counteract several cellular restriction factors to secure efficient replication[93]. The functions of HIV proteins are summarized in table 1, and discussed in relation to the HIV-1 life cycle (see HIV-1 replication cycle).

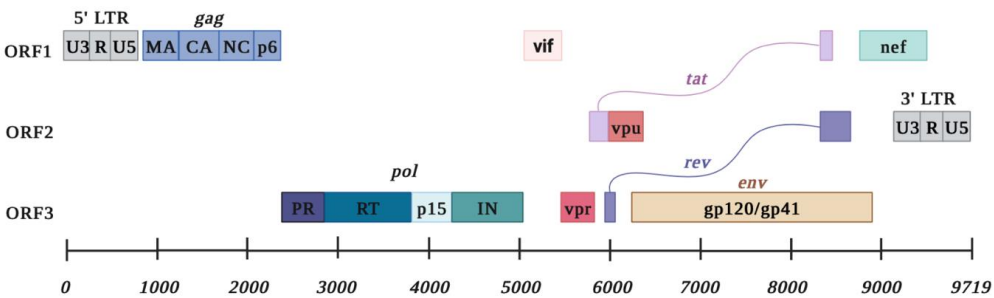


Figure 5. Schematic representation of HIV-1 genome organization.

The three-coding reading frame are depicted along with their open reading frame. The genome is composed of three structural genes (*gag*, *pol* and *env*) four accessory genes (*vif*, *vpr*, *nef*, and *vpu)* in and two regulatory genes (*tat* and *rev)*. Genome position numbering is based on HXB2 reference strain. Abbreviations: ORF, open reading frame; LTR, long terminal repeat, PR, protease; RT, reverse transcriptase; IN, integrase; vpu, virus protein unique; nef, negativity regulatory factor; rev; vif, viral infectivity protein; vpr, virus protein r; tat, trans-activator of transcription; regulator of expression of viral protein. The diagram was adapted with modification for clarity from www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html.

Table 1. Functions of HIV-1 proteins.

Gene	Size *	Protein	Position	Function
gag	P17	Matrix (MA)	790-1185	Myristilated protein, forming the inner membrane layer
	P24	Capsid (CA)	1186-1878	Formation of conical capsid
	P7	Nucleocapsid (NC)	1921-2085	Formation of the nucleoprotein/RNA complex
	P6	Core protein	2134-2289	Involved in virus particle release
pol	P10	Protease (PR)	2253-2549	Proteolytic cleavage of gag-pol precursor protein and viral enzymes
	P51	Reverse transcriptase (RT)	2550-3869	Transcription of HIV RNA into proviral DNA
	P66	RNase H	3870-4229	Degradation of viral RNA in the viral RNA/DNA replication complex
	P31	Integrase (IN)	4230-5093	Integration of proviral DNA into the host genome
env	gp120	Surface glycoprotein (SU)	6225-7758	Attachment of virus to target cell
	Gp41	Transmembrane (TM)	7758-9895	Anchoring of gp120, fusion of viral and cell membrane
tat	p14	Tat	5831-6045, 8379-8469	Binds TAR in presence of host cyclin T1 and CDK9 enhances RNA Pol II elongation on the viral DNA template
rev	Rev	Rev	5970-6045, 8379-8653	Binds RRE inhibits viral RNA splicing and promotes nuclear export of incompletely spliced viral RNAs
nef	P27	Nef	8797-9417	Promotes down-regulation of surface CD4 and MHC 1expression, blocks apoptosis, enhances viral infectivity, alters state of cellular activation
vpu	P16	Vpu	6045-6310	Promotes CD4 degradation and influences virion release, overcomes inhibitory effects of tetherin
vif	P23	Vif	5041-5619	Overcomes inhibitory effects of APOBEC3, preventing hypermutation and viral DNA degradation
vpr	p15	Vpr	5559-5850	Regulate viral and cellular gene expression, facilitates HIV infection of macrophages

Nucleotides numbered according to HXB2 subtype B reference strain (GenBank accession number K03455).
*Numbers represent the protein (p) or glycoprotein (gp) size in kilo Daltons (kDa)

HIV genetic diversity

HIV-1 is characterized by its rapid genetic evolution and high genetic variability, which are the result of a combination of factors[95]. First, HIV-1 has an extremely high mutation rate[96, 97]. Multiple viral and cellular factors influence both the rate and type of mutations produced during viral replication. However, RT is a major source of mutations (5.9×10^{-4} to 5.3×10^{-5} mutations per base per replication cycle)[98] because it is highly error-prone, owing to a lack of proofreading activity[97, 99-101]. Second, HIV-1 has a high replication rate, with a viral doubling time of 0.65 days and a viral generation time of 2.6 days[102]. Each day an estimated 10^{10}-10^{12} virions are produced in untreated individuals, resulting in an innumerable virus variant, often called quasispecies[102, 103]. The frequent recombination and natural selection further elevate its rate of evolutionary change[104]. Although the HIV-1 infection is thought to be initiated by a single virus or infected cell in 80% of heterosexuals, 60% in MSM and 40% in intravenous drug users (IDU)[105], within a single host, HIV-1 can acquire 5-10% diversity levels and be 10% divergent from the original variant(s) within a year after infection[106, 107].

HIV-1 recombination

HIV is a diploid virus with two genomic RNA molecules in each virion. The viral enzyme, RT, can switch between the two RNA templates with high frequency during reverse transcription. If the two RNA copies differ due to infection by two or more different HIV-1 subtypes (co-infection or super-infection), this results in the production of virions that pack an RNA molecule from both subtypes (heterozygous virions). When such virions infect new target cells, they may produce a mosaic genome via gene or gene fragment exchange caused by RT template switching, resulting in recombinant viruses[108, 109]. All progeny virions that are made after this will have this recombinant genotype (that carries regions from two genetically distinct parental strains) (**Fig. 6**).

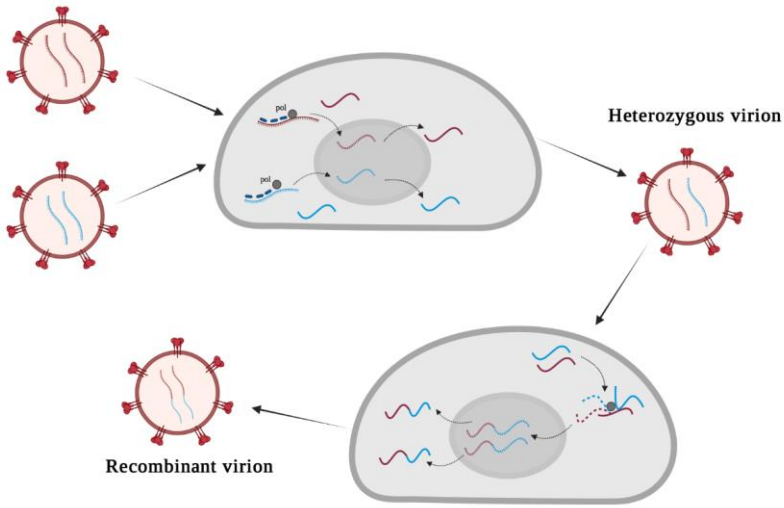


Figure 6. Mechanism of HIV-1 recombination.

HIV-1 is a diploid virus for which, when a cell gets infected by two genetically different HIV-1 viruses, the RT may use both RNA templates for the first strand synthesis resulting in heterozygous virion. When this virion subsequently infects a new cell, template switching occurs during reverse transcription, resulting in a recombinant virion. The diagram was adapted from Lam *et al*., 2010, with modification for clarity[110].

HIV-1 has one of the highest recombination rates of any genetic system (estimated rate of 2.8 crossovers per genome per cycle)[111]. Recombinant forms are classified as circulating recombinant forms (CRFs) if found in three or more epidemiologically unlinked individuals or unique recombinant forms (URFs) if there is no evidence of onward transmission. When recombination involves more than three subtypes, the term "cpx" is used[112, 113]. For example, CRF27_cpx and CRF18_cpx are composed of five or six different subtypes and unclassified segments[105].

The co-circulation of different HIV-1 variants is required for the generation of new recombinants (i.e. URFs), which redefined into CRFs if successfully transmitted in the population as described above. Co-infection and subsequent recombination between subtypes may be common in regions where different subtypes are prevalent, such as central Africa, Southeast Asia, and South America[114-117]. Recombination is a key feature of HIV-1, shaping its evolution, diversity, and adaptation[118-121]. Recombination is also believed to contribute to viral fitness[120, 122-126], drug resistance[109, 118, 127-129], immunological escape[130, 131], and disease progression[109, 118, 124, 132-134]. Early on, recombination between HIV-1 subtypes was identified as a major mechanism for HIV-1 group M diversification. Some recombinants, such as CRF01-AE and CRF02-AG, were present at the start of the HIV epidemic in Central Africa and subsequently spread to other regions where they play important roles in regional epidemics as well as globally[118]. The majority of

known recombinants, on the other hand, have only recently emerged. Recombination is ongoing in many places worldwide where different subtypes and CRFs co-circulate, referred to as "recombination hotspots," giving rise to URF[135-137]. A high prevalence of URFs (and the subtypes/CRFs involved) is found in West Africa, Central Africa, East Africa, South America, and Cuba[105, 117, 138].

HIV genetic variants and global distribution

Based on the phylogenetic analysis, HIV-1 has been classified into four genetically distinct groups M (main), O (outlier), N (non-M, non-O), and P. HIV group M is a major epidemic strain accounting for more than 98% of the global HIV epidemic, whereas groups N, O, and P have limited transmission and account for only 1-2% of all HIV-1 infection worldwide. HIV-1 group O isolates have been recovered from people in Cameroon, Gabon, and Equatorial Guinea; their genomes share approximately 65% identity with group M viruses [139]. Group O strains were responsible for more than 20% of HIV-1 infections in Cameroon early in the epidemic but are now responsible for only about 1%. The genetic diversity within group O is high and can be classified into five clades, I-V, which are genetically distant from each other; however, due to its limited spread beyond its origin, it lacks the subtype-like signal[105]. Group N account for a handful of cases, mainly in Cameroon, France and possibly Togo and only two known cases belonging to group P have been described in Cameroon.

The major HIV-1 group M is the most diversified genetically and further classified into ten distinct subtypes (A, B, C, D, F, G, H, J, K, and L), several sub-sub types (A1-A6, F1, F2.) and CRFs and URFs. As of April 2022, based on the LANL-HIV database

(https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html),118 CRFs and a large number of URFs have been reported worldwide.

The classification was initially based on *env* sequences; however, it applies to all regions of the genome. The genetic variation between subtypes at the amino acid level ranges from 17–35% but can be up to 42% depending on the subtypes and genome regions examined, whereas intra-subtype genetic variation ranges from 8–17% but can be as high as 30%[105, 115].

The geographical distributions of the HIV-1 subtype are heterogeneous and determined by complex factors, including social transmission networks, urbanization, transportation networks, migration, founder effects, and population growth[105, 140]. However, in some cases, HIV-1 subtypes can be linked to a specific epidemiological risk group or geographical region. These patterns of HIV-1 distribution are either due to accidental trafficking or due to a predominant route of

transmission that gives a subtype a strong advantage to be dominant in a particular country or region[141]. Molecular epidemiology studies have revealed that, except for central Africa, which contains the majority of the existing HIV-1 Group M subtypes and recombinants, there is a distinct geographic-demographic distribution pattern[141](**Fig. 7**).

For example, HIV-1 subtype A is the most prevalent form of HIV-1 in areas of Central and Eastern Africa and within member countries of the former Soviet Union. HIV-1 subtype A is primarily transmitted through heterosexual contact in East and Central African countries, whereas intravenous drug use is the most common source of infection in former Soviet Union countries.

Subtype C was the dominant subtype in South Africa, Ethiopia and South Asia (India), contributing to 89% of the infection[115]. Subtype C epidemics have also been reported in Brazil and China, where it is primarily associated with intravenous drug use[141]. Subtype B was the major subtype in Western and Central Europe, the Caribbean and Latin America, and North America, accounting for at least 75% of the infection[115]. CRF01_AE is the dominant strain in Southeast and East Asia, accounting for about 80% of the infection[115]. CRF02_AG is the dominant strain in West Africa, accounting for 46.2% of the infections[115].

According to a recent global survey of the distribution of HIV-1 subtypes and CRFs, HIV-1 subtype C accounts for the 46.6% of all HIV-1 infections worldwide, followed by subtype B (12.1%), subtype A (10.3), CRF02_AG (7.7%), CRF01_AE (5.3%), subtype G (4.6%), subtype D (2.7%). Subtypes F, H, J, and K combined accounted for 0.9% of all global HIV-1 infections. Other CRFs and URFs were each responsible for 3.7% and 6.1% of global infections, respectively, bringing the combined total of worldwide CRFs to 16.7% and all recombinants (CRFs along with URFs) to 22.8%[115]. Thus, there has been a global increase in the proportion of CRFs, a decrease in URFs and an overall increase in recombinants.

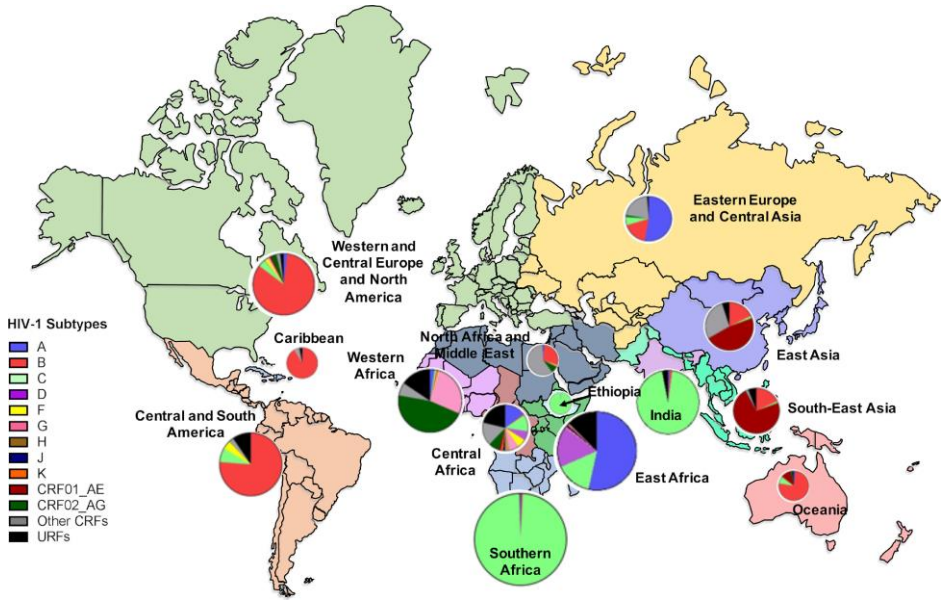


Figure 7. World map illustrating the prevalence of HIV-1 group M subtypes within each region.

Pie graphs depict the percentage of each subtype that circulates in each region, with the size of each pie representing the total number of infections in that region. The figure was adopted with permission from[142]

HIV replication cycle

HIV-1 infects cells that express its main entry receptor, CD4, and co-receptors. T-lymphocytes and macrophage are primary target, but the virus can also infect many other cells including monocytes, dendritic cells, microglia, and astrocytes. HIV-1 transmission is generally caused by virus exposure at mucosal surfaces, followed by virus replication in submucosal and locoregional lymphoid tissues, and finally by overt systemic infection [143]

The initial step of the HIV-1 replication cycle involves the binding and fusion of the gp120 to the CD4 receptors[144-147]. The binding induces a conformational change that creates a new recognition site on gp120 for the binding to co-receptors, particularly CCR5 and CXCR4[147-150]. HIV-1 are often classified based on their co-receptor usage as R5-tropic viruses that use CCR5, X4-tropic viruses that use CXCR4, and dual-tropic, R5X4 viruses that use both CCR5 and CXCR4 co-receptors. R5-tropic viruses preferentially infect cells of macrophage, dendritic and T-lymphocytes lineage, whereas X4-tropic viruses are generally restricted to T-lymphocyte lineage. The co-receptor switch from R5 to X4 virus is associated with increased CD4 + cell depletion and clinical deterioration, implying that coreceptor usage strongly influences disease progression[107, 134, 151]. Genetic mutations in the gene encoding

CCR5 are also associated with altered susceptibility to HIV-1 infection between different populations. Homozygosity for the 32 base-pair deletions (CCR5Δ32) in the CCR5 coreceptor gene is associated with a high level of resistance to HIV-1 infection[152]. About 1% of the Caucasian population inherits CCR5Δ32, which protects them from sexual transmission of HIV-1 infection, whereas individuals who are heterozygous for this deletion (about 20% of the population) are not protected from AIDS but have a delay in disease progression[152-155]. CCR5Δ32 mutation has not been found in people of African and East Asian populations[156, 157].

The binding to the co-receptor induces further conformational changes in gp41, which results in the exposure of the highly hydrophobic N-terminal fusion peptide of gp41, previously buried in the spike structure. The fusion peptide inserts into the target cell membrane, forming a stable six-helix bundle composed of two sets of heptad repeat regions (HR1 or N-terminal heptad repeat and HR2 or C-terminal heptad repeat), called "hairpin" conformation, resulting in virus-host cell into close physical proximity and allows membrane fusion and virion core entry [158](**Fig. 8**). The virus core, which includes the HIV genome, two copies of genomic RNA, the viral proteins CA, NC, IN, RT, and vpr, as well as the cellular protein cyclophilin A, enters the cell's cytoplasm and forms the pre-integration complex (PIC)[159, 160]. Upon entry into the host cell, the viral RT enzyme within the PIC reverse transcribes the single-stranded RNA genome to double-stranded complementary DNA (cDNA). Following reverse transcription, the PIC is actively transported through the cytoplasm into the host nucleus via nuclear pore complexes[159, 161] (**Fig. 8**).

IN catalyzes the integration of the double-stranded, linear viral DNA in the host genome, preferably in the active and thus open region of the human genome. IN catalyzes two critical steps in viral integration: the 3' end processing reaction, which removes GT-dinucleotide from the 3' ends of the proviral DNA, and the strand transfer steps, which involve viral integration to host DNA alongside host DNA repair enzymes to permanently join the HIV cDNA to the host chromosome. The viral cDNA that has been integrated is referred to as proviral DNA. During the infection, however, various unintegrated forms of HIV DNA, including one-LTR circles, two-LTR circles, and linear forms, are found within the nucleus. During integrase inhibitor therapy, more unintegrated forms are detected.

Following integration, the provirus will be transcribed, resulting in the expression of viral proteins required for viral particle formation, or it may remain silent and persist for the lifetime of the infected cell (long-lived, latent reservoir in the host). The latency explains the inability of viral therapies used to date to completely eliminate the virus from infected individuals, and it is the greatest challenge to a complete cure for HIV. The integrated HIV-1 provirus is the template for the viral RNA transcripts produced by the cellular RNA polymerase II enzyme. The activation of HIV transcription and gene expression from the integrated provirus is dependent on the activity of both cellular and viral factors. The binding of NF-kB and other transcription factors enables the proviral genome to be transcribed.

Transcription is also aided by the viral protein Tat, which binds to the transactivating responsive sequence (TAR), an RNA element responsible for viral transcription initiation and elongation from the LTR promoter. Unspliced or partially spliced transcripts are exported from the nucleus to the cytoplasm via active transport mediated by the viral Rev protein. The gpl60 Env precursor is translated within the endoplasmic reticulum and Golgi apparatus, whereas the Gag and Gag-Pol polyproteins are synthesised by cytoplasmic ribosomes. The Gag domain initiates translation, and Gag-Pol transcripts are generated via a frameshift process that allows the termination codon between the two genes to be bypassed[16].

Viral proteins are then transported to the inner surface of the plasma membrane, where they accumulate and condense to form an immature virion. As the particle exits the cell, it acquires a lipid coat containing mature TM and SU envelope glycoproteins, resulting in cellular death. The viral protease performs the final maturation step of proteolytic cleavage of the Gag-Pol polyprotein after budding[162]. The average generation time of HIV-1 has been estimated to be 1.2-2.6 days[102].

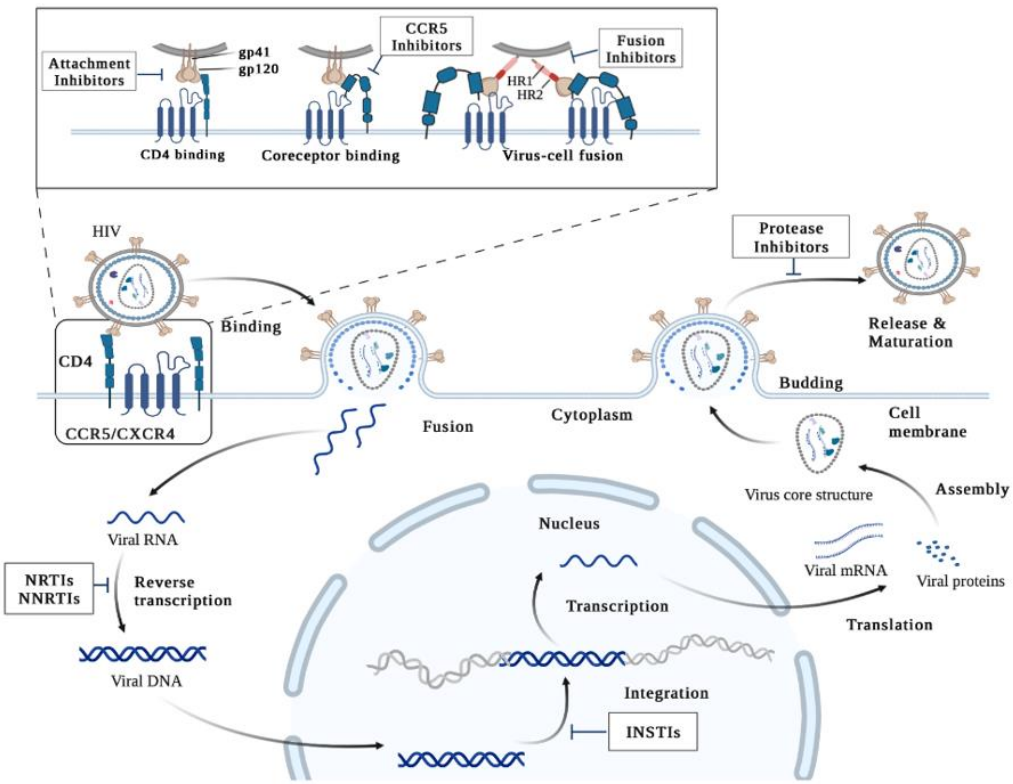


Figure 8. HIV-1 replication cycle and site of action of different antiretroviral inhibitors.
The infection begins when the envelope (Env) glycoprotein spikes engage the receptor CD4 and the membrane-spanning co-receptor, CCR5, leading to fusion of the viral and cellular membranes and entry of the viral particle into the

cell. Extracellular virions enter their target cell via a three-step process that includes binding to the CD4 receptor, binding to the CCR5 or CXCR4 coreceptors, or both, and membrane fusion. After fusion and uncoating, the viral RNA is then reverse transcribed into DNA. The pre-integration complex is imported into the nucleus, and the viral DNA is then integrated into the host genome. Mediated by host enzymes, HIV DNA is transcribed to viral mRNAs. These mRNAs are then exported to the cytoplasm, where translation occurs to make viral proteins. Finally, viral RNA is packaged in new capsid envelopes and released from the cell as newly formed, intact, mature infectious virions. Each step in the HIV-1 life cycle (HIV entry, reverse transcription, integration and protein maturation) is a potential target for antiviral intervention. The sites of action of clinical inhibitors are indicated in the white box. Abbreviation: nucleoside/nucleotide reverse transcriptase inhibitors, NRTI; non-nucleoside reverse transcriptase inhibitors, NNRTI; Integrase strand transfer inhibitor, INSTI.

HIV transmission

HIV is primarily transmitted through sexual contact, contaminated injection equipment and blood transfusions, and mother-to-child transmission during pregnancy or delivery. Primary modes of transmission vary by region and population groups. Sexual contact accounts for roughly 80% of all HIV infections worldwide while injection drug use accounts for approximately 10%. However, this percentage is much higher in countries with large IDU populations, such as Eastern Europe and Central Asia. In sub–Saharan Africa, the transmission mode is mainly heterosexual, while in south and Southeast Asia, the HIV-1 epidemic started among injection drug users and sex workers and subsequently spread to the general population through heterosexual transmission[163].

Factors that influence the risk of infection are the amount of infectious virus particles in infected body fluids, co-infections such as other STIs (notably genital ulcers of any cause[164], herpes simplex type-2 infection[165], herpes simplex type-2 infection [165], and bacterial vaginosis [166]), behavioral factors, the extent of exposure to the body fluid, and the route of transmission[167-170]. Transmission events generally occur when the source partner has a high viral concentration (>3.5 Log_{10} copies/mL)[171]. A 1 log_{10} increase in plasma HIV-1 RNA increases the risk of sexual transmission by 2.4 times[172], while a 0.7 log_{10} decrease in plasma viral load is estimated to reduce HIV-1 transmission by 50%[173]. Successful ART suppresses viral replication (plasma HIV-1 RNA levels to < 50 copies/mL), resulting in lower concentrations of HIV in blood [174] and other biological fluids such as sperm[175], vaginal fluids[176], and anal mucosa[177]. Undetectable viral load means the virus cannot be transmitted through unprotected sexual intercourse (known as U=U, undetectable equals untransmissible)[172, 178-182]. Individuals undergoing treatment have a lower risk of transmitting HIV vertically[183], sexually[172], and by sharing needles[184]. Consequently, effective early treatment and behaviour modification programmes, counselling for serodiscordant couples, harm reduction programmes for injecting drug users, and male circumcision will all help to reduce HIV transmission[172, 178, 179, 181]. The risk of HIV transmission through sexual contact is lowered by 60% in men who have undergone male circumcision[185].

Pathogenesis of HIV-1 infection

HIV-1 infection is characterized by the progressive destruction of CD4-expressing T lymphocytes, macrophages, and monocytes, followed by a loss of immunocompetence. The clinical course of HIV infection is classified into three stages: acute (or primary), asymptomatic, and AIDS (**Fig. 9**). The acute phase of the disease corresponds to the time between the detection of viral particles in blood serum and plasma and the production of specific antibodies.

HIV transmission across mucosal membranes is typically established by a single founder virus with distinct phenotypic properties such as usage of CCR5 rather than CXR4 for entry[186], enhanced interaction with dendritic cells, and resistance to interferon-α[187]. In the event of sexual transmission, virus-infected cells are a plausible source, as they can be present in vaginal or seminal fluids in considerably greater numbers than free virus particles. After being exposed to the mucosal surface, the virus is carried to the draining lymph node either as a virus-infected cell or as a free virus attached to a dendritic cell, where it is extensively replicated. For the first few days to several days, the virus cannot be detected in the plasma; this period is known as the "eclipse phase," and it usually lasts 7 to 21 days[160, 167]. During the acute phase, viral replication is high, reaching up to 10^{10} copies/ml, particularly in the gut-associated lymphoid tissue (GALT), where large numbers of CD4+ T cells (usually memory cells) are infected and depleted. During this phase, 30 to 60% of CD4+ T cells in the gut are destroyed, either directly or indirectly via bystander effects[188-191]. The virus establishes long-term reservoirs in lymphoid tissue as well as latently infected resting CD4+ cells during this phase[192, 193]. The virus's persistence in these reservoirs is a major impediment to antiviral treatment's ability to eradicate the virus from the body. Within two to six weeks after infection, approximately 50 to 70% of individuals develop a flu-like illness (an acute HIV syndrome) characterized by fever, swollen lymph nodes, a sore throat, arthralgia (joint pain), myalgia (muscle aches), headache, fatigue, weight loss, and occasionally a rash. These symptoms usually lasts 7 to 10 days[160, 194].

Acute infection has been shown to be a significant risk factor for HIV transmission due to the high viral load combined with unmodified risk behavior[195-197]. As a result, early diagnosis and ART will significantly impact both HIV transmission prevention and individual health outcomes[198]. Furthermore, early diagnosis and ART initiation will reduce the viral reservoir [199, 200] and improve partner notification outcomes[201].

The initial peak of viremia is greatly reduced within a few weeks of infection as the individual mounts a vigorous cellular and humoral immune response. The appearance of virus specific CD8+ cytotoxic T cells and neutralizing antibodies, which will destroy the virions via phagocytosis, is evidence of these responses. The period between infection to the detection of antibodies is called the serological

window period. These immune responses are reflected by the new rise in CD4+ cell count and a drop in viremia to a lower steady state (the “virologic set point”), the absolute levels of which vary between individuals but are usually between 10^3 and 10^5 copies/ml. The rate of disease progression is determined by viral set point values; the higher the set point of plasma HIV-1 RNA, the faster the patient will lose CD4+ T cells and progress to AIDS[202, 203].

The asymptomatic period, the chronic stage of infection, begins when the immune response to the HIV-1 infection is fully developed. During the asymptomatic stage, the HIV-1 virus continues to replicate at low levels causing the gradual depletion of CD4+ T cells, and the infected individual may not experience any clinical symptoms. The period of the asymptomatic stage varies significantly between individuals. In the absence of treatment, most patients with HIV infection progress to AIDS within 7 to 10 years, but some individuals, referred to as rapid progressors, who progress to AIDS within 2-3 years and account for 10–15% of the HIV infected population. While about 5% to 15% of HIV-infected individuals maintain a low viral load and a high CD4+ cell count for long periods, possibly indefinitely, without antiviral therapy, they are called long-term non-progressors (LTNPs). Among these LTNPs, there are a few individuals who have a low viral load, almost undetectable in the absence of treatment, and they are known as "elite controllers (EC)"[204].

The difference in asymptomatic phase duration and disease progression rate is most likely due to a combination of host and viral factors. Some naturally attenuated viruses, including nef deletion, are associated with slower disease progression[205]. Possession of a specific MHC class I allele (certain HLA types) is also linked to different prognoses: HLA-B57 and HLA-B27 are linked to slower progression, whereas HLA-B35 is linked to faster progression[206]. Individuals who have a mutation in the chemokine receptor CCR5 have a delay in disease progression[152-155].

The progressive decline in CD4+ T cell levels eventually leads to immune system exhaustion, and the immune system fails to control the HIV infection. AIDS is the end stage of the disease which can be characterized by the breakdown of host defense, associated to progressive damage to the populations of CD4+ lymphocytes, an increase in plasma virus and the presence of one or more AIDS-defining illnesses, such as opportunistic infection and specific malignancy, which usually occur after the CD4 dropped below 200 cell/mm^3.

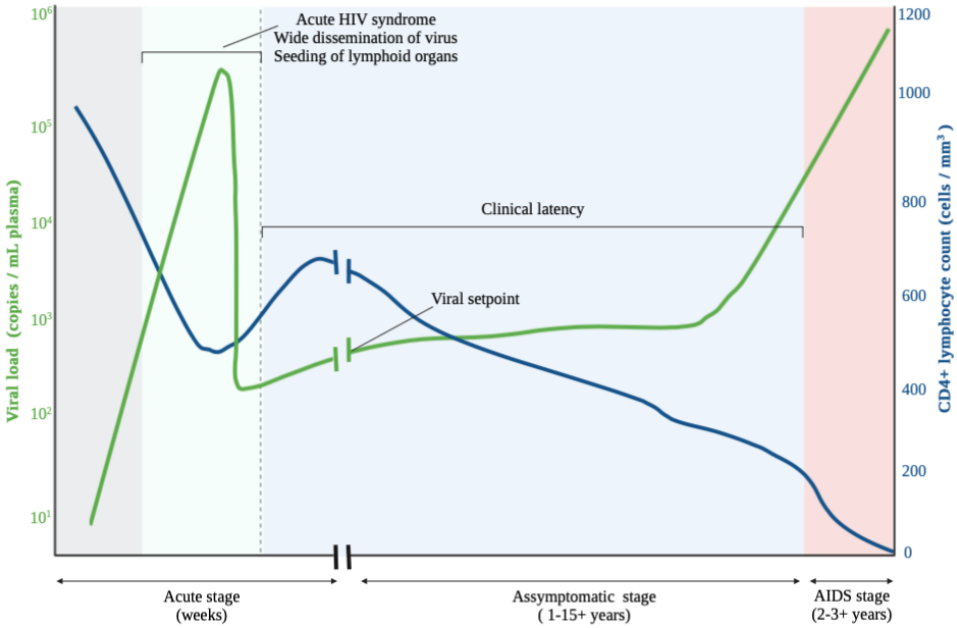


Figure 9. The natural course of HIV-1 infection in the absence of treatment.

During the acute stage, there is a rapid increase in viremia and a decrease in CD4+ T cell count.Following the acute stage, a relatively stable viral load (viral setpoint) is achieved, which lasts throughout the asymptomatic stage. During the early stages of AIDS, there is a rapid increase in viremia, a sharp decline in CD4+ T cells, and the onset of AIDS-defining illnesses. The diagram was adapted from Maartens *et al.*, 2014, with modification for clarity[171].

Antiretroviral therapy

Since the development and approval of the first antiretroviral drug, azidothymidine (AZT), in 1987, significant progress has been made in the treatment of HIV infection. The clinical benefit of this drug was limited for only two years due to the emergence of HIVDR. In the years that followed, dual therapy with newly introduced NRTI of didanosine (ddl) or zalcitabine (ddC) provided some initial benefit[207], but the effect was again short-lived due to HIVDR. The treatment of HIV-1 infection was revolutionized in the mid-1990s, as the standard of care in HIV management advanced to include the administration of combination antiretroviral therapy or highly active antiretroviral therapy (HAART) that utilizes drugs from more than one class of HIV-1 drugs. The main goals of ART are to maximally and durably suppress plasma HIV-1 RNA, which enables immune system reconstitution and maintenance, reduced HIV-1-associated mortality and morbidity, improved health-related quality of life, and HIV transmission prevention[208-210].

ART has transformed HIV/AIDS from a lethal condition to a treatable and potentially preventable chronic disease. ART has contributed to both individual treatment success and population-level HIV transmission reductions. Between 2004 and 2014, ART is estimated to have prevented 7.8 million deaths and 30 million new infections in LMICs[210]. As of June 2021, 28.2 million people were receiving ART worldwide. Current treatment guidelines recommend lifelong ART immediately following HIV diagnosis regardless of CD4+ T cells count. Prevention of mother-to-child transmission (PMTCT) was one of the earliest and most significant successes in the field of HIV prevention[208]. It is estimated that the implementation of PMTCT services saved 1.2 million lives and 2.5 million HIV infections[208, 209]. Currently, two-drug (tenofovir/emtricitabine (TDF/FTC, Truvada) pre-exposure prophylaxis (PrEP) or three-drug post-exposure prophylaxis (PEP) oral regimens are used for HIV prevention and have been shown to be effective when used correctly in persons at high risk of HIV acquisition (for example, MSM with multiple sexual contacts and female sex workers[179, 180]. The impact on the HIV epidemic control has been demonstrated when these HIV treatment and prevention strategies are scaled up. Importantly, initiation of ART immediately or shortly after HIV diagnosis in people with HIV who have HIV-negative partners (discordant couples) has been shown to reduce the risk of HIV transmission by 96%[179, 180].

The therapeutic arsenal continues to improve with the advent of new classes of drugs targeting viral entry and integration. To date, the U.S. Food and Drug

Administration (FDA) has approved and recommended more than 40 antiretroviral drugs for HIV treatment, which are divided into seven classes: (1) nucleoside/tide reverse transcriptase inhibitors (NRTIs), (2) non-nucleoside reverse transcriptase inhibitors (NNRTIs), (3) protease inhibitors (PIs), (4) integrase strand transfer inhibitors (INSTIs), (5) Attachment Inhibitors (AIs), (6) CCR5 antagonists and (7) fusion inhibitors (FIs) (see below paragraphs).

Over the past two decades, NNRTI-based triple-combination therapy has been the predominant first-line treatment in LMICs. NNRTIs were initially combined with NRTIs, including lamivudine (3TC), zidovudine (AZT), or stavudine (D4T). However, due to the increasing PDR to NNRTI, dolutegravir (DTG)-based treatment is currently the preferred first-line regimen. According to 2018 WHO updated guideline, the preferred first-line regimen for adults and children are a combination of TDF + 3TC (or FTC) +DTG and ABC + 3TC + DTG, respectively therapy[208]. More complex combinations with these or other classes (2 NRTIs + ritonavir-boosted lopinavir (LPV/r) or ritonavir-boosted atazanavir (ATV/r), or 2NRTIs + DTG) are used for second-line or salvage therapy[208, 211].

CCR5 antagonist

CCR5 antagonists bind to the CCR5 coreceptor of CD4+ T cells, induce a conformational change that impedes CCR5 interaction with HIV gp120, thereby preventing initialization of the gp41-TM-mediated membrane fusion[160, 212]. Maraviroc (MVC) is the only CCR5 co-receptor antagonist in clinical use, although others are in development. MVC is only effective against R5-tropic viruses, limiting clinical use and requiring prior co-receptor testing (**Fig. 8**).

Mechanism of resistance: HIV resistance to CCR5 antagonists develops either from the outgrowth of pre-existing CXCR4-using viruses or from the ability of CCR5-using HIV-1 to use the antagonist bound form of CCR5. MVC resistance can develop through the selection of mutations within different regions of gp120 and a shift in tropism from R5-tropic to X4-tropic, particularly in patients with pre-existing X4 variants.

Fusion inhibitors (FIs)

Enfuvirtide (ENF) selectively inhibit the function of gp41. FIs are peptide mimetics of the HR2 that bind to HR1 and prevent the conformational changes of gp41 that are required for membrane fusion to occur. ENF is the only fusion inhibitor currently in clinical use, but it requires twice-daily intravenous administration, limiting its clinical use[213](**Fig. 8**).

The mechanism of resistance is not fully understood. However, mutations in the HR1 region of gp41 disrupts HR1-HR2 interaction which may increase discrimination between the FIs and the native HR2 from CCR5 to CXCR4 or dual

tropic[214]. The major ENF resistance mutations in gp41 are G36D/E/V, V38E/A, Q40H, N42T and N43D[215].

Nucleoside/tide reverse transcriptase inhibitors (NRTIs)

NRTIs inhibit HIV RNA-dependent DNA polymerase (RT) by acting as structural analogues for purine or pyrimidine when DNA is reverse transcribed inside cells leading to premature DNA chain termination and inhibition of viral replication. NRTIs lack a 3′-hydroxyl (3’-OH) on their ribosome ring and act as chain terminators when they are incorporated into elongating viral DNA by RT[216] (**Fig. 8**). Currently available NRTIs include, lamivudine (3TC), emtricitabine (FTC), abacavir (ABC), tenofovir (TDF), Tenofovir alafenamide (TAF) and zidovudine (AZT).

Mechanism of resistance: There are two mechanisms of resistance to NRTIs, discriminatory mutation and primer unblocking mutation. Primer unblocking mutations also referred to as thymidine analogue mutations (TAMs), facilitate the phosphorylytic excision of an NRTI-triphosphate from viral DNA[217]. TAMs include M41L, D67N, K70R, L210W, T215F/Y and K219Q/E. The discriminatory mutation enables RT to distinguish between NRTIs and the cell’s own dNTPs, thus preventing NRTIs from being incorporated into viral DNA. The most common discriminatory mutations include M184V/I, K65R, K70E/G/Q, L74V/I, Y115F and the Q151M complex of mutations.

Non-nucleoside reverse transcriptase inhibitors (NNRTIs)

NNRTIs are non-competitive inhibitors that bind the substrate pocket of the reverse transcriptase and induce conformational changes in reverse transcriptase, thereby reducing polymerase activity and impeding proviral DNA synthesis[218, 219](**Fig. 8**). NNRTIs are effective against HIV-1 but not HIV-2 because HIV-2 RT contains amino acids that confer intrinsic resistance to NNRTIs. Tyrosine residues were found in HIV-1 RT at codons 181 and 188, but HIV-2 contains isoleucine at 181 and leucine at 188, both of which can prevent NNRTIs from binding to HIV-2 RT[220]. Currently available NNRTI are nevirapine (NVP), delavirdine (DLV), efavirenz (EFV), etravirine (ETR), rilpivirine (RPV), and doravirine (DOR).

Mechanism of resistance: Amino acid mutations near the drug-binding pocket of NNRTIs can be selected to induce drug resistance under the selective pressure of NNRTIs. The majority of the amino acid mutations that confer resistance to NNRTIs are found within the HIV-1 polymerase domain (especially codon groups 100 - 108 and 181 - 190). Because all NNRTIs have an overlapping binding pocket, cross-resistance to different NNRTIs develops quickly[215]. A single mutation, for example (L100I, K101P, Y181C/I, G190A) can result in resistance to most NNRTIs such as NVP, EFV, and ETR. The early generation NNRTIs (DLV, NVP, EFV) have a low genetic barrier to resistance that resistance-associated mutations can

quickly induce high drug resistance. For example, a single mutation such as K103N and Y181C can result in a significant loss of NVP and EFV potency[221]. The new generation of NNRTIs (e.g., rilpivirine, doravirine) have a high genetic barrier may be effective against HIV-1 strains resistant to early-generation NNRTIs. Doravirine and rilpivirine, for example, remain active against HIV-1 strains with drug resistance mutations such as K103N[222]. The common NNRTIs resistance mutations are L100I, K101E/P K103N/S, V106A/M, Y181C/I/V, Y188L/C/H, G190A/S/E and M230L[215].

Integrase strand transfer inhibitors (INSTIs)

HIV-1 integrase is a 32-kilo Dalton (kDa) protein which plays a vital role in the HIV-1 replication cycle by catalysing the two distinct reactions termed: 3'-end processing and strand transfer. During the 3' processing, IN removes two nucleotides from the 3' ends of both viral DNA strands, and strand transfer covalently links the viral and host DNA. IN consists of three structural and functional domains: the N-terminal domain (NTD) (aa:1–50), which contains a highly conserved histidine–histidine–cysteine–cysteine ($H_{12}H_{16}C_{40}C_{43}$) motif that coordinates zinc binding and favours multimerization of the IN subunit; the catalytic core domain (CCD) (aa: 51–212), which contains the catalytic triad $D_{64}D_{116}E_{152}$ (known as the DDE motif) that plays an essential role in IN enzymatic activity; and the C-terminal domain (CTD) (aa: 213–288), which is involved in binding to viral and cellular DNA, and in protein oligomerization and interactions with the reverse transcriptase.

INSTI block the HIV-1 integrase strand transfer steps by binding to the metal cation (Mg^{2+}), blocking the enzyme active site and prevent the formation of the covalent bond with host DNA, thus, inhibiting the incorporation of viral DNA into the host genome (**Fig. 8**). There are currently five US FDA -approved drugs belonging to INSTI: raltegravir (RAL), elvitegravir (EVG), DTG, bictegravir (BIC), and cabotegravir (CAB)[223]. RAL and EVG were the first-generation INSTIs and have relatively low genetic barriers for resistance and extensive cross-resistance between them. The major resistance mutations for EVG and RAL are R263K, N155H, Q148HKR, S147G, F121Y, E92Q and T66I and the accessory mutations include T97A, E92G and T66AK[224]. DTG, BIC and CAB are second-generation INSTIs with a higher genetic barrier to resistance[225, 226]. Following the global increase of PDR NNRTIs, WHO recommends using DTG-based regimens in both treatment-naïve and treatment-experienced patients. Structural and functional characterization studies of DTG have shown that DTG has high flexibility and antiviral activity, attributed to the length and flexibility of the linker connecting the tricyclic metal-chelating core and difluorophenyl ring[226] to adjust its position even in the presence of mutations. DTG has a prolonged residence time at the active site with a half-life of 71 hours for dissociation from the wild-type integrase-DNA complex, which contributes to the high efficacy and higher genetic barrier[226, 227]. DTG is a potent

drug with the inhibitory quotient (IC_{90}) of 0.064 lg/mL, which is 17-fold below plasma concentrations observed at the end of the dosing interval for a once-daily, 50-mg dose[226, 228]. DTG has a wide therapeutic index and longer elimination half-life (14 hours) and appears to have good pharmacokinetic forgiveness in case of suboptimal adherence[226, 229].

Mechanism of resistance: Despite the high genetic barrier to resistance of DTG, there is still potential for the emergency and transmission of HIVDR to DTG[230-235]. The major mutations associated with DTG resistance are R263K, Q148HKR and G118R, with some minor mutations (i.e., N155H, G140AS, E138AKT, T66K, E92Q and F121Y) also being accessory to DTG resistance. Some risk factors associated with DTG resistance include poor treatment adherence, low CD4 + T-cell count or high VL at the time of DTG initiation, and drug-drug interactions from co-treatment[224, 236].

Protease inhibitors (PIs)

PI are competitive inhibitors of the protease enzyme and bind at the catalytic site with high affinity and thereby block its activity. Inhibition of HIV protease enzymes results in immature and non-infectious viral particles[214, 237]. PIs are usually co-administered with a drug that boost their plasma drug level; either low dose ritonavir or cobicistat (potent inhibitor of CYP3A4 metabolism). Currently available PIs are atazanivr (ATV), daunavir (DRV), fosamprenavir (FPV), saquinavir (SQV) and tipranavir (TPV) (**Fig. 8**).

Mechanism of resistance: Resistance to PIs occurs primarily due to mutations within or proximal to the catalytic binding site of the drug, which reduces the affinity of the protease to the PI. PIs have a high barrier to resistance, and the accumulation of multiple PI mutations is required for clinically relevant resistance. Mutation in the *gag* and *env* gene are also associated with PI drug resistance[214]. The major PI mutations are D30N, V32I, M46IL, G48V/M, I50V/L, I54V/T/A/L/M, L76V, V82A/T/F/S, I84V, N88S and L90M[215].

Attachment inhibitors (AIs)

Fostemsavir (FTR) is the only attachment inhibitor approved by FDA as of 2022. FTR bind to gp120 and stabilizes gp120 in a conformation, and prevents the conformational changes required for the eventual exposure of gp41 for fusion. Due to the novel mode of inhibition with no cross-resistance with other antiretroviral drug classes, FTV was approved for use in highly treatment-experienced patients who had limited treatment options due to drug resistance to currently approved drugs (**Fig. 8**). However, treatment-emergent gp120 genotypic substitutions at four key sites S375, M434, M426, and M475 have been linked to decreased susceptibility to fostemsavir[238].

Post-attachment inhibitors: Ibalizumab is the only FDA-approved post-attachment inhibitor. Ibalizumab binds to the CD4 extracellular domain, thereby preventing conformational changes of the gp120 complex that are essential for viral entry. Ibalizumab has a broad effect on R5 and X4-tropic strains, with no evidence of drug-drug interactions or cross-resistance[238].

Antiretroviral therapy in Ethiopia

In Ethiopia, the ART program was launched as a fee-based system in 2003[239], and free ART was launched in 2005 in selected health facilities with support from the government, the US President's Emergency Plan for AIDS Relief (PEPFAR), and the Global Fund to Fight AIDS, Tuberculosis, and Malaria[240]. In 2008, Ethiopia decentralized the service and made it available in an increasing number of health centers and hospitals across the country[240]. The integration of the service at lower-level health facilities, as well as task shifting to low- and mid-level health workers, has greatly aided in the scaling up of ART in Ethiopia[241]. As of 2020, 1123 health facilities, hospitals and health centres are providing ART services. In 2017, Ethiopia adopted the WHO recommendation to initiate treatment for all HIV+ patients at the time of diagnosis, regardless of CD4 count status[242]. As of 2020, about 465,457 adults (Age >15) and 17,670 children (age <15 years) were receiving ART, with an approximately 79% national coverage. Adult ART coverage has reached 80.5%, but coverage for children living with HIV remains low (40.03 %)[90, 241].

Following the WHO recommendation, the Ethiopian national ART program, has adopted the DTG-based regimen as the preferred first- and second-line treatment for all populations. The preferred standard first-line ART regimens for adolescents (age 10-19 years) and adults is a fixed-dose combination of TDF+3TC+DTG, known as tenofovir-lamivudine-dolutegravir (TLD). For children (age <10 years), the preferred first-line regimen is ABC+3TC+DTG. However, there are alternative first-line regimen for adult including, TDF + 3TC + EFV or AZT + 3TC + DTG or AZT + 3TC + DTG[241]. For children, the alternative first-line regimen includes ABC+ 3TC+LPV/r or AZT+3TC+DTG. The preferred second-line ART regimen for adults is AZT+3TC+ ATV/r or LPV/r, while for children, the preferred second-line regimen is AZT+3TC+LPV/r. Boosted PI + two NRTI combinations is the preferred second-line ART regimen for adults and adolescents (AZT+3TC+ ATV/r or LPV/r) and children (AZT+3TC+LPV/r). Two NRTI + DTG can also be used as a second-line regimen if it is not used in the first line[241].

HIVDR is a serious threat to HIV epidemic control in Ethiopia, as in many resource-limited settings, due to suboptimal patient monitoring and insufficient access to viral load testing for early detection of virological failure. Furthermore, there are various barriers to an effective ART program that may lead to the development of HIVDR,

such as poor ART adherence, a limited HIV care specialist, drug stockout (irregular drug supply), and a lack of HIV drug resistance testing. Despite the WHO's recommendation that HIV treatment expansion be accompanied by a comprehensive assessment of HIVDR emergence and transmission, Ethiopia lacks a nationally representative HIVDR prevalence estimate for ADR and PDR. However, the few studies that have been conducted have revealed an increase in PDR, as well as detection of major NRTI and NNRTI DRMs among patients with virological failure with the scale-up ART in the country.

In 2008, a TDR survey conducted in accordance with WHO guidelines among ART-naive individuals attending public health facilities in Addis Ababa revealed a low level of TDR (5%)[243]. TDR prevalence was 3.3% in 2009 and 5.6% in 2011 in two other studies conducted at Gondar University Hospital among ART-naive patients seeking ART[244, 245]. Another study, based on samples collected from seven university hospitals across the county between 2009 and 2011, showed 3.9% TDR[246]. A study by Tadesse *et al*. showed 14% of PDR among treatment naïve HIV infected children, with 9% associated only with NNRTI and 5% having dual NRTI and NNRTI DRMs[247]. Similarly, the recent nationally representative HIVDR survey in Ethiopia has shown a 13.2% prevalence of PDR to NNRTI (NVP or EFV) [248].

Similarly, following the expansion of ART coverage in Ethiopia, the few studies available have also revealed a significant increase in the incidence of virological failure and the accumulation of major acquired NRTI and NNRTI DRMs. A prospective cohort study conducted in Gondar to assess the treatment outcome of ART revealed an increase in virological failure and accumulation of major acquired NRTI and NNRTI DRMs over time, implying that prolonged ART exposure drives the emergence of mutant variants[249]. Similarly, a study done among adults receiving health center–based ART in Adama, Ethiopia, showed 65.3% of virological failure during the first year of ART was due to acquisition of DRMs and all were associated with NNTRI while 35.7% were associated with NRTIs. M184V/I, K65R and K103N were the most common NRTI and NNRTI-associated HIVDRMs detected in this study[250]. In another study, based on data collected from seven hospitals from different parts of the county in 2009-2011, DRMs were detected in 76.6% and 66.7% of patients with VLs >1000 copies/mL after six and twelve months of ART, respectively[251]. A recent study among patients on a second-line regimen from two HIV clinics in central Ethiopia showed that 80% of the virological failure were associated with HIVDR[252].

HIV-1 drug resistance

ART has been widely implemented around the world, resulting in significant reductions in HIV-1 mortality and incidence. To help achieve these goals, WHO recommended that all HIV-infected individuals begin ART as soon as possible after diagnosis (universal test and treat) and that pre-exposure prophylaxis (PrEP) be considered for people at high risk of HIV infection. However, there are still significant gaps in the quality of ART service delivery in many settings, including poor ART adherence, high patient attrition, unreliable drug supply chains, and suboptimal viral suppression rates. As a result, the dramatic increase in ARV use is likely to increase the emergency and transmission of HIVDR, posing a serious threat to the long-term efficacy of ART and posing a threat to the 2030 goal of eradicating AIDS as a public health concern[253]. HIVDR limits the number of effective drugs, increases the potential for onward transmission, and compromises survival[254, 255]. This is especially significant in low- and middle-income countries, where routine viral load monitoring is lacking, treatment options are limited, and HIVDR testing is neither practical nor routinely recommended for patient monitoring. HIVDR to ART can be acquired (ADR) when there is viral replication in the presence of a drug or can develop as a result of infection with a drug-resistant viral strain, (transmitted drug resistance (TDR)).

Mechanisms of HIV-1 drug resistance

HIVDR is invariably evolutionary driven phenomenon. Due to the high rate of HIV-1 viral replication, combined with the high error rate of reverse transcriptase and frequent recombination, a swarm of different but genetically related viral variant circulate in an infected individual. Existing virus replication models predict that any single mutant and many double mutants can be generated daily by viral replication. Each viral variant in the quasispecies has a different replicative capacity and drug susceptibility[256, 257]. Resistance-conferring mutations have low replication capacity or fitness in comparison with the wild-type (WT) virus, and thus WT variant predominates in the absence of drug selective pressure. In the presence of therapy, suboptimal treatment mutant strains can emerge as the dominant viral population (selection for mutants with a fitness advantage). Mutant selection occurs at a rate proportional to replication level and relative fitness advantage in that particular environment. Persistent viremia in the presence of drug selective pressure results in the accumulation of DRMs that increase resistance or improve viral fitness. Because DRMs usually impose a fitness cost on the virus in the absence of therapy, ART discontinuation usually results in the rapid decay of mutants and the re-emergence of the WT virus. Such mutants persist as minority variants in viral quasispecies and viral reservoirs and can reappear if drug selective pressure is applied again[256, 258, 259]. The viral quasispecies can withstand subsequent immunologic or pharmacologic

pressure by generating escape mutants in advance and storing evolved resistant variants. The likelihood of developing HIVDR is determined by the antiretroviral regimen's relative potency and the degree of ongoing replication (replication capacity) in the presence of therapy. A highly effective regimen that reduces viral replication to minimal levels is associated with slow resistance accumulation, while an incompletely suppressive therapy tends to increase the virus's selective pressure, which rapidly accumulates DRMs[256]. Incompletely suppressive therapy could result from poor adherence, incorrect dosing, poor absorption, or reduced drug levels due to drug-drug interactions. The emergence of DRMs is also associated with the genetic barrier to resistance of the ART regimen, which is a function of the number of mutations required to reduce viral susceptibility. NNRTIs, some NRTIs (3TC, FTC), and first generation INSTIs have a low genetic barrier to resistance, PIs and second-generation INSTIs require multiple mutations before drug susceptibility is compromised.

Impact of HIV-1 drug resistance

As resistance mutations accumulate, drug susceptibility decreases, progressively weakening the effectiveness of ART. Continued replication in the presence of the drug leads to even higher levels of resistance to each drug administered and progressive cross-resistance to other drugs in the same family. As a result, the therapeutic arsenal available for salvage therapy is reduced, resulting in the prescription of more complex, expensive, and frequently poorly tolerated regimens. HIVDR-associated virological failure creates a vicious circle in which ARV options are reduced, consecutive treatment lines are associated with progressively reduced duration of antiviral efficacy, and each new virological failure is associated with resistance accumulation. This vicious circle can result in subjects developing viruses that are resistant to all classes of drugs. Resistant viruses in genital secretions, blood, or milk can be transmitted to ART naïve individuals during sexual activity, needle-sharing, childbirth, or nursing. Hence emergency of HIVDR has a subsequent effect on treatment outcome both at induvial level and population level. Because different factors contribute the emergency and transmission of HIVDR, the prevalence and scope of the problem vary geographically. Patients in high-income countries are regularly monitored for viral load, and resistance testing is performed before ART initiation and if viral load increases indicate that the treatment is failing. This allows for a quick switch to a different drug regimen, lowering the risk of HIVDR accumulation and transmission. These approaches have resulted in stability and even a decrease in TDR in developed countries in recent years.

In contrast, resistance testing is not routinely available or recommended for PDR testing or resistance testing at the time of virologic failure on first-line ART due to the high costs in many LMICs. Treatment is based WHO recommended standard first-line and second-line regimen with limited treatment option. Furthermore,

several logistical issues, such as interruptions in drug supply and poor retention in care, all contribute to the development of HIVDR. The selection of second line regimens is not based on resistance testing in many LMICs, which frequently leads to a potentially suboptimal second-line regimen.

Pretreatment drug resistance

The term pretreatment drug resistance (PDR) defines DRMs detected in people with HIV before they start ART, resulting from either transmission of a drug-resistant strain (TDR) or previous exposure to ARVs, including prevention of mother-to-child transmission (PMTCT), pre-exposure prophylaxis (PrEP), post-exposure prophylaxis, or interrupted first-line ART. A recent meta-analysis suggests that TDR increased gradually over time and was associated with the duration of the rollout of ART. PDR is associated with poor virological outcomes, impaired immune recovery, reduced durability of first-line regimens, and increased mortality[260]. People with PDR were three times more likely to experience VF, and the clinical impact was even greater for those with NNRTI-associated PDR[260]. The prevalence of PDR has increased in several LMICs nations as the number of HIV patients receiving ART has increased.

Gupta and his colleagues, in their meta-analysis, showed that PDR has increased in several LMICs. They analyzed 385 datasets, which included 56,044 HIV-infected adults from 63 countries. The absolute prevalence of PDR increased by 0.3% in Asia and 1.8% in Southern Africa between 2015 and 2016. The odds of developing PDR are increasing by 23% in Southern Africa, 17% in Eastern Africa, 17% in Western and Central Africa, 11% in Latin America and the Caribbean, and 11% in Asia each year[261]. Similarly, the rapid increase in PDR was evident in a meta-analysis that included 19 studies covering 2617 children from 13 sub-Saharan African countries.

According to 2021 WHO HIVDR report, the prevalence of PDR to NVP and EFV in patients initiating first-line ART exceeded 10% in 21 of 30 reported surveys to WHO. The prevalence of PDR to NVP and EFV was three times higher in ART initiators with prior drug exposure compared to ART initiators without prior drug exposure. Among the ten nations that assessed PDR to integrase strand transfer inhibitors (DTG), only South Sudan detected an extremely low prevalence of DTG resistance (0.2%). In contrast, the prevalence of PDR has been declining in high-income countries over the last decade and has stabilised at around 10%[262].

The most common mutation reported was M184V/I which is associated with 3TC, and FTC followed by TAMs D67N, M41L, which confer resistance to AZT. The prevalence of PDR to boosted PI and INSTI is very low, reflecting their restricted use. According to this report, a PDR prevalence of 1.6% for TDF, 1.7% for FTC, and 1.8% for 3TC was documented, justifying the use of these drugs. In the same

report, the prevalence of PDR to NVP and EFV among newly diagnosed infants in ten sub-Saharan counties was 42 to 49%.

Because of the increase in pre-treatment drug resistance, particularly to NNRTIs, WHO has recommended DTG-based combination ART as first-line treatment. Although DTG has a higher barrier to resistance than NNRTIs, there are concerns that without adequate VL monitoring and drug-resistance testing, HIVDR to DTG will remain a concern[263-268].

HIV-1 drug resistance testing

HIVDR genotyping testing has been used for selecting therapeutic drugs, evaluating treatment efficacy, monitoring drug-resistant strain spread, and establishing new prevention strategies. The two types of assays used to assess HIVDR are genotypic assay (Sanger sequencing) and phenotypic assay.

Genotypic assay

Sanger sequencing is a population-based method that generates a single consensus sequence representative of the most common bases at each nucleotide position using dideoxy chain termination chemistry. HIVDR genotypic testing produces a nucleotide sequence of the enzymatic targets of ART, most commonly PR, RT and IN. The nucleotide sequence is then translated to amino acid sequence and compared to the wild-type subtype B laboratory strain (HXB2)[217]. The difference in amino acid sequence between the sequenced clinical virus and the reference wild sequence is reported as a list of mutations by the different HIVDR interpretation algorithms[217]. There are several HIVDR interpretation algorithms available, including Stanford HIV Drug Resistance Database (HIVdb), HIV Genotypic Resistance-Algorithm Deutschland (HIV-GRADE), French National Agency for AIDS Research (ANRS), and Rega[269]. Output from the HIVdb genotypic resistance interpretation system includes a list of penalty scores for each HIVDR mutation in a submitted sequence, estimates of decreased NRTI, NNRTI, PI and INSTI susceptibility, and comments on each DRMs in the submitted sequence[270].

Sanger sequencing has been the gold standard method as part of commercial kits or in-house-developed and validated protocols for HIVDR genotyping. It has been shown to be a highly reproducible and interpretable method producing high-quality sequence data, with short turnaround times and relatively simple workflows and data interpretation[271]. It does not, however, reliably detect mutations that are occurring at less than 20% of the viral population within the viral pool (low-abundance DRMs variants)[215, 272]. Low-abundance DRMs variants have been linked to an increased

risk of virological failure, the accumulation of HIVDRM, and impaired immune system recovery[273]; the effect is significant when DRM-bearing variants are present at frequencies ≥1% frequency[274, 275].

The next-generation deep sequencing (NGS) technologies have increased sensitivity and resolution for the detection of low-abundance DRMs variants (about 1% prevalence) with higher throughput and at lower cost[273]. The field is rapidly evolving, and various NGS platforms are now available [276]. However, using NGS for clinical monitoring is difficult due to the massive amount of data generated, the complexity of analysis, and the lack of standardized clinical interpretation that allows clinicians to act on the results for patient care[277]. Furthermore, implementing NGS for HIVDR genotyping in resource-limited settings is difficult due to infrastructure and equipment requirements and costs and other related issues[278].

However, the introduction of NGS has transformed the fields of virology and molecular epidemiology. It has provided an insight into viral pathogenesis, improved diagnosis, disease detection, instrumental in vaccine design (SARS-CoV-2), prevention, and treatment of many viral diseases. Several NGS sequencing platforms are currently available from various manufacturers, including Illumina (Illumina, CA, USA), Thermo Fisher (Thermo Fisher Scientific, Waltham, MA, USA), Pacific Biosciences (PacBio) (PacBio, CA, USA) and Oxford Nanopore Technologies (ONT) (ONT, Oxford, UK)[271].

Phenotypic assay

Phenotyping in vitro susceptibility assays that measure ARV susceptibility in cell culture. Susceptibility is usually reported as the ARV concentration that inhibits HIV-1 replication by 50% (IC_{50}). The IC_{50} of a patient's virus is compared to that of a drug-susceptible reference strain and expressed as a ratio, referred to as fold change, of the IC_{50} of the patient's virus relative to the reference control. Nearly all susceptibility assays use recombinant viruses created by inserting PCR-amplified patient virus gene segments (PR/RT, IN and env) into the backbone of a wild-type laboratory clone[217, 279]. It is, however, labour intensive and time-consuming, taking at least six weeks from initial specimen collection to data generation. It is thus reserved for drug development, drug resistance research, or complex clinical cases[217].

Phylogenetic analysis

Since the days of Charles Darwin, evolutionary relationships between taxa or species have been inferred from phenotypic differences or similarities. Initially, these trees were drawn by hand, with the branching order between taxa determined by observed phenotypic differences or similarities. Two critical technological advances in the late 1950s and early 1960s provided a new impetus to modern phylogenetics. These were advances in molecular biology (the composition of nucleic and amino acid sequences) and the development of large, centralized computers capable of handling complex computations. Scientists set out to develop algorithmic means of analyzing genetic data to infer evolutionary relationships now that genetic information and computational power are readily available[280].

The use of phylogenetic analyses in many fields of study has increased over the last decade as the cost of sequencing has decreased, and access to sequence data from online repositories like GenBank and GISAID has increased. Phylogenetics has become more useful due to the recent advances wealth of genetic data generated by modern sequencing technology and Bayesian phylogenetics methods[281]. The recent success of Bayesian phylogenetics has relied heavily on the effective integration of multiple data sources, including spatiotemporal traits and sequence phylogenies. As a result of this integration, epidemiologists have gained new insights into epidemic dynamics and the determination of numerous parameters that cannot be estimated using traditional tools[281].

Advances in molecular biology, such as DNA sequencing, computer technology, and evolutionary biology have transformed the field of HIV research. Throughout this PhD thesis work, a variety of phylogenetic and phylodynamic methods were used, so it is necessary to briefly review some of the fundamental concepts of phylogenetics and phylodynamic analysis.

Phylogenetic tree

The phylogenetic approach has been used to study the evolutionary relationship of different organisms. A phylogenetic tree, also known as a phylogeny, is a tree-like diagram used to show the evolutionary relationship among various groups of organisms based on genetic information. A phylogenetic tree consists of nodes that are linked together by branches. Nodes can be either internal or external. The

phylogenetic terminal nodes (tips of the tree or leaves) represent sequences from the genetic entities under investigation. The internal nodes represent the common ancestors of the sampled tips, while the root represents the most recent common ancestor (MRCA) of all taxa in the tree. A horizontal branch length indicates the genetic relatedness between the different ancestors and their descendants in the tree, often measured in base substitutions per site. The purpose of the vertical branches in a tree is only to aid the visual interpretation of the phylogeny and to enhance the readability of the tree, and it therefore do not carry any information on evolutionary distance between the analysed sequences (**Fig. 10**).

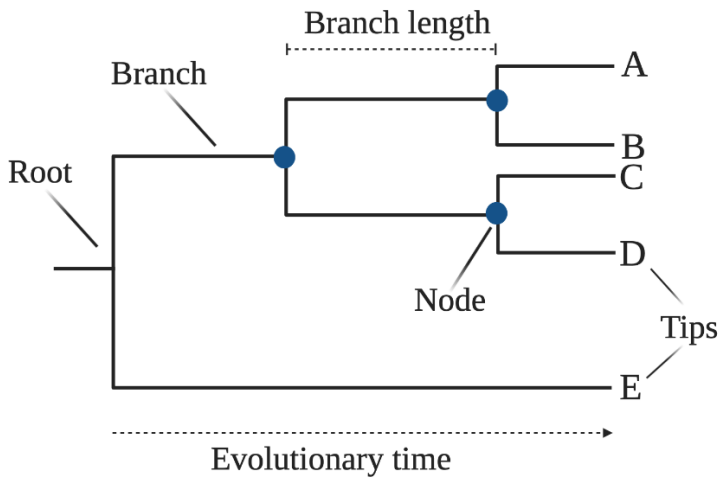


Figure 10. Diagrammatic representation of a phylogenetic tree.

Tips (terminal nodes) represent actual sample under investigation. The internal nodes (blue circles) represent hypothetical ancestors for the tips. The root is the common ancestor of all the taxa in the tree. The horizontal lines are branches and represent evolutionary changes measured in a unit of time or genetic divergence.

A phylogenetic tree can be rooted or unrooted. Unrooted trees do not have an evolutionary direction and only specify the relationship among taxa, whereas rooted trees contain information regarding ancestor-descendant relationships of the sequences. A rooted tree can be reconstructed by adding a distantly related taxon or outgroup in the dataset. If a suitable outgroup is unavailable, the middle point of the most extended branch on the tree can be used as the tree's root.

Phylogenetic methods

Phylogenetic tree-building methods are generally classified as distance methods or character-based methods (based on how the sequences are compared to infer the tree

topology, e.g., a matrix of pairwise genetic distances or discrete character states)[110, 282]. Several distance-based methods for inferring phylogenies have been developed, including the Fitch-Margoliash method[283], the unweighted group method with arithmetic methods (UPGMA)[284], Minimum Evolution method (ME)[285], and Neighbour Joining method (NJ)[286]. Maximum likelihood (ML)[287], maximum parsimony (MP)[288], and Bayesian methods[289] are common examples of character-based methods.

Distance-based methods

Distance-based methods of phylogenetic tree reconstruction, also known as algorithmic methods, involve converting the aligned sequences into a distance matrix (representing an estimate of the evolutionary distance between sequences), which is then used for tree reconstruction. Each method employs a different algorithm or criteria to efficiently infer trees. NJ applies a clustering algorithm to a distance matrix to generate a fully resolved phylogeny tree[286]. Distance-based methods are much faster compared with character-based methods, especially when dealing with large sequence datasets.

The disadvantage of using distance-based methods is that it does not provide information on the cause of the variations between the sequences, such as the specific position at which the bases differ and whether such differences are due to transitions or transversions[290]. Another issue is that the branch lengths and topologies inferred by distance-based methods are point estimates, making it impossible to evaluate alternative hypotheses for the dataset under consideration.

Character-based methods

Maximum parsimony method

The MP of tree construction is the most simplistic character-based method of tree inference. MP is based on the assumption, that the simplest explanation that explains the greatest number of observations is preferred over more complex explanations[291]. MP selects the tree which requires the least number of nucleotide substitutions to reflect the divergence amongst sequences[292].

The first step in parsimony inference is to identify all informative sites, which are sites in sequence alignment that allow for distinction between different sequences, followed by calculating the least number of substitutions at each informative site. The sum of the number of changes across all informative sites, called the parsimony length of a tree, for each possible tree allows to identify the most parsimonious tree, i.e., the tree with the lowest number of nucleotide substitutions. The MP score is dependent on the minimum number of mutations that could possibly produce the

data; however, as more-divergent sequences are analysed, the degree of homoplasy increases due to multiple substitutions at the same site. The true evolutionary tree becomes less likely to be the one with the least number of changes and can be erroneously inferred to be too closely related, a phenomenon called long-branch attraction (LBA)[293]. MP does not use all of the sequence information, instead focusing on the informative sites of the alignment. MP is less appropriate if the mutation rates are highly variable or the sequences themselves are highly divergent.

Maximum likelihood method

ML is currently the most widely used method due to improved computational power and software implementations, as well as the development of increasingly realistic models of sequence evolution (see the nucleotide substitution models section below). The ML algorithm searches for the tree topology (branching pattern) with the highest probability of reflecting the relationship of the observed sequences (sequence alignment). The approach is based on complex statistical theory and makes use of the concept of likelihood. Likelihood (L=P (D|H)), is defined as the probability (P) of observing data (D) given a hypothesis (H). In phylogenetic reconstruction, D represents the sequence alignment of interest, and H represents a given phylogenetic tree. A maximum likelihood approach will compute the likelihood of all possible trees for the specified alignment and choose the one(s) with the highest, or maximum, likelihood as the final ML tree for that data set. Thus, the likelihood of a tree corresponds to the probability of that tree describing the sequence alignment patterns given a specific model of nucleotide substitution.

An exhaustive search of all individual trees is theoretically possible when the number of sequences to be compared remains small. However, the number of possible trees grows rapidly with the number of sequences in the data set; for example, there are four possible unrooted trees for three sequences, but there are 2.03×10^6 possible unrooted trees for ten sequences. As a result, when the number of candidate trees becomes too large, it becomes computationally impossible to evaluate each one, and an exhaustive search cannot be considered. Instead, a heuristic strategy will be used but gives no guarantee of finding the optimal tree. A heuristic search can be compared to a 'hill-climbing process', in which an initial tree is generated and then modified by rearrangement until the tree, or trees, with the most likely topology is achieved. In ML, tree space can be considered as a two-dimensional landscape of hills and valleys with heights reflecting tree likelihoods; the tree with the highest peak in this landscape is interpreted as having the highest likelihood. However, if the starting tree is some distance away from the true ML tree, there may be several intervening lower peaks in this landscape that are incorrectly determined as the ML tree; therefore, to reduce computation time, one popular approach to tree construction is to use a NJ tree as the starting topology for a ML tree search. Some perturbation strategies, such as subtree pruning and re-grafting (SPR), nearest neighbor interchange (NNI), and tree bisection and

reconnection (TBR), are widely used to accelerate the heuristic search of the ML tree[294]. The key benefit of ML is that it allows the users to control the assumption (select the model of molecular evolution) for the computation of the data. However, such a dependency on a model might also be a disadvantage, as a specific model of evolution can bias the search result, and ML trees may not be valid if the model is not carefully selected. Moreover, ML remains a computationally intensive method and producing a ML tree rapidly becomes a time-consuming process as the number of sequences increases. Because almost every possible tree is tested, finding the ML tree can be computationally intense with large sequence datasets.

Bayesian inference method

Although Bayesian phylogenetic methods were introduced in the 1990s, their strength and ease of use have made them a popular method in molecular epidemiology[288, 295, 296]. It is widely used for molecular dating[117, 297, 298], to infer demographic history[117, 138, 299], viral phylodynamics[300], phylogeographic analysis of virus spread in humans[301-303] and inference of phylogenetic relationships among species or populations[304-307].

Bayesian methods, like ML, involve an explicit model of sequence evolution and search for trees that correlate best with the sequence alignment under a given model of substitution. Nevertheless, maximum likelihood searches for the tree that, under a hypothesis (i.e. a tree topology), maximizes the probability of observing the data (i.e. the sequence alignment), whereas Bayesian reasoning works the problem differently and searches for the tree that maximizes the probability of observing that tree given the data and the model of substitution[295]. In other words, ML methods search the tree that maximizes the probability of the data given the tree P (Data | Tree), whereas Bayesian inference searches the tree that maximizes the probability of the tree given the data P (Tree | Data).

The Bayesian framework is built on Bayes's theorem, which states that the posterior probability distribution is given by the product of the prior distribution and the likelihood, divided by the probability of the data[308]. In phylogenetics, the theorem takes the following form:

$$\mathbf{P(\theta \mid D) = P(\theta)\, P(D \mid \theta) / P(D)}$$

where P (θ| D) is the probability distribution of the parameters given the data (posterior probability), P(θ) is the probability distribution of the parameters (prior probability), P (D | θ) is the probability of the data given the parameters (likelihood), and P(D) is the marginal probability of the data which requires multidimensional integration over all possible trees which is effectively indecipherable. Instead, the probability of the data is assumed to be constant, and a Markov chain Monte Carlo (MCMC) simulation is used to approximate the posterior probability by sampling a large number of samples from its target distribution[309, 310].

Metropolis-Hastings algorithm is the most popular method to simulate MCMC in Bayesian phylogenetics[309, 310]. To do this, the MCMC algorithm starts with a totally random tree and set of values for all associated model of parameter and calculate the posterior probability. It then proposes changes to one or more model parameters and calculates the posterior probability again. A posterior probability ratio is calculated between the proposed and initial states. If the ratio of the posterior probability of the current state is greater than the random number, the new state (tree) is accepted and becomes the current state; otherwise, it is rejected, and the chain remains in the current state. The process continues for a number of generations (typically millions of steps) until it converges to a stationary distribution, or the user determines that enough samples have been drawn. Newly accepted proposed states gradually form a chain of states during the process. The proportion of time that any phylogeny has been visited is used as a valid approximation for the highest posterior density (HPD) for that phylogeny[110, 311]. Maximum clade credibility tree, a tree with the highest posterior density from the tree distribution, will be selected (**Fig. 11**). The confidence in the final tree topology can be represented as a proportion by the frequency at which clades appear on trees in every posterior estimate. A value of 1 indicate high confidence in the topology. Due to the stochastic nature of the MCMC algorithms, the chain's early moments spent in the chain tend to fluctuate haphazardly to convergence on the posterior. To avoid skewing the final distribution, it is common to discard early posterior estimates. This is known as the burn-in period. Typically, the first 10% of a chain (10% burn-in) are removed. However, depending upon analysis, results should always be checked, and the appropriate burn-in period determined. To improve the efficiency of Bayesian inference and reduce the risk of being stranded in local maxima, performing multiple Markov chains needs to be performed in parallel.

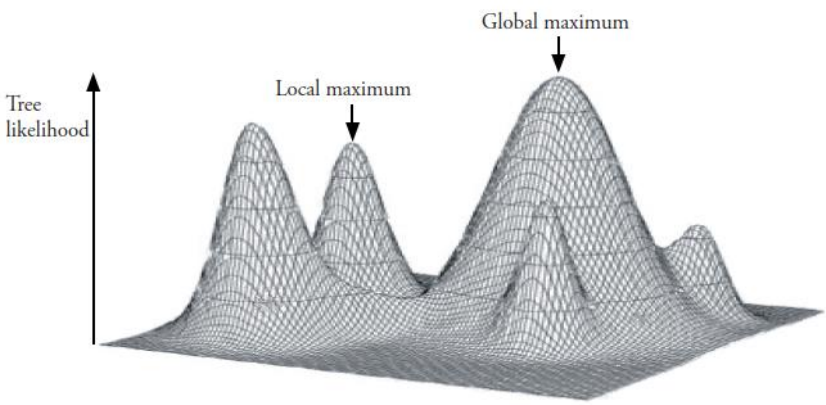


Figure 11. Phylogenetic tree landscape.
The figure depicts a possible tree landscape with several local minima and maxima and a global maximum where the most likely tree will be found. The figure was adopted with permission from Joakim Esbjörnsson[312].

Critical steps in phylogenetic analysis

Before attempting to infer the topology of a phylogenetic tree, several critical steps must be completed and considered. In the following section, I will discuss several of the most common concepts and steps to be considered in phylogenetic analysis.

Sampling consideration

Appropriate sampling is the key to accurate phylogeny and parameter estimation. Sampling considerations include the number of taxa or sequences and the number of nucleotides or characters per sequence. The first step in any phylogenetical analysis is to select genomic data of the pathogen that has genetic variability (phylogenetic signal) to reconstruct the epidemiological relationships of the population under study[313]. A sufficient increase in sequence sampling improves phylogenetic estimation significantly[308, 314-317]. Other critical steps include ensuring uniform spatial and temporal sampling, allowing enough time between consecutive sample collections to observe measurable evolution, and taking genomic recombination into account. Recombination often cause a significant overestimation of substitution rate heterogeneity and loss of the molecular clock[318], hence,it is critical to identify and eliminate putative recombinant sequences before performing phylogenetic analyses[319].

HIV phylogenetic analysis typically employs the PR/RT region, which has a large number of sequences and has been shown to contain sufficient information to study HIV transmission[320]. However, a significant disadvantage of using PR/RT is that they do not detect recombinant sequences that occurred outside of those regions. Recombination events can have a significant impact on phylogenetic trees, making estimation of population histories or event timings less reliable. The assumption of a strictly bifurcating genealogy (i.e., where one descendant has two ancestors only) is violated when recombination occurs because sequences have different phylogenetic histories in the various regions of their locus and evolve along with a set of correlated trees rather than a single tree. Whole-genome sequence analysis will provide the most accurate phylogenetic information and may help to depict the epidemic better, but its use is limited due to the high cost[321, 322]. Most phylogenetic analyses are designed to determine the evolutionary relationship between newly sequenced data and other sequences. To establish such an evolutionary relationship, reference sequences that are closely related to the newly sequences must be obtained and included in the analysis. The Basic Local Alignment Search tool, or BLAST, is the simplest way to find homologous sequences[323].

Sequence alignment

Sequence comparison is only valid when the sequences have a common ancestor (homologous). Sequence alignment is a technique for correctly positioning each nucleotide or amino acid position in relation to its homologous sites. Rows in a matrix are typically used to represent aligned sequences of nucleotides or amino acid residues. Sequence alignment allows researchers to identify and locate evolutionary changes, such as deletions or insertions, made by different lineages since their common ancestors[324].

A sequence alignment can identify three types of base differences: matches, mismatches, and gaps. A mismatch occurs when at least one substitution has occurred since the two sequences diverged, whereas a gap represents a deletion or insertion in one of the compared sequences (**Fig. 12**).

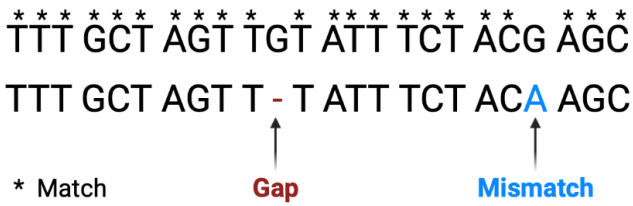


Figure 12. Sequence alignment.

When the same base is found at a specific position, the sequences are said to match; otherwise, they are said to be mismatched; a gap indicates that one of the compared sequences has undergone a deletion or insertion.

Sequence alignment is frequently overlooked in phylogenetic analyses, but misalignment can lead to inaccurate estimates of evolutionary divergence[325]. Manual alignment of sequences with low genetic divergence is possible; however, several sequence alignment programmes are freely available online. The most often used include, BioEdit[326], CLUSTAL X[327], MUSCLE[328] and MAFFT[329]. Phylogenetic analysis is highly dependent on the sequence alignment and a manual visual check is strongly advised to ensure that the nucleotides or amino acids are properly aligned[314].

Nucleotide substitution models

Substitutions or evolutionary models play an important role in the analysis of molecular sequence data[330]. Counting the number of differences between DNA or protein sequences for a specific gene or genes is the simplest way to determine divergence. However, because multiple substitutions or hits occur at the same site, the extent of observed genetic dissimilarity between two sequences is curvilinear rather than linear with time. The likelihood of the same site undergoing more than

one change increases as the number of substitutions increase and failing to account for these multiple hits may result in an underestimation of the true evolutionary distance between the two sequences. Substitution models are statistical models that are used to estimate the true number of substitutions that have occurred in the past given the pattern of sequence variation observed in the alignment[330]. Substitution models reduce the complexity of the biological mutation process to simpler patterns that can be characterized and predicted with only a few parameters.

To date there are several substitution models each with its own assumptions. The relative complexity of these models is measured by the amount of biological, biochemical, and evolutionary information that they integrate. The substitution models developed to date include two types of parameters: base frequency and base exchangeability. Base frequency considers the relative frequency of the four bases (A, G, C, and T) across all sequence sites. It is believed that allowing for some bases to emerge more frequently than others when substitutions take place reflects the compositional restrictions nucleic acids are subject to, such as G-C content or secondary structures. For instance, the HIV-1 genome has a strong preference for G to A transitions[331, 332]. Base exchangeability describes the tendency of bases to be substituted for one another. Transitions (purine to purine, or pyrimidines to pyrimidines), for example, have been shown to occur at a higher rate than transversions (substitutions from purines to pyrimidines, or vice versa). Base exchangeability reflects the biochemical similarity that bases share and its effect on mutational bias.

The first and simplest substitution model is the Jukes and Cantor model (JC69) which assumes that all nucleotides occur in equal frequency (25%) and the nucleotide rates of exchange are equally likely[333]. However, it is well known that certain bases occurs more likely than others, and that transition mutations are more common than transversion mutations[334, 335]. Kimura's two-parameter model (K2P) assume, the base frequency is equal along site, but rates of change differ between transitions and transversions[334]. While Felsenstein (F81) extended the JC69 model to include different nucleotide frequencies[336]. A number of models were later developed by adding an extension the original models, Hasegawa-Kishino-Yano (HKY85), which assume that nucleotides occur at different frequencies and that transitions and transversions occur at different rates[337]. The general time-reversible (GTR) model, also known as the general reversible (REV) model, allows all six pairs of substitutions to occur at a different rate, and all substitutions are reversible[338-340] (**Fig. 13**).

In phylogenetic analysis, in addition to using a specific model of nucleotide substitution, one must account for variation in substitution rates across sites. All the above substitution models (JC69, F81, HKY85, and GTR) assume that different sites in a sequence evolve in the same way and at the same rate. However, such assumptions may be incorrect because some regions of a coding sequence may be more conserved due to their importance function or in determining protein

secondary structure. In the case of HIV, for example, the *env* gene is more variable than the *gag* and *pol* genes, and the env V1-V2 and V3 regions (variable regions) contain more substitutions than other more conserved regions (e.g. *env* C2 and C3)[341]. Such rate heterogeneity can be accounted for by assuming that the rate for every site is a random variable that can be computed from a statistical distribution. Such nucleotide substitution rate heterogeneity across sequences is typically described by the gamma distribution in evolution. Models with a gamma distribution of rate heterogeneity are usually denoted by the suffix "+ Γ"[338].

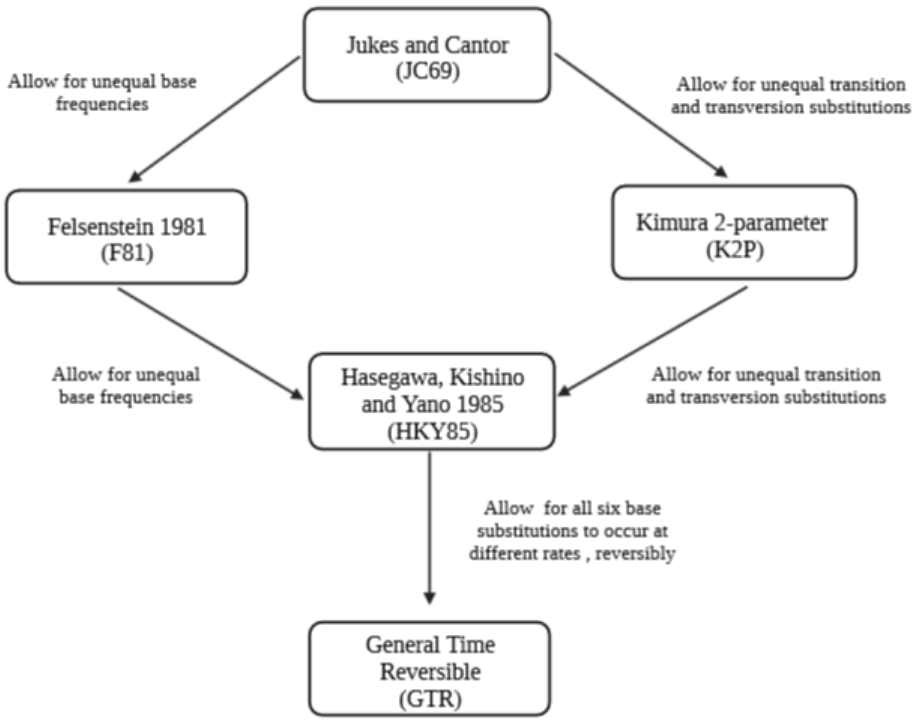


Figure 13. **A comparison of nucleotide substitution models**.
Each model is essentially a development of the JC69 model.The F81 model allows for variable base frequencies while maintaining a constant transition to transversion rate, whereas the K2P model allows for different transition to transversion rates while maintaining a constant base frequency. The HKY85 model is a hybrid of the F81 and K2P models in that it allows for variable base frequencies and a variable transition to transversion rate. The GTR model is an extension of the HKY85 model in that it allows all six pairs of substitutions to occur at different rates and that all substitutions are reversible. The The diagram was adapted from Lam *et al*., 2010, with modification for clarity[110].

Model selection

Although an array of nucleotide substitution models of increasing complexity has been developed and described, selecting an appropriate model of substitution is one of the major problems in phylogenetic reconstruction. Substitution model misspecification might lead to false phylogenetic inferences[330, 342, 343]. Therefore, the selection of the best-fit substitution model became an essential stage in the pipeline of phylogenetic inference[344, 345], which also increased the popularity of substitution models in phylogenetics.

One possible solution to model selection for constructing phylogenies could be the arbitrary use of complex, parameter-rich models[55, 346]. However, with the use of complex models, the analyses become computationally difficult and also increases the error with which each parameter is estimated[346]. The best-fit model of substitution to the dataset can be chosen through rigorous statistical testing[346]. There are mainly two types of statistical tests for model of substitution: tests intended to compare two different models using the likelihood ratio test statistics (LRT), and tests of the overall adequacy of a particular model. The LTR is a statistical evaluation of goodness-of-fit between two models of substitution. It yields a likelihood ratio statistic which corresponds to the ratio of the likelihood of the two models[346]. The LRT compared models must be hierarchically nested. This means that one model must be derived from the other or be a special case. The most complicated model can only differ from the simpler one by adding one or more parameters. There are several computer programs available that search for best model to fit the data and estimates the corresponding parameters[346]. For this PhD thesis work jModelTest was used.

Assessing topological confidence

Because the construction of a phylogenetic tree cannot guarantee the correct phylogeny, determining the robustness of the obtained topology is an important step in phylogenetic analysis[347]. Bootstrapping the reconstructed tree is one way of estimating the robustness of the reconstructed phylogeny. The bootstrap method involves repeating resampling with replacement from the original samples to create new subsets pseudo-alignments. These pseudo alignments are then subject to the same analysis as the original samples. Resampling with replacement means that some characters/data from the original samples can appear in the bootstrap sample multiple times, while others may not appear at all. The process of creating pseudo-alignments is repeated many times, often 1000 times and trees are then generated from this multiple pseudo-alignment. An estimate of the reliability of a branch from the original tree will be based on how many times it has been found in pseudo-trees. A bootstrap value, which is indicated on the branch, indicates the proportion of time that branch was found in all pseudo-trees. A bootstrap value of 100, for example, indicates that the branch associated to it was present in all pseudo-trees, and is thus

extremely robust. Although bootstrap values ≥ 70% have been proposed to indicate strong support for a cluster because bootstrap values are conservative measurements, the bootstrap approach has the disadvantage of being difficult to choose a reasonable significance cut-off[348].

The Shimodaira-Hasegawa-like approximate likelihood ratio test (aLRT-SH) is an alternative method to bootstrapping. The aLRT-SH method uses a likelihood ratio test to determine whether a particular branch is significantly longer than zero or not. In comparison to bootstrap, aLRT-SH allows for quick computation of branch support, and aLRT-SH values ≥ 0.90 are typically considered significant[135]. The aLRT-SH algorithm is used by the majority of fast ML tree estimation programmes, including PHYML and IQ-tree[349-351].

Phylodynamics

One of the most active and successful areas of research in modern evolutionary biology is the investigation of how and where viruses enter and spread throughout human populations. Many viruses, particularly those with RNA genomes, evolve quickly, allowing genetic variation to accumulate within an epidemiological time frame and thereby enhancing the utility of viral genomes obtained over time as an asset for epidemic analyses[300, 352]. To put it another way, because RNA viruses evolve so rapidly, the epidemiological and ecological processes that shape their genetic diversity happen on the same timescale as mutations are fixed in the viral populations[353]. As a result, genetic variation in RNA viruses, from a given population, can be used to infer viral evolution patterns, processes, and dynamics, offering a unique molecular perspective on their ancestry and change mechanisms[354, 355]. Indeed, the rapidity with which RNA viruses evolve allows for the resolution of phylogenetic relationships between isolates sampled only days apart, providing information on forensically important questions[356]. Phylodynamic approaches integrate phylogenetic reconstruction into a population genetics framework to estimate population size and epidemiological parameters (e.g., the reproductive ratio, which represents the average number of onward transmissions from each case). Phylodynamics is an interdisciplinary field that seeks to understand the evolution and transmission of infectious diseases. Population genetics (mutation rates, recombination, and selection), ecology (spatial and temporal distributions of the virus and host), and phylogenetics are examples of such disciplines[357].

The combination of larger genetic data sets and computing power allows for the development of more sophisticated and efficient phylogenetic and phylodynamic methods. This has greatly improved our understanding of the emergence, spread, and origins of viral infections in human populations[117, 138, 298, 358]. Coalescent and birth-death models (see below paragraphs) are widely used within the phylodynamic

framework to estimate key population and epidemiological parameters from pathogen sequence data.

The coalescent model

Coalescent inferences are based on the idea that present-day populations have a genetic signature that is encoded in their genome data that can be used to reconstruct their historical population dynamics[354, 355]. According to the coalescent theory, in the absence of selection, sampled lineages are expected to randomly choose their parent as we travel back in time. When two descendants choose the same parent, it is said that their lineages have coalesced (**Fig. 14**). The rate at which lineages coalesce is determined by the number of lineages coalescing (the more lineages, the faster the rate) and the population size (the more parents to choose from, the slower the rate). For example, by randomly sampling two viruses from a small population of viruses, the probability that they coalesce on a common ancestor sometime in the recent past is high, i.e., the short coalescent time correlates with the small population size from which they were drawn. In contrast, given a large virus population, the coalescent time of any two randomly sampled viruses will likely be longer. The coalescent approach aims to describe a stochastic process that allows for the inference of a population's historical states from the genealogy of individuals randomly sampled from it. Looking backward in time, the number of ancestral sequences decreases as the lineages coalesce, until all lineages coalesce into the most-recent-common ancestor of all the samples.

The temporal distribution of the internal nodes or coalescent events in the phylogeny, can be used to parameterise tree structures and draw inferences about the effective population size over time[299]. The effective population size is directly proportional to the number of individuals who contribute offspring in subsequent generations and is almost always less than the actual population size. Coalescent methods employ demographic models that describe the changes in effective population size over time. This assumption of a demographic model is crucial for the coalescent model, just as phylogenetics inference is closely tied to molecular evolution modelling. Previously, deterministic demographic models (parametric models) such as constant size, exponential growth, logistic growth, expansion growth model[299] were used to describe the population history. However, most population histories are far more complex and cannot be adequately described by simple parametric means. This resulted in the development of non-parametric methods for deducing demographic histories from sequence data. The coalescent Bayesian Skyline[299] and Skygrid[359] models, are popular non-parametric models that do not impose fixed assumptions of a growing or declining population. These models have been particularly useful in estimating the real incidence of undetected infections in populations[360] and approximating epidemic curves when the past population size is unknown.

The coalescent framework has been extensively used in recent years to study the epidemiology and population dynamics of viral epidemics. These include: the inter-host evolutionary dynamics and longitudinally sampled HIV-1 virus *env* genes[361], the estimation and pandemic growth H1N1 in the USA[362], sexual transmission and the phylodynamics of HIV-1 amongst the MSM population in the UK[363]in Kenya[302, 358], and the HCV epidemic reconstruction in Egypt[299]. The classical coalescent model is based on an idealised Wright-Fisher population, with no recombination, selective pressure, overlapping generations, significant immigration, or emigration[364, 365]. It has now been updated to include more parameters and tree priors for varying population growth[366] and continuous (overlapping) generations[367, 368].

One of the coalescent model's limitations is that it cannot tell whether changes in the effective population size (i.e., the number of infected people) are the result of an increase in incidence or a decrease in prevalence. Second it approximates the population dynamics, by assuming that a small random sample was drawn from a large background population. However, during epidemics or large cohort studies, the percentage of infections that are sampled can be very high[369, 370].

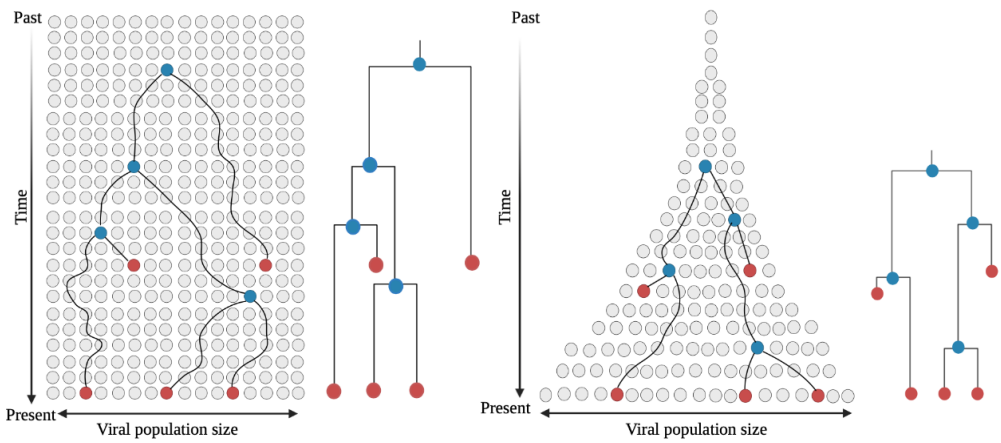


Figure 14. Illustration of the coalescent model used to estimate viral population size over time.

Panel (A) shows the scenario of phylogenetic relationships of individuals drawn from a constant-sized viral population growth and panel B represent from an exponentially growing viral population. The red dots represent individuals sampled at different time points (heterochronic), the blue dots their hypothetical common ancestor of sampled viruses (known as coalescent event), and the grey dots individuals from the non-sampled viral population. Moving back in time from the present, we follow the number of lineages in the genealogy in each generation. This value decreases when two lineages share a common ancestor (a coalescence event). The probability of a coalescence event occurring at a given time is inversely proportional to the population size at that time. This relationship is reflected in the size of the branches of the phylogenetic tree and can be used to estimate the demographic history of the population. The figure was simplified based on Drummond *et al*. (2003)[371].

Birth-death model

The birth-death model, in contrast to the coalescent model, is a forward-in-time model based on the birth-death process. The birth-death process starts with a single infectious individual at the time of origin, $t = 0$. The birth (λ) refers to the rate at which an infected individual will infect another individual (transmission rate), while the death (μ) is the rate at which an infected individual become noninfectious. Up on becoming noninfectious, an individual is sampled with probability of δ and does not remain infectious[370]. The noninfectious state of sampled individuals can be caused by several factors, such as successful treatment, behavior change or death. The birth event (transmission) corresponds to a bifurcation in the tree, while a death event or being sampled (becoming noninfectious) is reflected by the truncation of a lineage (**Fig. 15**). A sampling rate is used to capture the sampling of infected individuals from this transmission tree (or sampling probability). To obtain the reconstructed phylogenetic tree, all edges with no sampled descendants are removed from the transmission tree. As a result, the transmission tree describes the sequence of transmission among infected individuals, whereas the phylogenetic tree describes the evolutionary relationships between the viral samples obtained from each individual. In this manner, the birth-death model generates a transmission tree (**Fig. 15**), that describe the epidemiological process, starting from a single individual at time at time t, and as a function of the sampling (δ), birth (λ), and the death (μ) rates.

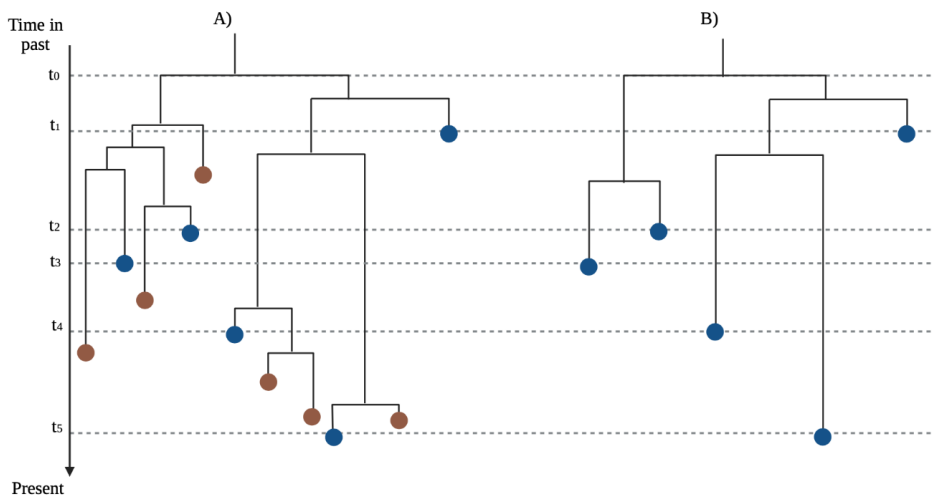


Figure 15. Schematic tree representing the birth-death process.
(A): A transmission tree produced by the birth-death process. The horizontal dashed lines represent sampling events (at time t1-t5). Sampled tips are marked in blue, and becoming non-infectious(death), (lineage that stop growing) is marked as gray color. An individual is considered non-infectious after being sampled. (A): represents a transmission tree, in which all lineages with no sample descendants are pruned. (B): represent phylogenetic tree obtained by suppressing all unsampled tips from the transmission tree. Adapted from Stadler *et al.* 2013, with modification for clarity[370].

Birth-death model can be used to determine the effective reproductive number ($R = \delta/\lambda$), which is a key epidemiological parameter. The effective reproductive number (R_e) is defined as the number of expected secondary infections produced by an infected individual at any given time during the epidemic while the basic effective reproductive number (R_0) assumes in a completely susceptible population. Both refer to the same at the start of an epidemic[372]. The R_e parameter is used to describe temporal variation in the transmission potential of infectious diseases, $R_e > 1$ indicates that the epidemic is growing, $R_e < 1$ indicates that the epidemic is declining, while $R_e = 1$ indicates that the epidemic is stabilizing[370]. The birth-death model can be used to examine different processes, from epidemiological to macroevolutionary procedures. Depending on the application they have different interpretation, they correspond to speciation and extinction rates on a species level and on epidemiological context, they correspond to transmission and becoming un-infectious rates.

Application of HIV-1 molecular epidemiology to public health

Classification and dating the origin of HIV-1

Phylogenetic analyses have shed light on the historical roots of the HIV-1 and HIV-2 pandemics[45, 55, 373], as well as the connections between HIV and other simian lentiviruses and the categorization of HIV diversity within HIV-1[38, 53]. Through the utilization of phylogenetic methodologies, it has been possible to recognize and characterize the cross-species transmission. Phylogenetic analysis has greatly improved our understanding of the HIV-1 group M's early spread. Although the first human case of HIV-1 was reported in the United States in 1981, phylogenetic analysis revealed that HIV has been circulating in humans since the early twentieth century, with relatively slow growth until around 1960[42]. However, after 1960, HIV entered a secondary phase of faster exponential growth and geographical expansion. The use of phylogenetic and phylogeographic approaches in tandem was critical for tracking the founder events that resulted in the geographic spread of pandemic HIV-1 trains. According to the findings, most HIV-1 pandemic lineages spread from the Congo basin to neighbouring regions in southern, eastern, and western Africa before spreading outside of Africa[42].

Global and regional patterns of HIV-1 spread

The global dissemination of pandemic subtypes and CRFs resulted in local epidemics with varying sizes and geographic distributions. Since temporal changes in the spatial dispersion and population size of HIV-infected individuals leave an

imprint on HIV genetic diversity and phylogenetic patterns, model-based phylodynamic inference methods were used to track the phylogeographic and demographic history of the HIV-1 epidemic within a specified area[117, 374]. Phylodynamics could offer crucial epidemiological insights (epidemic drivers or risky behaviours) regarding HIV-1 epidemics affecting a vast geographical area[135, 303, 358]. This method was used, for instance, to determine the spatiotemporal dynamics of the HIV-1 subtype A variant that dominates the epidemic in the former Soviet Union (A_{FSU})[375] and the non-pandemic subtype B variants that are prevalent in the Caribbean region (B_{CAR})[376]. Phylodynamics studies were also used to determine the dynamics of HIV lineage dissemination at the country level. For example, HIV-1 has been circulating in Brazil for 20-30 years prior to their detection by the public health surveillance[377]. Phylodynamics analyses also aided in elucidating the origin and population dynamics of HIV-1 lineages spreading within small communities by shedding light on their origin. An outbreak of HIV in children from a Libyan hospital in 1998, for instance, was suspected to have originated from the malicious intervention of foreign medical personnel[378]. However phylogenetic analysis revealed that the HIV-1(CRF02_AG) outbreak affecting Libyan children originated from a single viral introduction from West Africa before March 1998 and that many of the HIV infections had already occurred before the arrival of foreign medical personnel, ruling out their involvement in the initial transmissions[378].

Investigation of an HIV-1 transmission network

Due to the absence of viral RT proofreading activity and rapid replication, mutations accumulate in HIV genomes on an epidemiological timescale. This not only indicates that we can track the spread of HIV across vast territories over the course of decades but also that genetic diversity accumulates quickly enough to reconstruct the viral transmission network, which describes the history of infections at the level of individual cases. The fundamental assumption is that closely related viruses in a phylogenetic tree indicate that the hosts are connected through a common source, a direct or short chain of transmissions. In recent years, the use of phylogenetic analysis in conjunction with the vast availability of HIV sequences has become an increasingly important area of study for reconstructing HIV transmission networks uncovering who and where HIV infection is spreading and estimating the rate of HIV transmission[135, 358]. In this regard, the majority of studies that attempt to infer potential transmission chain or network among HIV-infected patients have relied on HIV-1 partial *pol* gene generated from the routine HIVDR genotyping for clinical monitoring. In countries with large data sets, phylogenetic analysis can also be used to infer source population for HIV infection, with the combination of sociological, demographic, and epidemiological data. In South Africa, a large-scale study was done in KwaZulu-Natal to identify the key mode of sexual networks driving local HIV transmission. The study found that older men are the source for HIV infection among young women[379]. According to a phylogeographic study conducted in

Uganda, viral strains from the general population were found to migrate to HIV-hyperendemic, fishing communities, suggesting that fishing communities were a reservoir for viral strains from the general population, not their source[380]. Germany using phylogeographic analyses found that Cologne-Bonn was the HIV transmission hotspots. In this study, when comparing individuals with and without links, the authors demonstrated that individuals in clusters tended to live closer to one another[381]. Such studies highlight the importance of phylogenetic analyses in determining which groups are most vulnerable to HIV infection and where prevention is most likely to be effective.

Monitor the dynamics of local HIV-1 transmission

One of the most important applications of molecular epidemiology analyses on HIV sequences in the population is to identify and characterize transmission clusters. Phylogenetic analysis has long been used to determine HIV linkage and infer a possible network among populations[135]. A transmission cluster is a group of people who belong to the same transmission chain and have similar viral sequences derived from a common ancestor as a result of multiple infections in a short period of time. In phylogenetic tree transmission cluster corresponds to a specific branch (or monophyletic clade) with high support (70–99 %) and sufficiently small genetic distances (0.5-4.5)[322, 348, 382, 383] . However, the genetic distance used for the cluster definition depend on the objective of the study. If the primary goal is to identify all possible transmissions associated with a given case, a higher genetic distance threshold can be used. Using a genetic threshold of 0.5% for HIV-1 phylogenies corresponds to approximately 2–3 years of viral evolution separating any putative transmission cluster's sequences while a 1.5% threshold would detect networks with a maximum of 7–8 years of viral evolution[322]. In recent years, a new simplified genetic distance (HIV-TRACE[384] and Cluster Picker[382]) and phylogeny-based approach for inferring potential transmission networks in real time for proper interventions has been developed, and it is increasingly being used in large sequence datasets in the United States, China, and Europe[322]. Transmission network analysis, in conjunction with other epidemiological studies, can be used to identify high risk groups (diagnosed/ undiagnosed HIV-1infected and uninfected individuals) as well as clinical and social behavioral factors associated with HIV transmission, which can then be used to implement targeted preventive interventions[322]. The approach demonstrated a great potential to reduce HIV transmission among MSM and high-risk groups[385]. This transmission network-based intervention strategy is currently being implemented as a key tool for HIV epidemic control in the United States, Canada, and China[322].

Quantifying critical epidemiological parameter

Surveillance systems have been the backbone of public health efforts to track infection cases and their distribution by time, person, and location[386]. However,

traditional epidemiologic methods have inherent limitations. Despite the simplicity of metrics, it may be challenging to calculate key epidemiological variables such as R_e, reproduction rate, prevalence, and incidence in certain circumstances. People with HIV are frequently diagnosed late, the timing of the transmission event cannot be estimated, especially in the absence of testing histories, and identifying index cases is nearly impossible. As a result, it is challenging to estimate incidence and reproduction rates. Biases and other sources of error in epidemiologic research and public health practice can also distort associations and lead to erroneous or misleading conclusions[386].

Phylogenetic analysis has been used for many years to infer a potential transmission chain or network among HIV-infected patients. Recent advances in phylodynamics allow for the quantification of transmission dynamics as well as the estimation of key epidemiological variables such as R_e, reproduction rate, prevalence, and incidence[386, 387]. Stadlers *et al*. have used the birth-death skyline plot to estimate the R_e for the subtype B HIV epidemic in UK[370]. Similarly, Novitsky *et al*. has used the birth-death-model to characterize the HIV epidemic in Botswana[388]. The birth-death model has been also used to characterize the transmission dynamics between and within subpopulation (transmission risk group) and understand who the epidemic drivers are[389].

Assessing the impact of an intervention

An early ART prevention trial's effectiveness in preventing HIV-1 transmission among serodiscordant couples has been evaluated using phylogenetic analysis. In this study, linkages between HIV-infected individuals and their seronegative partners who participated in the early initiation of an ART prevention trial were investigated using phylogenetic analysis. The majority of new infections (76%) were found to be linked to their partners, while 18% of seroconversions were unlinked[390]. In a multi-center study that looked at the risk of HIV transmission among people on ART with viral loads less than 200 copies/ml who engaged in condomless sex, phylogenetic analysis revealed no link between newly infected people and their seropositive partners, implying that the infection came from another HIV-infected person. The application of phylogenetic techniques in ART prevention trials can aid in the estimation of linkages between index-partner pairs and, consequently, the evaluation of the efficacy of various HIV prevention interventions[391]. Phylogenetic is gold standard for estimating linkages between HIV-infected individuals. It has been utilized in various prevention trials, and it can also be utilized to evaluate the efficacy of prevention strategies such as PrEP and treatment as prevention TaSP.

Out-break investigation

One goal of genomic epidemiology is to derive epidemiological and emergence dynamics from virus genome sequences obtained over short epidemic timescales[392]. Phylodynamic analysis have been used to detect and characterize an epidemic potential of an outbreak[393]. For instance, the molecular epidemiology of the people with injecting drug use (PWID) outbreak in Athens identified early transmission networking and its temporal changes. Additionally, when combined with traditional epidemiology, the molecular epidemiology revealed factors (immigration, homelessness, and unemployment) that could potentially be associated with the outbreak. These findings helped to design the intervention program, ARISTOTLE programme (seek-test-treat and retain intervention programme) to improve HIV testing, referral, and treatment in Athens[394].

Aims of this doctoral dissertation

The general objective of this thesis work was to investigate the HIV-1 epidemic in Ethiopia, with specific focus on HIV-1 genetic diversity, transmission dynamics and antiretroviral drug resistance

Specific objectives:

- To characterize the molecular epidemiology of HIV-1 in Ethiopia with specific focus on transmission dynamics and evolutionary history of HIV-1 subtype C in Ethiopia using state- of-art phylogenetic and phylodynamic approach.

- To assess the prevalence of transmitted HIV drug resistance (TDR) among newly HIV-1 infected young adults using the WHO threshold surveillance approach and to describe the molecular epidemiology of HIV-1 in terms of genetic diversity, transmission clusters and drug resistance mutation (DRMs) transmissions within clusters in Gondar, Ethiopia.

- To characterize the pattern of HIV drug resistance mutation (HIVDRMs) during the initial year of antiretroviral treatment in HIV-1 positive adults receiving care at Ethiopian health centers and investigate the impact of tuberculosis on DRMs.

- To determine viral load nonsuppression (VLN) rates, HIV drug resistance (HIVDR) prevalence, and associated factors among female sex workers (FSWs) in Ethiopia.

- To investigate HIV-1 integrase (IN) genotypic profile to evaluate the prevalence of pretreatment drug resistance (PDR) mutations and natural occurring polymorphisms (NOPs) that might affect the genetic barrier to the emergence of resistance in integrase strand transfer inhibitor (INTSI)-naïve patients in Ethiopia infected with HIV-1 subtype C.

Materials and methods

Study populations and sequences dataset

Paper I

A combined data set of newly sequenced and publicly available Ethiopian HIV-1 *pol* sequences were used for this study. The newly HIV-1 *pol* sequences were generated from plasma samples collected from treatment naïve participants enrolled for the HIVDR survey in St. Paul General Specialized Hospital located in Addis Ababa, Ethiopia, in 2011. This was established according to the WHO-recommended survey methodology[395]. A total of 150 treatment naïve adults (age ≥18) eligible to start ART were consecutively enrolled and included in the study. Blood specimens and basic sociodemographic information were collected at baseline and every six months from each participant. From the 150 baseline specimens, 144 were successfully sequenced and used for this study. A comprehensive dataset of all publicly available Ethiopian HIV-1 subtype C *pol* sequences (matching pos. 2,243–3,326 relative of HXB2) were retrieved from the Los Alamos National Laboratory (LANL) HIV Sequence database (http://www.hiv.lanl.gov). The sequence quality control program of the LANL HIV sequence databank was used to remove sequences that had stop codons, frameshifts or were of poor quality. We retained one sequence per each patient and for patients with multiple sequences we chose the oldest sequence. A total of 1132 Ethiopian HIV-1 subtype C *pol* sequences which were collected from different parts of Ethiopia between 1986 and 2017, were downloaded from LANL and used in the study.

Paper II

The specimens used for this study were obtained from a cross-sectional study done among ART-naïve adults (age ≥ 18 years) in Gondar's two major Voluntary Counselling & Testing (VCT) clinics located 700 km north of Addis Ababa. The study was done according to the WHO-recommended threshold survey methodology[396] and was aimed to evaluate the prevalence of TDR in Gondar. Based on the WHO-recommended inclusion criterion, only individuals with a new HIV diagnosis, aged 18-25, no history of pregnancy, no ART exposure and no HIV related illness were included. Between August 2011 and December 2013, 84

participants were recruited for the study. Blood samples and basic sociodemographic data were collected and used for the study.

Paper III

For this study, participants were identified from a longitudinal cohort, which included 812 ART naïve adult participants (age ≥ 18 years), recruited from five public health centers (Mojo, Adama, Geda, Dhera, Wolenchiti), which provide ART services to residents of Adama town and the surrounding rural and sub-urban areas of Oromia Region, Ethiopia. These five health centers are located in the uptake area near the highway connecting Addis Ababa and Djibouti. This corridor is a high-risk area for HIV infection in Ethiopia[397]. All participants in the cohort were underwent intensified sputum-based bacteriological case-finding for active TB at inclusion. Blood samples were collected at inclusion and at subsequent follow-up visits, scheduled at months one, six, and twelve and then biannually for viral load and HIVDR testing. A total of 621 individuals with viral load results at six and/or twelve months after starting ART were used for this analysis.

Paper IV

This study was part of a larger cross-sectional study that was done in 2014 to evaluate HIV prevalence in FSWs in Ethiopia. Data collection was done using the respondent-driven sampling technique (RDS). The study was done in 11 cities, including Addis Ababa, Bahir Dar, Mekelle, Adama, Diredawa, Gambella, Shashemene, Kombolcha, Semera/Logia, Metema and Hawassa. Seven study sites, including Addis Ababa and Mekelle, Bahir Dar, Hawassa, Adama, Gambella, Dire Dawa, are regional capitals where many female sex workers reside. The four other sites (Metema, Shashemene, Logia and Kombolcha) are transport corridor cities to Addis-Djibouti, Addis-Moyale, Addis-Metema and Addis-Mekele which are also home to more FSWs (hotspot areas) (**Fig. 16**). All FSWs who lived in the eleven selected cities were considered to be sources population. Long distance truck drivers were also included in the survey as they were considered a high-risk group.

For this study, FSWs were defined as 'women who engage in sexual activity under the condition of receiving financial or other benefits. For the survey, the inclusion criteria were being able to receive money or other benefits for having sex with at least four people in the past 30 days, being over 15 years old, being properly recruited by a peer (presenting the coupon) and consenting to the interview and blood draw. A total of 4900 FSWs were included in the survey, of which 1172 were HIV-1 positive. This study used 1154 of the FSWs that tested HIV-1 positive and had viral load results.

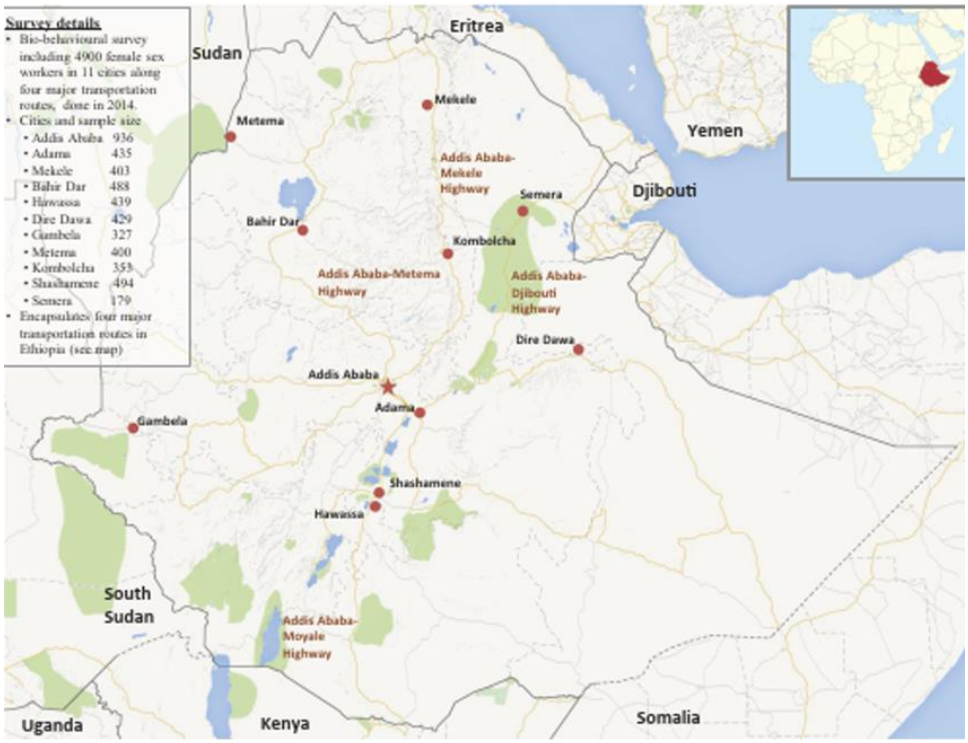


Figure 16. A map of Ethiopian cities and towns that were included in a 2014 study of HIV drug resistance among female sex workers.

Details of the study are shown in the box. This figure was modified from Google Maps (https://www.google.com/maps/place/Ethiopia).

Paper V

For this study we used samples collected from HIV-1-infected individuals as part of a nationwide HIVDR survey in Ethiopia. Based on WHO recommendations, a cross-sectional nationwide survey was undertaken in 2017 among treatment-naive patients and patients on first- and second-line regimens in 40 selected health facilities from all over the country. For this study we used 460 IN sequences collected from INSTI-naive individuals as part of the survey.

Methodology in details

HIV-1 drug resistance genotyping

For **paper (I-IV)** a 1084-bp HIV-1 *pol* fragment, which contained amino acids 6-99 and 1-251 of PR and RT respectively, were amplified by using CDC in-house assay which was later commercialized as ABI HIV-1 genotyping kit (Thermo Fisher Scientific, Waltham, MA)[398]. A premixed Big Dye terminator sequencing reagent (Applied Biosystems, Foster City, CA) were used to sequence the PCR products. ABI 3500xl and ABI 3730 Genetic Analyzer were used for sequencing. Sequence assembly and editing were performed using Standalone RECall V 2.0 HIV-1 sequencing analysis tool[399]). In **paper V**, the integrase region was genotyped using an in-house developed and validated assay for IN[400]. Each sequence was manually proofread to ensure good quality. The primers used are listed in the table 2.

Table 2. Primers for the amplification and sequencing of the HIV-1 *pol* region.

Primers	Sequence 5′–3′	Position (HXB2)	Description of primers
KVL068	AGGAGCAGAAACTTWCTATGTAGATGG	3854–3880	RT-PCR
KVL069	TTCTTCCTGCCATAGGARATGCCTAAG	5955–5981	RT-PCR
KVL070	TTCRGGATYAGAAGTAAAYATAGTAACAG	4013–4042	Nested
KVL084	TCCTGTATGCARACCCCAATATG	5243–5266	Nested/ Sequencing
KVL076	GCACAYAAAGGRATTGGAGGAAATGAAC	4161–4188	Sequencing
KVL082	GGVATTCCCTACAATCCCCAAAG	4647–4669	Sequencing
KVL083	GAATACTGCCATTTGTACTGCTG	4750–4772	Sequencing
PrtM-F1	TGA ARG AIT GYA CTG ARA GRC AGG CTA AT	2057-2085	RT-PCR
RT-R1	ATC CCT GCA TAA ATC TGA CTT GC	3370-3348	RT-PCR
Prt-F2	CTT TAR CTT CCC TCA RAT CAC TCT	2243-2266	Nested/ Sequencing
RT-R2	CTT CTG TAT GTC ATT GAC AGT CC	3326-3304	Nested/ Sequencing
SeqF3	AGT CCT ATT GAR ACT GTR CCA G	2556-2577	Sequencing
SeqR3	TTT YTC TTC TGT CAA TGG CCA	2639-2619	Sequencing
SeqF4	CAG TAC TGG ATG TGG GRG AYG	2869-2889	Sequencing
SeqR4	TAC TAG GTA TGG TAA ATG CAG T	2952-2931	Sequencing

Public sequence dataset and sequences quality control

For all newly generated sequence dataset of HIV-1 *pol* sequences (**paper I, II** and **V**), we retrieved all publicly available homologous HIV-1 subtype C sequence from the Los Alamos National Laboratory (LANL) HIV Sequence database (http://www.hiv.lanl.gov). For **paper I** and **II**, sequences matching pos. 2,243–3,326 relative of HXB2 and for paper V, sequence matching pos. 4230–5093 relative of HXB2. All sequence details, including the year of collection and source country, were collected. All sequences generated for this PhD study including those obtained from LANL were checked for sequence quality using the WHO tool (https://sequenceqc-dev.bccfe.ca/who_qc) and the Quality Control program of the Los Alamos HIV sequence database (https://www.hiv.lanl.gov). Sequences with poor quality, stop codon and frameshift were removed. For patients with multiple sequences, we kept only one and chose the earliest sequence.

Multiple sequence alignments

Multiple sequence alignments were performed using CLUSTAL X[327], and MAFFT version 7[329]and were then visually inspected and manually edited using BioEdit V7.0.9.0[326] until a perfect codon alignment was obtained (**paper I, paper II** and **paper V).** We removed positions of mutations that cause or contribute to HIVDR from the alignment to avoid the effects of drug-induced convergent evolution (**paper I**).

Subtype and intra-subtype recombination analysis

We used the online tool, REGA v3.0[401], COMET[402], RDP ver.3.5.13[403] for initial exploratory HIV-1 subtyping and recombination detection. Potential intra-subtype recombinant were identified using jumping profile Hidden Markov Model (jpHMM)[404, 405]and Simplot ver. 3.5.1[406] as described below. Subtyping was further confirmed by ML phylogenetic tree analysis using references sequences from HIV-1 subtype (A-K) and recombinant virus downloaded from the Los Alamos Database. At least two subtypes C clades, (C-EA and C'-ET) and their recombinant forms are circulating in Ethiopia[407-409]. Further molecular analysis revealed that the Ethiopian C-EA strain was similar to those found in East African countries, while the C'-ET strain was more closely related to strains from countries in Southern Africa (C-SA). According to Delatorre's and Bello's studies, based on phylogenetic relationships, ten subtype C clades (termed C1-C10) were identified[410]. Clades C1 through C9 were mostly composed of sequences from southern Africa. The C10 clade was made up of sequences from East or Central Africa[411]. The C10 clade includes the Ethiopian C-EA strain.

The Ethiopian subclade C’-ET clade was a distinct subclade among the major Southern African Clades. Thomson and Fernandez Garcia classified the C’-ET as

belonging to the Southern African C9 Clade, but it is a distinct phylogenetic subcluster[411]. Based on prior studies, the major African subtype C strain can be broken down into three distinct clades: C-SA, C-ET, and C-EA (**Fig. 17**).

We first created a non-recombinant reference data set of the different subtype C clades to identify potential intra-subtype C mutations. The dataset was based on the publication from Thomson, Fernandez Garcia and Delatorre[410, 411]. The sequences, polymerase region corresponding 2248-3309 of HXB2 were retrieved from Los Alamos HIV sequence database (http://www.hiv.lanl.gov).

The data set was first screened for putative recombinant sequence using an iterative version the phi-test, RDP v.3.44[403, 412] before the phylogenetic analysis. ML phylogenetic tree using the GTR+I+G substitution model was constructed with Garli v2.0[413] . A aLRT–SH test implemented in PhyML v3.1[414] was used to obtain the branch support and A aLRT-SH value ≥ 0.9 was considered significant[414, 415]. The phylogenetic tree as shown in figure 17, clearly showed that the sequences were divided into distinct clades.

The clade sequences were used to construct a scoring matrix as described in the jpHMM documentation (http://jphmm.gobics.de/)[416]. For the other HIV-1 subtypes we used polymerase sequences derived from the HIV reference data set from Los Alamos HIV sequence database (http://www.hiv.lal.gov). To optimize and validate the settings of jpHMM parameter setting, three training sets were used: 1) An artificial data set that contained chimeras of polymerase sequencings derived from combination of clade specific sequences were assembled. The first 340-bp comprised of C-EA and followed by 340-bp from of C'-ET and the last 340-bp were from C-EA sequences.

Twenty different chimeric sequences, each with a different combination of clade fragments, were created and screened with jpHMM using different jump (j), beam-width (bw) parameters. These artificial chimeras were correctly identified as recombinant sequences by a bw=1e-10 and an j=9.5 e^{-03}; 2). We also used the same settings to screen a number previously identified intra-subtype C (C-EA/C'-ET) recombinants[409, 410]. In all cases, these sequences were identified as recombinant sequences; 3) We further screened the reference sequences to their own matrix with the addition of reference sequences using the above parameters and, in all cases, the correct subtype and clade identity was found. All sequences identified as putatively recombinant sequences were further analyzed using Simplot version 3.5.1 and phylogenetic analyses. For **paper I and paper II**, we screened all the dataset for putative intra-subtype recombinant as described above.

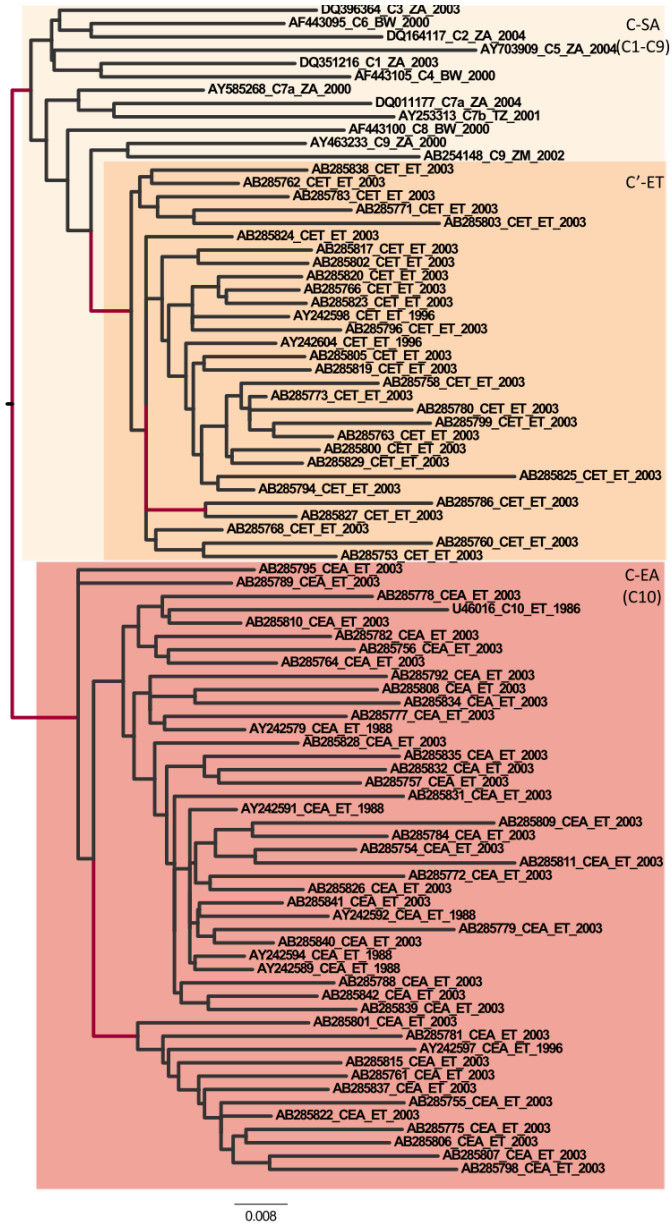


Figure 17. Maximum likelihood phylogenetic tree of the reference data set, derived from Garli v2.0. Significantly supported branches are highlighted in red (aLRT-SH≥0.9). Each supported group of sequences (C SA, C'-ET and C-EA clades) is highlighted in a coloured box.

Testing the molecular clock

We checked temporal signal or 'clocklikeness' of the dataset by a root-to-tip regression of genetic distance against sampling year using TempEst v1.5.3[417].

HIV-1 drug resistance analysis

To determine the prevalence of transmitted HIV drug resistance, the Stanford Genotypic Resistance calibrated population resistance tool, version 6.0 (https://hivdb.stanford.edu/cpr) was used by applying the WHO surveillance transmitted HIV drug resistance mutation (SDRM) list. Classification of PDR level was done based on WHO recommended threshold survey method (low (< 5%), moderate (5%-15%) and high (≥15%)[418] (**paper II** and **paper V**).
To determine ADR and mutation score, the current available version of Stanford HIVdb (https://hivdb.stanford.edu/hivdb/) was used (**paper III, IV** and **V**). For details, please refer to the individual papers.

Phylogenetic investigation

We used the online version of PhyML[414]for an initial ML phylogenetic tree construction with the GTR+I+Γ nucleotide substitution model (using estimated proportion of invariable sites and four gamma categories) and NNI plus SPR to estimate the tree topology. SPR branch-swapping algorithm was used for Heuristic tree. We used the aLRT-SH (approximate likelihood ratio test Shimodaira–Hasegawa-like) implemented in PhyML to determine the branch support and aLRT-SH value ≥0.9 was considered significant[414]. For transmission cluster analysis, ML phylogenetic trees were constructed using IQ-TREE[350] using GTR+I+Γ best fitting nucleotide substitution model as selected by jModelTest v2.1.7 (**paper I** and **paper II),** and the phylogenetic trees was visualized using FigTree v1.4.3[419]. In order to determine if phylogenetic clustering was associated with geography (**paper I)**, viral sequences were divided into six geographical regions (sequence collection locations). Bayesian Tip-association Significance testing (BaTS) program were used to determine the strength of the association between geographic location and phylogeny[420].

Phylogenetic investigation of transmission events

For **paper I** and **paper II**, we defined a transmission cluster in the ML phylogeny from root to tips with an aLRT SH-support of ≥ 0.9 that had a majority (≥ 80%) of sequences from (Gondar/Ethiopia) as Gondar/Ethiopia transmission cluster.

We further classified transmission clusters based on their size (number sequences/cluster), into dyads, with two sequences, medium size, (three to fourteen sequences), and large clusters (≥ fifteen sequences)[135, 421]. For **paper II**, we defined, putative drug resistance transmission cluster as a cluster sharing >33% of DRMs[422].

Evolutionary and phylodynamic analysis

Epidemiological and evolutionary parameter of the selected clusters were estimated by employing Bayesian Markov Chain Monte Carlo phylogenetic inference as implemented in BEAST 1.10.4 and 2.6.2. We used a Bayesian Skygrid coalescent tree prior to estimate changes in effective population size (N_e) through time. The epidemic growth rate (r, years^{-1}) was obtained by using a logistic growth coalescent tree prior that best fit to the demographic signal contained in the datasets[359, 423, 424]. To quantify the change in effective reproductive number (R_e) (epidemic growth through time), analyses was done for each cluster using the birth death skyline model (BDSKY) as implemented in BEAST2 v 2.6.2[387, 425, 426]. We used a lognormal distribution, LogNorm (0,1) prior for the become uninfectious rate (δ) in units per year with δ = 0.2 as mean distribution (i.e., the inverse of the time duration of being infectious in a unit of years). For R_e, we used LogNorm (0,1) with the upper bound of 10. We employed a different sampling probability (*rho*) prior for each year to account for the uneven number of sequences per year. The change in R_e was estimated for six equally spaced interval between the tMRCA and the most recent sampling year.

Analyses were performed using the GTR+I+Γ4 nucleotide substitution model and the temporal scale of evolutionary process were estimated using an uncorrelated relaxed molecular clock model with an underlying lognormal distribution with normal priors. In both cases, the analysis was done by using the GTR+I+Γ4 nucleotide substitution model. The age of most common ancestor, evolutionary rate (nucleotide substitutions per site per year, s/s/y) and other phylodynamics parameter was determined by using relaxed uncorrelated molecular clock model with an underlying lognormal distribution with normal priors. We ran three independent MCMC chains for each of the phylodynamic approaches until all associated parameters converged.

Convergence of each MCMC runs were inspected by using Tracer v1.7.5[427]. Effective Sample Sizes (ESS) > 200, after the first 10% burn-in were considered as convergence or good mixing. Results, from the independent multiple chains (log file and corresponding trees) were combined using LogCombiner, to ensure stationarity and good mixing[428]. The Maximum Clade Credibility (MCC) trees from the posterior distribution of trees was summarized using a Tree Annotator and were visualized in Fig Tree v1.4.3[419]. To plot the results of the BDSKY analysis, we used R's bdskytools package (https://github.com/laduplessis/bdskytools).

HIV-1 subtype C integrase polymorphism and conservation analysis

For this study, 453 HIV-1 subtype C integrase sequences were used. We first performed the multiple sequence alignment using MAFFT version 7[329] and then the alignment was edited manually until perfect alignment was obtained using BioEdit V7.0.9.0[326] . The nucleotide sequences, that has been aligned were translated into an amino acid sequence. Then, each amino acid along the 288 IN positions was compared to HIV-1 subtype B reference sequence (GenBank accession number: K03455) and thoroughly examined for the presence of primary mutations, nonpolymorphic and polymorphic mutations associated with resistance to INSTI. Each amino caid prevalence at each IN position was calculated and compared to the HIV-1 subtype B reference sequence (GenBank accession number: K03455). For this study NOP was defined as substitutions within the HIV-1 IN that occurred in ≥1% of the sequences. Positions with ≥20% substitutions were considered as highly polymorphic, while those with ≤0.5% substitution were considered highly conserved. This definition was used to calculate the NOPs for N-terminal domain (NTD), catalytic core domain (CCD) and C-terminal domain (CTD).

Generation of consensus HIV-1 integrase sequence

To comprehensively characterize the polymorphism (variability) in the IN sequences, we retrieved global subtype B and C IN sequences from the HIV Los Alamos National Library (LANL) database that matched the area (HXB2: 4230-5093 relative to HXB2 gene). To avoid overestimating variant calls and to validate those sequences used in our analysis were from INSTI-naive individuals, we only used sequences available prior 2007 (before FDA approved INSTIs). The online Quality Control program was used to verify the quality of all HIV-1 sequences. The analysis excluded sequences that had frameshifts, stop codons and/or poor quality. Only one sequence was retained per patient. If a patient had multiple sequences, the first sequence was chosen and used. IN's consensus amino acid sequence was created for the Ethiopian HIV-1 Subtype C, the global HIV-1 Subtype B and the global HIV-1 C sequence using BioEdit V7.0.9.0[326]. Both amino acids were represented for positions where they occur at higher frequencies than 30%. The consensus letter representing the most common amino acid was the first letter at the consensus. Furthermore, to assess the effect of prior exposure to ART on IN gene NOPs, consensus amino acid sequences of IN were generated from ART-experienced and ART-naive patients and compared. We also compared the consensus sequences of IN from patients who had one or more major HIVDRMs to a protease inhibitor, NRTI and/or NNRTIs (HIVDR Group) with those without major HIVDRMs in their respective protease/reverse transcriptase (PR/RT).

Docking of integrase strand transfer inhibitor to HIV-1 integrase

The crystallographic structure of full-length HIV-1 IN (6u8q.pdb) was obtained from the Protein Data Bank (www.rcsb.org)[429]. This structure includes a DNA fragment and dolutegravir (DTG). To visualize the PDR and ADR HIV-1 sequences, the monomer of 6u8q structure was modified by using UCSF-Chimera at 12 amino acid positions, then it was used in the following. In order to analyse the effect of PR and RT associated HIV drug resistance on the structure of HIV-1 IN, docking analysis was performed by using Autodock Vina (Vina) (Version 1.1.2)[430, 431]. To prepare the structure of IN, first the DNA fragment and water molecules were removed from crystal structure, then DTG (ligand) and IN files were saved separately. To create pdbqt files for docking with Vina, MGL Tools (Version 1.5.7rc1) was used[430, 431]. Vina uses two methods for local optimization: quasi-Newton and Broyden-Fletcher-Goldfarb-Shanno (BFGS) methods[430, 431]. Ligand was docked to binding site cavity, which is the catalytic site in the monomer of HIV-1 IN using x = 211.63 Å, y = 205.453 Å, and z = 171.895 Å Cartesian coordinates. To specify the certain grid positions 50 Å x 40 Å x 40 Å grid box dimensions were used. Docking calculations were performed with exhaustiveness option of 8 (average accuracy) and an energy range of 3. Re-docking DTG to the modified crystal structure of HIV-1 IN was performed to validate the docking method.

Genetic barrier to integrase strand-transfer inhibitor resistance

We first identified all wild-type triplets and their prevalence in our dataset of Ethiopian HIV-1 subtype C IN sequences (n=453) and global subtype B IN sequences (n=1884) to determine the extent of natural diversity at each selected position. Following that, we calculated a genetic barrier score for each wild-type triplet to evolve to a resistant amino acid at the chosen position. The genetic barrier was determined by adding the number of transitions and/or transversions required to evolve to any major drug-resistance substitution. As described by Nguyen *et al.*, we used a score of 1 for transition (C↔T and A↔G), 2.5 for transversion (A↔C, G↔C, A↔T, G↔T), and 0 for no change. We compared the genetic barrier to evolution of INSTI HIVDR substitution between HIV-1 subtype B and subtype C[432].

Statistical methods

Statistical tests were performed using SPSS 24 (IBM Corp., Armonk, NY, USA). HIVDR prevalence was determined with a confidence interval (CI) of 95% using the Wilson method. Categorical variables were compared using the 2-tailed Fisher's exact test, while continuous variables were compared using the Mann-Whitney 2-tailed U test. Trends over time were analyzed using the linear by-linear test for association (**paper II** and **paper III).** Logistic regression analysis was employed to

identify potential risk factor (**paper III** and **paper IV)**. In all the papers (**paper I-V**), p-value <0.05 was considered statistically significant.

Ethical approval

Scientific and ethical approval was granted by the Research and Ethical Clearance Committee of the Ethiopian Public Health Institute, and the National Health Research Ethics Review Committee of Ministry of Science and Technology of Ethiopia. All participants included in this thesis work have provided written informed consent

Data availability

All the newly generated sequences for this PhD thesis work have been deposited in the GenBank with accession numbers: OM302554–OM303013, MH324937-MH325003, OL598713-OL598856.

Main findings and discussions

Paper I

Background: Ethiopia is among sub-Saharan country that has been hard hit by the HIV epidemic. Although previous studies have indicated that the epidemic is dominated by subtype C, the evolutionary and temporal dynamics of HIV-1 in Ethiopia are not well scrutinized. It is crucial to understand the epidemiological and evolutionary patterns of HIV in order to monitor its spread, evaluate, and implement HIV prevention strategies. Ethiopia, like many low-income countries, has sparse and incomplete HIV epidemiological data, making HIV epidemic surveillance difficult. However, with the increased availability of HIV genetic sequencing data and the development of phylogenetic and phylodynamic tools, molecular epidemiology analysis can now be used to describe the transmission dynamics and evolutionary history of HIV. In this study, we used a combined data set (n=1276) of newly sequenced HIV-1 *pol* sequence (n=144) and HIV-1 subtype C *pol* sequences (n=1132) retrieved from LANL collected from different regions of Ethiopia between 1986 and 2017, to elucidate the evolutionary trajectories and temporal dynamics of the HIV-1 epidemic in Ethiopia, we used state-of-art phylogenetic and phylodynamic methods, including both Bayesian coalescent and birth–death models to estimate the dynamics of the effective population size (N_e) and reproductive numbers (R_e) through time for the HIV-1 epidemic in Ethiopia

Major findings

- The ML phylogenetic tree identified two distinct and well-supported clades (C-EA and C'-ET) indicating two independent introductions of HIV-1 to Ethiopia from the eastern and southern African countries, respectively.
- Our transmission cluster analysis showed that the Ethiopian sequence tend to form a large cluster. However, using our definition of cluster, three large well-supported clusters of 259 (C-EA-259), 153(C'-ET-153) and 148 sequences (C-EA-148) were detected. These clusters were mixed with geographical locations, indicating intermixing of HIV epidemics in Ethiopia.
- Our molecular dating analysis showed, HIV was introduced to Ethiopia in 1975 (95% HPD: 1970–1979) a decade before the first case reported. The median year of tMRCA (95% HPD) for C-EA-259, C-EA-148 and C'-ET-

153 were estimated, 1975 (95% HPD: 1970–1979), 1976 (95% HPD: 1963–1985) and 1983 (95% HPD: 1975–1988) respectively

- The median estimated evolutionary rate was in the range 1.76–1.82×10−3 substitutions/site/year for the three clusters, with overlapping 95% HPD intervals.
- For all the three clusters, the maximum R_e values (6.13, 3.93 and 4.88, for C-EA-259, C-EA-148 and C'-ET-153, respectively) was detected during the early period of the epidemic indicating an expanding epidemic growth. The R_e remained consistently high ($R_e > 1$) until the beginning of the 1990s and dropped below the epidemiological threshold ($R_e < 1$) at the mid-1990s and remained below one till recent year.
- Basic reproductive number (R_0) was in the range 4.0–5.0 for all the three clusters, all with overlapping 95% HPD intervals.
- Bayesian skygrid inference showed a rapid rise in the effective population size (N_e) in all three clusters from the initial introduction period until shortly before the year 2000, followed by a decline and stabilization in N_e until recent years.
- The median growth rate using for C-EA-259, C-EA-148 and C'-ET-153 was 0.66, 0.61 and 0.80 $year^{-1}$ respectively.

To our knowledge this is the first comprehensive study that employed the state of phylogenetic and phylodynamic method to describe the HIV epidemic in Ethiopia. We showed that the two subtype C clades (C-EA and C'ET) circulating in Ethiopia are result of at least two independent HIV introductions from the East and southern Africa countries in 1975 (1970–1979) and 1983 (1975–1988), respectively. Our finding is in consistent with the previous studies that showed distinct phylogeography subdivision of HIV subtype C circulating in southern, East, and central African countries[408, 409, 411]. In our study we showed that the Ethiopian sequences, both C-EA and C'-ET clades, were forming large cluster with the basal root dominated by sequences from Burundi for the C-EA and southern Africa countries for C'-ET clade suggesting that C-EA has its origin from Burundi while the C'-ET clade has its origin from southern Africa countries. The exact source country for the C'-ET was challenging due to the high intermixing within southern Africa sequences. Although the geographical proximity associated interconnectivity has been linked with HIV transmission in Africa[433, 434], our result suggest other population movement might have played a role in HIV-1 subtype C introduction to Ethiopia. Similar observation has been shown in other part of the Africa[435].

Our molecular dating analysis showed HIV-1 has been circulated in Ethiopia for a decade before the first case was reported. Similar finding has been reported for other countries, for example in USA, HIV has been estimated to be introduced in 1969 (1966–1972), while the first case was reported in 1981[436]. Interestingly our estimate coincides with the previous estimate of HIV-1 introduction Ethiopia[410] and to other

east African countries[410, 437] and also the large population migration from Burundi which might have played a role in the HIV-1 C-EA clade dissemination in eastern Africa countries[410].

Our phylodynamic analysis showed that the HIV-1 epidemic in Ethiopia were characterized by an expanding epidemic growth ($R_e > 1$) from the beginning of the epidemic until the mid-1990s, followed by a sharp decline ($R_e < 1$) in HIV-1 transmissions. This was consistent with the routine serological data showing an increase in HIV infection among FSWs in the capital city (Addis Ababa) and the major cities along the trading route in Ethiopia during the early 1990[77, 438, 439].The rapid epidemic increase in the early years was most likely due to a lack of HIV awareness, high mobility among FSWs, high-risk sexual behavior, and high STI prevalence in the risk groups and the general population[79, 81].

Although, there is no national representative incidence trend data, the different studies done and retrospective serological data have shown the decline in the HIV-1 incidence since the 1995, supporting our findings[77, 86, 87]. Interestingly our result showed the decline in HIV epidemic occurred after 1995, ten years before ART was introduced in Ethiopia and corporate well to the UNAIDS incidence and prevalence estimates. The epidemic decline coincides well with timing of the different HIV prevention and public health awareness program implemented in Ethiopia. This highlights the significant impact of behavioral intervention and public awareness in reducing the HIV transmission[75, 82-84]. Several studies have shown a significant decline in the HIV prevalence following behavioral intervention. A study done among MSM in Europe and North America and heterosexual in Thailand have showed a substantial decline in the HIV incidence, following behavioral intervention[440-443]. Similarly, a study done in Zimbabwe and Uganda a significant decline in prevalence following the behavioral intervention[444, 445]. Although it is well known that ART has significantly contributed to lowering HIV transmission, mortality, and maintaining the epidemic's decline[179], our findings highlight the importance of scaling up behavioral and risk reduction interventions alongside ART in the HIV/AIDS control strategy.

Paper II

Background: With the increased access to ART, the emergency and transmission of HIVDR is inevitable. TDR has been a major concern as it will lead to decreased population-level efficacy of standard first- and second-line ART regimens[395]. HIVDR testing is routinely done before ART initiation and during virological failure to guide clinical management in high-income countries. However, this is not feasible or affordable in many resource-limited countries. Instead, WHO developed a minimal resources HIVDR threshold survey that will help to classify the prevalence of TDR, guide for selection of ART regimens and the necessary HIVDR

prevention intervention[418]. This method allows for the classification of drug resistance among HIV-infected individuals into three categories: low prevalence (< 5%), moderate prevalence (5-15%) or high prevalence (≥15%). Low prevalence indicates that there are no modifications needed to the existing standard drug regimens, while high resistance rates suggest that changes to the current regimens are necessary. Moderate prevalence should be considered a warning sign and should prompt consideration of alternative treatments[418]. In this study, we used the WHO HIVDR threshold survey method to assess transmitted HIVDR in Gondar. We also performed phylogenetic and phylodynamic analysis using a combined dataset (newly generated and retrieved all publicly available Gondar HIV-1 subtype C *pol* sequences) to describe HIV-1 temporal dynamics and transmission cluster in Gondar.

Major findings

- Three of the 47 consecutively collected and sequenced specimens contained major HIVDRMs, indicating a moderate TDR level in Gondar. However, four (6%) of the 67 successfully sequenced samples were found to have major HIVDR mutations.
- All the HIVDRMs (two K103N, one G190S and one Y181C) were associated with NNRTIs and no NRTI or PI-associated HIVDRMs were identified among the sequenced samples.
- We identified 28 clusters (21 dyads, six medium-sized clusters, and one large cluster) which indicated multiple HIV introductions into Gondar, followed by local spread.
- Both the C-EA and C'-ET clades are circulating in Gondar, but the C-EA clade was the main circulating clade in Gondar.
- According to our Bayesian coalescent analysis, HIV-1 was introduced to Gondar between 1980 and 1990.
- We identified a transmission cluster with HIVDRMs (G190A) that showed a high rate of onward transmission.

This study was the first TDR survey done among young ART-naïve individuals in Gondar using WHO recommended guideline. In this study we also included sequences obtained from previous HIVDR studies done in Gondar (n=301) to comprehensively describe the HIVDR trend and molecular epidemiology of HIV in Gondar.

The overall prevalence of TDR (6%) detected in our study after 8 years of ART roll-out was consistent with observations of increased TDR prevalence after ART implementation in sub-Saharan Africa 6–8 years after ART roll-out[254, 262, 446]. Although direct comparison of the temporal trend is difficult due to the different methods used, the prevalence of DRM among different age groups in Gondar was not different, despite the fact that TDR differences between age groups and gender

have been reported from other sub-Saharan African countries. All of the DRMs identified in this study were linked to the NNRTIs (EFV and NVP). This finding is not surprising given the low genetic barrier of these drugs to resistance development and their widespread use as part of first-line ART regimens. Furthermore, the mutation detected in our study were K103N, G109S, and Y181C, which have reported to be the most common NNRTI-associated mutations in all world regions and HIV subtypes[262]. In comparison to previous in Gondar studies, we did not find DRMs associated with PIs and NRTIs, which could be attributed to the generally small sample size seen in threshold studies[244, 245, 447].

Our phylogenetic analysis revealed multiple HIV introductions in Gondar, with the oldest introduction in 1980 followed by local transmission, a phenomenon previously described for other local HIV epidemics. In this study, we discovered a link between clustering and increased transmission of viruses containing NNRTI DRMs. We discovered one cluster with the G190A mutation that had DRM transmissions for at least eight years. The G190A mutation is a slowly reverting HIVDR mutation that has the potential to persist and spread when found in a population with frequent transmission[448, 449].

Paper III

Background: Because of the large number of HIV+ patients in need of ART, WHO recommends decentralizing HIV care alongside ART scale-up[450]. HIV care decentralization allows more people to receive HIV treatment. It also facilitates patient access to care in areas where hospitals are difficult to reach. However, health centers have fewer resources and care providers are with lower levels of training than hospitals. The majority of patients in Ethiopia receive care in health centres, and many patients who begin ART in health centres have advanced disease. A previous study from the same setting showed that around 20% have active TB at the time of ART initiation which may jeopardize virologic suppression and increase the risk of HIVDR[451]. In this study, we aimed to assess the emergence of HIVDR and associated factors among patients receiving ART in five health centers during the first year after ART initiation. HIVDR testing was done on all samples with viral load ≥ 500 copies/mL at six and/or 12 months and on baseline samples from those with ADR mutation during ART.

Major findings

- Among the 729 subjects who started ART during follow-up, 621 individuals had VL data at six and/or 12 months after starting ART, of which 101 (16.3%) had VL ≥500 copies/ml.

- HIVDRMs were identified per sequence from 98 samples obtained during ART (VL ≥ 500 copies/mL) and major HIVDRM were detected in 64/98 (65.3%) of participants.
- Among the detected HIVDRMs per sequence 64 (100%) conferred resistance to NNRTI and 35 (54.7%) had both NRTI and NNRTI resistance mutations.
- Among the 64 patients with ADR, pre-ART resistance testing was done for 56/64 (88%) patients, and PDR was found in 7/56 (12.5%), and all were associated with NNRTI (K103N, K181V, G190A) while one patent has dual-class DRMs (D67N, T215C, K219E).
- Low mid-upper arm circumference and high pre-ART VL, and lower CD4 T-cell count were associated with an increased risk of DRMs acquisition.
- Although 12/64 (18.8%) patients with HIVDRM had active tuberculosis, TB was not associated with the development of HIVDRMs.

In this study we showed that HIVDR attribute for the 65% of the virological failure among patient on first-line ART during the first six to 12 month. Similar findings have been reported among patients receiving care in different hospitals in Ethiopia[251, 452, 453]. Thus, our finding highlights on the importance of strengthening adherence support and virological monitoring in Ethiopia to maximize the ART outcome and minimize the emergency and transmission of HIVDR. In this study, we also showed that TB coinfection is not associated with an increased risk of ADR; this might be due to the regular follow-up and adherence support for TB patients.

Paper IV

Background: In Ethiopia, FSWs have been at high risk of HIV infection since the beginning of the epidemic in Ethiopia and have been identified as key drivers of HIV transmission[74, 454]. ART was made free in Ethiopia in 2005, and it has since been expanded to include all who tests positive for HIV. Despite HIV prevention and treatment advancements, Ethiopia still has limited access to regular virologic monitoring. This delays the identification of patients experiencing treatment failure and increases the risk of HIVDR and further transmission of HIVDR[254, 455]. This is more apparent among FSWs who have limited access to HIV treatment and prevention. FSWs with high mobility are also less likely to have access to regular virologic monitoring and HIVDR testing[456, 457]. Given the potential risk of transmission to the general population, we aimed to study the prevalence of viral load non-suppression (VLN), HIVDR prevalence and associated factors among FSWs in Ethiopia.

Major findings

- 1172 (24%) of the 4900 total participants in the survey were HIV-positive.
- ART uptake (FSWs who were HIV-positive and receiving ART) was 20.7 % based on self-report.
- Among the 381 ART-naïve participants with genotyping results, 63 (16.5%) had one or more major HIVDRMs.
- PDR was most prevalent against NNRTIs (55/381 (14.4%), 10.5% (40/381) against NRTI, and 9.2% (33/381) against dual-class DRMs (NRTI/NNRTI). There was no PI-associated mutation detected.
- The 90% of NNRTI PDR mutations were associated to five DRMs (K103N, Y181C, G190A/E/S, K101E/P, and V106M).
- M184V and TAMs (M41L, D67G/N, K70R, L210W, T215F/Y, and K219E/Q) were the most common NRTI DRMs, accounting for 58.7% (37/63) and 27.0% (17/63) of the NRTI PDR, respectively.
- 59 (24.7%) of 239 participants receiving ART were not virologically suppressed. DRMs were detected in 29 (74.4%) of the 39 (66.1%) specimens which were successfully genotyped. Of these, 29 (100%) had NNRTI DRMs, 23 (79%) had NRTI DRMs, and 23 (89%) had dual-class DRMs.
- The genotypic susceptibility scores of individual antiretroviral drugs revealed that many of the specimens had high levels of resistance to several of the most frequently used first-line ART drugs in Ethiopia. The majority of specimens (69.0%) demonstrated high-level resistance to 3TC and TDF, NVP (100%), EFV (86.2%), and rilpivirine (RPV) 51.7%.
- VLN was associated with age 35 years and above, being forced into selling sex and CD4+ T cell count < 350 cell/mm^3. ADR and PDR were associated with CD4+ T cell count < 350 cell/mm^3.

In this study, we showed FSWs in Ethiopia have low ART uptake, which could be attributed to their high mobility and the stigma associated with sex work and HIV[457, 458]. Similar finding has been reported from other African countries countries[459-463]. Improving ART uptake will not only benefit FSWs health but will also lower the risk of HIV-1 transmission to their clients and the general population. Hence our finding suggests the need of programmatic intervention targeting FSWs to reduce onward HIV transmission to the general population[464-466].

We also found a high virological failure among FSWs on ART, this could be due to a combination of factors, such as stigma, low adherence to ART, as well as high mobility, that prevent FSWs from accessing the HIV care continuum [457, 458]. Several studies in sub-Saharan Africa have also found high levels of virological failure among FSWs, which is consistent with our findings[459, 467-469]. Among patients with VF, a high proportion of FSWs were found to have multiple mutations to the various ART drugs commonly used in Ethiopia. This could be due to prolonged exposure

to a failing regimen as a result of a lack of regular viral load monitoring [470, 471]. Aside from the limitations in the selection of effective treatment regimens for VLN patients, the high prevalence of HIVDR detected among study participants highlights the potential risk of HIVDR transmission to the general population.

Furthermore, when people with multiple DRMs are switched to second-line therapy, there is a risk of introducing functional monotherapy, which may be associated with a significant risk of subsequent virologic failure and the emergence of HIVDR[244, 245, 447, 472]. We also found a high PDR among FSW when compared to the general population, emphasizing the vulnerability of FSWs to the emergency and transmission of HIVDR, as well as the risk of onward transmission to the general population. Similarly, a high level of PDR has been reported among communities and groups with high-risk behavior[473, 474].

Paper V

Background: Following the global increase of PDR to NNRTIs, the WHO recommended the transition from NNRTI to INSTI-based regimens in both treatment-naive and treatment-experienced patients[475]. Many LMICs, including Ethiopia, have already switched to dolutegravir (DTG-based) regimens[242]. DTG has a high genetic barrier to resistance, a good safety profile, and a low drug-drug interaction potential. However, subtype-associated differences in naturally occurring polymorphisms (NOPs) have been linked to the development of different mutational pathways. As a result, different HIV-1 subtypes have different levels of HIVDR against INSTIs[476-479]. In this study, we aimed to examine HIV-1 subtype C IN genotypic profiles to determine the prevalence of PDR and NOPs that could affect the genetic barrier to resistance in INTSI-naive HIV-1 patients in Ethiopia.

Major findings

- Regardless of previous ART (NNRTI, NRTI, and/or PI) exposure, no DTG-associated HIVDRMs were detected among INSTI-naïve individuals. However, we found E92G in one patient specimen, and accessory mutations in 20 (4.3 %) specimens.
- No difference in the prevalence of accessory mutation among the ART-naïve and ART-experienced patient were observed.
- A high similarity was also observed in the comparison of the consensus IN sequence for ART-experienced and ART-naive patients.
- An overall of 64.9% (187/288) of the IN amino acid positions of the HIV-1 subtype C, were conserved (<1.0% variability). The N-terminal domain (NTD), the catalytic core domain (CCD) and the C-terminal domain (CTD) were 60% (30/60), 66.1% (107/162), and 66.8% (50/76) conserved, respectively.

- The majority of amino acids involved in key functions of the enzyme and the catalytic triad $D_{64}D_{116}E_{152}$ were fully conserved
- Subtypes B and C had similar genetic barriers to DTG resistance at selected amino acid positions, except for subtype C having a higher genetic barrier to G140C and G140S mutations than subtype B, indicating that the Q148H/K/R DTG resistance pathway is less selected in subtype C.
- Docking analysis of the DTG revealed that the PR- and RT-associated HIVDRM had no effect on the native structure of the HIV-1 IN, implying that DTG could be used as a salvage therapy for patients who have developed resistance to drugs targeting these enzymes.

In this study we found no major INSTI DRMs among INSTI naïve patients, which was not unexpected and consistent with other studies that found no or very few major INSTI mutations in INSTI-naive patients[480-485]. We also showed that the PR- and RT-associated HIVDRM had no effect on the structure of the HIV-1 IN, implying that DTG could be used as a salvage therapy for patients who have developed resistance to drugs targeting these enzymes. However, the detection of INSTI-associated accessory mutations and NOPs, which can affect the IN-protein function, the genetic barrier to INSTI resistance and susceptibility, warrant the need for continuous surveillance of INSTI resistance.

Clinical trial data and observational studies in a setting with routine viral load monitoring and HIVDR testing have shown that the DTG-based regimen performs exceptionally well in ART-naïve and ART-experienced patients without significant background resistance. Other clinical trials and observational studies, on the other hand, have reported the emergency of DTG resistance among patients on first-line DTG-containing therapy and treatment-experienced adults, highlighting that with the widespread implementation of DTG-based therapy, the gradual development and transmission of HIVDR against INSTIs will be unavoidable, rendering existing therapies ineffective and increasing the risk of virological failure[263-268]. This is especially true in LMICs, where patient monitoring is suboptimal and access to viral load monitoring for early detection of virological failure is limited. Furthermore, switches between regimens are often implemented without viral load testing. Many patients might accumulate NRTI resistance and be on functional DTG monotherapy, which might blunt the effectiveness of this regimen, thereby increasing the risk of virological failure and emergence and transmission of drug-resistant[486]. Studies done in Togo have shown switch to DTG-based first-line therapy without viral load testing will lead to 50% of adults and adolescents and almost all children with virological failure to a functional DTG monotherapy. While for those with virological failure, switching to second-line regimen will result in 30% of functional DTG monotherapy due to the accumulated NRTI mutation while on the first-line and second-line. Furthermore, in this study, 12% of the INSTI-naïve patient were

found to harbor INSTI resistance NOPs, whose long-term effect in the context of functional monotherapy is unknown[487].

A prospective cohort study in Malawi showed that among three patient with virological failure who were switched to TLD, two of these individuals developed treatment-emergent resistance to DTG (R263K, G118R) after only six months on TLD[488]. Another recent Malawi's National HIV Treatment Program study showed that among the 27 patients with virological failure while on DTG based therapy, 8 (30%) had DTG resistance, highlighting the risk of DTG resistance emergency among patients transitioning from NNRT to DTG-based therapy without viral load testing at the time of switch[489]. A recent systematic review of the genetic mechanism of DTG resistance by Rhee *et al*. and Cevik *et al*. have identified the risk of functional monotherapy among highly treatment-experienced patients, which leads to a high risk of virological failure and emergency of DTG resistance[490, 491]. A recent study in Botswana also showed that 32% of treatment-experienced patients with failing INSTI-based regimens have DRMs to DTG and other integrase inhibitors and of which 36% have 4-class multidrug-resistant[265]

Due to the recent introduction of DTG-based therapy in Ethiopia, data regarding its effectiveness among treatment naïve and ART-experienced patients in Ethiopia is lacking. However, the rates of HIVDR to TDF and 3TC at the time of virologic failure among patient on NNRTI-based regimens in Ethiopia is high[250, 492]. As a result, more research is needed to better understand the impact of NNRTI, NRTI, and PI-associated DRMs on clinical outcomes of DTG-based therapy in Ethiopia.

Conclusions, limitations and future perspectives

A comprehensive understanding of local HIV-1 epidemics is essential for monitoring the spread, designing, implementing, and evaluating HIV prevention strategies. Routine surveillance systems have been the backbone of public health efforts to track infection cases and their distribution over time, person, and location[386]. On the other hand, traditional epidemiologic methods, are labour intensive, expensive, difficult to collect, and the estimates derived from them are very error prone due to variations in reporting rate and intensity of surveillance[386]. Furthermore, it may be difficult to calculate key epidemiological variables such as R_e, reproduction rate, prevalence, and incidence. However, recent advances in the field of phylodynamics have made it possible to use phylogenetic methods to quantify transmission dynamics as well as key epidemiological parameters such as R_e, reproduction rate, prevalence, and incidence solely from viral sequence data[386, 387]. Thus, in **paper I**, we employed the state-of-the-art phylogenetic and phylodynamic tool to describe the evolutionary history and dynamics of HIV epidemic in Ethiopia to provide evidence to inform future HIV control and prevention efforts. In this study, we used both coalescent and birth-death phylodynamics models to characterize the dynamics of the HIV-1 epidemic in Ethiopia. To our knowledge, this is the first study to use a comprehensive dataset of *pol* sequence collected over 30 years (1986-2017). Our finding showed that the HIV epidemic in Ethiopia originated from two independent introduction in the mid-1970s and mid-1980s which coincides with the estimates in other Eastern Africa countries including, Uganda, Kenya and Tanzania[410]. Furthermore, the phylodynamic analyses revealed that the Ethiopian epidemic dynamics were characterised by an expanding epidemic growth from the beginning of the epidemic until the mid-1990s, followed by a sharp decline in HIV-1 transmissions, which is consistent with the routine surveillance data and UNAIDS estimates[493]. The R_e decreased many years before the introduction of ART and coincided with early behavioural, preventive, and public health awareness campaigns implemented in Ethiopia[493]. This highlights the significant impact of behavioral intervention and public awareness in reducing the HIV transmission[75, 82-84]. Several studies have shown a significant decline in the HIV prevalence following behavioral intervention. A study done among MSM in Europe and North America and heterosexual in Thailand have shown a substantial decline in the HIV incidence,

following behavioral interventions[440-443]. Similarly, a study done in Zimbabwe and Uganda found a significant decline in prevalence following the behavioral intervention[444, 445].

Phylodynamics can be used to estimate changes in viral population size over time and to investigate how various factors (epidemiological, evolutionary, behavioural, and so on) influence these changes. Such analyses are frequently performed on sets of HIV genetic sequences gathered as part of national or regional HIV drug resistance surveillance efforts. When combined with epidemiological, sociodemographic, or behavioural data, it can be used to identify and monitor HIV transmission clusters, as well as identify which populations are most likely to spread the HIV epidemic and target subsequent public health interventions. Thus, extending the methodology to other risk groups would provide valuable information on HIV-1 dynamics and evolution within and between risk groups. This information will be useful for evidence-based prevention strategies and will contribute to various areas of HIV prevention and control prevention among high-risk groups.

Despite the fact that current treatment regimens are very effective at suppressing viral load and reducing transmission rates, HIV continues to spread, indicating that significant challenges remain to be overcome[66]. A comprehensive HIV management strategy should include not only ART but also social support. Future research should combine genetic and epidemiological data to assess patterns and causes of HIV transmission and develop strategies that influence individual and societal behaviour, which will have a significant impact on HIV transmission control. Ethiopia's HIV-1 subtype C epidemic has a similar history to that of other east African countries such as Kenya, Uganda, and Tanzania[410]. However, our findings and previous research have shown that the HIV-1 epidemic in Ethiopia is dominated by subtype C, whereas in neighbouring countries such as Kenya, different subtypes such as subtype A, C, and D are circulating[494]. Future, molecular, and epidemiological research should aim to identify factors that contribute to such distinctiveness.

Nonetheless, the study of molecular sequences in an epidemiological context is not without pitfalls, and the accuracy of evolutionary or historical estimates is sensitive to various levels of bias, such as sample size, patient demographics, and data collection period. Furthermore, many ethical issues surround the use of HIV-1 sequence data, such as data privacy and informed consent[495, 496]. Phylogenetic studies should balance the public health significance and the ethics.

Phylogenetic tools have been demonstrated to be most effective in resource-rich countries with a concentrated HIV epidemic and are regarded as a key tool in HIV-1 epidemic control[497]. However, several obstacles must be overcome in order to improve the practice of molecular network-guided targeted interventions in developing countries. In most developing countries, including Ethiopia, HIVDR testing is limited to surveillance activities, and very few sequences are available in

the public database; additionally, the sequences in the database lack the necessary sociodemographic, risk factor, and geographic information. The initiative, like the Phylogenetics And Networks for Generalized HIV Epidemics in Africa (PANGEA-HIV) consortium, which aimed to generate near-full-length HIV-1 sequences from across Sub-Saharan Africa, needs to be strengthened[498].

Over the last few decades, the rapid expansion of ART has significantly reduced the risk of transmission and improved the survival and quality of life of HIV-infected patients. However, global evidence suggests that the rapid expansion of ART is associated with a rapid and ongoing increase in PDR and ADR, posing an imminent threat to the planned elimination of HIV as a public health threat. In this PhD thesis work, we assessed the prevalence of PDR in different population groups (young newly HIV-1 infected adults (**paper II**), general population (**paper III**) and risk group (FSWs) (**paper IV**). Our finding indicated that HIVDR will be an eminent challenge for the HIV epidemic control in Ethiopia. The sampling time (delay in sampling might lead to reversion) and the use of population-based sanger sequencing, which fail to detect minority variant present in less than 20% of the total population might have affected the prevalence reported in our study. Studies using NGS, which allows for the detection of resistant variants accounting for about 1% of the viral population, may provide a more accurate estimate of PDR[273]. In **paper III**, we showed that 65% of the virological failures occurring during the first six and/or 12 months on ART were due to HIVDR, highlighting the importance of adherence support and regular virological monitoring in Ethiopian ART program. Before switching treatments, the Ethiopian ART guideline recommends three months of adherence counselling and viral load re-testing. In this case, patients could be on a failing regimen for at least three months, resulting in additional HIVDRM accumulation. It will be interesting to investigate the clinical significance of adherence counselling after the first virological failure.

Since the beginning of the HIV epidemic in Ethiopia, FSWs have been a key driver of HIV transmission in the country[77, 499]. In our study **paper IV**, we demonstrated that FSWs have suboptimal ART uptake, high virological non-suppression, and HIVDR levels, indicating that FSWs may be at high risk of transmitting HIVDR to their clients and the general population. Several factors may have contributed to the low ART uptake and emergency of HIVDR among FSWs, including HIV stigma and sex work, FSW mobility that prevents FSWs from accessing the HIV care continuum, and low retention in care[457]. In addition, a lack of regular viral load monitoring, a high level of violence among FSWs, and poor ART adherence may have contributed to the emergency and accumulation of HIVDR among FSWs[492, 500]. Improving FSWs' access to ART not only improves their survival and health, but it also lowers the risk of transmission to the general population. Our findings emphasise the significance of FSWs targeted programmatic intervention in Ethiopia to improve ART access, maximise the benefit of ART, and limit the spread of HIV

and HIVDR. However, because we relied on participant self-report to determine ART status, there is a possibility that we misclassified some people who had previously been exposed to ART but were afraid to disclose it for fear of being discriminated against.

In **paper V**, we showed that, there were no DTG-associated HIVDRMs, regardless of prior ART exposure, among INSTI-naïve patient in Ethiopia. Our docking analysis of dolutegravir also revealed that HIVDRMs associated with PR and RT did not affect the native structure of the HIV-1 integrase. In general, our findings support the implementation of a wide scale-up of DTG-based regimes in Ethiopia. However, the detection of polymorphism and accessor mutation that can contribute to the INSTI resistance warrant the need of regular virological monitoring and INSTI resistance testing among patient on DTG-based therapy. In this study we used population-based Sanger sequencing methods, which might underestimate the prevalence of INSTI DRMs among our study participants. Our findings and conclusions are based on patients with no INSTI exposure, further study among patients on INSTI-based therapy is required to investigate the clinical significance of the HIVDRMs associated with PR and RT on DTG-based therapy.

Acknowledgements

This PhD thesis would not have been possible without the help of several people who either directly contributed to the PhD study or had a positive impact on my educational path. I am grateful to each and every one of them. First and foremost, I'd like to thank Prof. Patrik Medstrand, my PhD supervisor, for his excellent advice, motivation, enthusiasm, unwavering support, and patience throughout my studies. Your extensive knowledge and experience, as well as your dedication, have had an impact on both my academic studies and my daily life. I was extremely fortunate to have you as a supervisor who genuinely cared about both my work and my personal life. You were not only the supervisor, but also the big brother. Thank you for giving me the tools I needed to move forward with more determination, passion, and conviction than I've ever had before. Words cannot express how grateful I am for your assistance; only "Thank you" will suffice. I honestly hope that our working relationships and friendships last for many years. Thank you so much for everything.

Prof. Per Björkman, my co-supervisor, you were the primary initiator of this long PhD journey and have been incredibly supportive throughout. Your extensive knowledge, motivation, and enthusiasm for research are a true source of inspiration for me. You are an excellent scientist and supervisor. Thank you for your noble guidance, critical advice, constructive feedback, and encouragement. Thank you for being there for me during the difficult time. Thank you for your assistance in advancing Ethiopia's research capacity and health programme.

Joakim Esbjörnsson, my co-supervisor, thank you for your encouragement, willingness to share your knowledge and experience, and insightful suggestions, all of which have contributed to the development of my scientific perspective. Above all, for creating, together with Marianne, Patrik, and the team, an amazing scientific environment.

Prof. Marianne Jansson, you are a diligent, kind, and extremely passionate immunology expert who consistently comes up with intriguing research questions. You demonstrate an admirable commitment to your research. Thank you for fostering a scientific environment, and I'll never forget you saying, "I'll send you the publication," at the conclusion of every discussion.

I would like to express my heartfelt gratitude to all the co-authors for their significant scientific contributions, which enabled the publication to be completed

successfully and on time. Special thanks to Denise Kühnert and Luis Roger Esquivel Gomez of the Max-Planck Institute for the Science of Human History in Jena, Germany, for their assistance with phylodynamic analysis.

I would like to express my heartfelt gratitude to all the brilliant friends and colleagues who have inspired me over the years:

Sara Karlson and Malin Neptin, you are the team builders and the backbone of the virology team. Thank you for creating a happy and social environment, as well as for your generosity and genuine concern for the students in the virology team.

My PhD colleagues, Jamirah Nazziwa, George Makau, and Emil Johansson, thank you for our fruitful and encouraging discussions, which have helped me to develop my perspective and attitude toward achieving my research goal. Emil, you're up next; best of luck.

The Postdoctoral team, Angelica Palm, Zsófia Szojka, Nordine Bakouche, Sviataslau Sasinovich, and Amin Hassen: - thank you for being such an amazing, supportive group of people and for sharing your diverse career paths. Sviat, - thank you for all your selfless assistance, for introducing me to the various software programmes, and for always being there for me in times of need. Zsófia, - thank you for being an excellent co-author, a kind and good listener, and for your invaluable assistance in developing the figures for this thesis work.

I am also grateful to Claus Bohn Christiansen, Luis Sarmiento-Pérez, Eva Maria Feny, and Fariba Vaziri-Sani, for the many thoughtful discussions.

I would like to take this opportunity to thank the Adama research team members: - Anton Repalu, Olof Elvstam, John Walles, Fregenet, Adugna Negussie, Selmawit Mekuria, Illili Jemal, Niclas Winqvist, Sten Skogmar, Johannes Thorman, Oskar Olsson. Special thanks to Anton, for your contributions, thoughtful discussions, kind support, and a lot of fun events during my PhD studies.

I would like to express my heartfelt gratitude to the CDC-Atlanta staffs, Clement Zeh, Joy Chih-Wei Chang, Arthur Ramos, Sehin Birhanu, Wagar, Nick Wagner, Duping Zheng, Joshua DeVos, for your kindness and hospitality, and giving me the opportunity to do my research work in your facilities. Working with you and sharing your experience and knowledge was an incredible experience.

I am extremely grateful to, Yenew Kebede and Taye Tolera, who have an incredibly very supportive persons in this journey. Thank you for making my journey smooth. I'd also like to express my heartfelt gratitude to Prof. Anette Agardh, for your kindness, wonderful care, and wholehearted support in establishing my family in the new environment.

I would like to thank my institute, Ethiopian Public Health Institute (EPHI), for providing me with the opportunity to further my education, as well as to my colleagues, Minilik, Kidist, Abebe, Habteyes, Atsbeha, Eleni, Wude, Gebremedhin,

Tesfaye and many others, for their encouragement and support throughout my journey.

I am grateful to Kebede Baissa and Gemechis Wakene for your kindness, unwavering support, and being there for my family when I needed it the most. Working with you was a pleasure that I will never forget. Menbere, Hiwot, and Merima, thank you for the enjoyable time we spent working together.

My heartfelt gratitude goes to my Atlanta family, Ahmedin, Hawa, Abdu, Fedila, Nadia, and Hammim, for your warm welcome and wonderful care in making my stay in Atlanta so enjoyable. It will always be etched in my memory and remain a part of me.

I would like to express my heartfelt appreciation to Lund University, the Swedish Research Council (Vetenskapsrdet), and donations to Lund University for providing funds to support studies in this thesis, without which this thesis could not have been completed.

Finally, I want to thank my family for your unending love and support, which keeps me motivated and confident. You've been my walking stick and driving force. I will be eternally grateful to my dear wife, Hirut Abebe, and my children, Emmahus and Maya. Hirute, I appreciate your unending love, understanding, and encouragement, as well as taking on all our family's responsibilities and bearing the burden. This book is dedicated to you.

Reference

1. Gottlieb MS, Schroff R, Schanker HM et al. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N Engl J Med* 1981; **305**: 1425-31.
2. Masur H, Michelis MA, Greene JB et al. An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. *N Engl J Med* 1981; **305**: 1431-8.
3. Durack DT. Opportunistic infections and Kaposi's sarcoma in homosexual men. *N Engl J Med* 1981; **305**: 1465-7.
4. Davis KC, Horsburgh CR, Jr., Hasiba U et al. Acquired immunodeficiency syndrome in a patient with hemophilia. *Ann Intern Med* 1983; **98**: 284-6.
5. Ragni MV, Lewis JH, Spero JA et al. Acquired-immunodeficiency-like syndrome in two haemophiliacs. *Lancet* 1983; **1**: 213-4.
6. Vieira J, Frank E, Spira TJ et al. Acquired immune deficiency in Haitians: opportunistic infections in previously healthy Haitian immigrants. *N Engl J Med* 1983; **308**: 125-9.
7. Ammann AJ, Cowan MJ, Wara DW et al. Acquired immunodeficiency in an infant: possible transmission by means of blood products. *Lancet* 1983; **1**: 956-8.
8. Curran JW, Lawrence DN, Jaffe H et al. Acquired immunodeficiency syndrome (AIDS) associated with transfusions. *N Engl J Med* 1984; **310**: 69-75.
9. Harris C, Small CB, Klein RS et al. Immunodeficiency in female sexual partners of men with the acquired immunodeficiency syndrome. *N Engl J Med* 1983; **308**: 1181-4.
10. Oleske J, Minnefor A, Cooper R, Jr. et al. Immune deficiency syndrome in children. *Jama* 1983; **249**: 2345-9.
11. Rubinstein A, Sicklick M, Gupta A et al. Acquired immunodeficiency with reversed T4/T8 ratios in infants born to promiscuous and drug-addicted mothers. *Jama* 1983; **249**: 2350-6.
12. Gilmore NJ, Beaulieu R, Steben M et al. AIDS: acquired immunodeficiency syndrome. *Can Med Assoc J* 1983; **128**: 1281-4.
13. Jaffe HW, Bregman DJ, Selik RM. Acquired Immune Deficiency Syndrome in the United States: The First 1,000 Cases. *The Journal of Infectious Diseases* 1983; **148**: 339-45.
14. Barré-Sinoussi F, Chermann JC, Rey F et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 1983; **220**: 868-71.

15. Gallo RC, Salahuddin SZ, Popovic M et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* 1984; **224**: 500-3.
16. Ratner L, Haseltine W, Patarca R et al. Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* 1985; **313**: 277-84.
17. Coffin J, Haase A, Levy JA et al. What to call the AIDS virus? *Nature* 1986; **321**: 10.
18. Clavel F, Guyader M, Guétard D et al. Molecular cloning and polymorphism of the human immune deficiency virus type 2. *Nature* 1986; **324**: 691-5.
19. Peeters M, Jung M, Ayouba A. The origin and molecular epidemiology of HIV. *Expert Rev Anti Infect Ther* 2013; **11**: 885-96.
20. Sharp PM, Robertson DL, Gao F et al. Origins and diversity of human immunodeficiency viruses. *Aids* 1994; **8**: S27-S42.
21. Apetrei C, Kaur A, Lerche NW et al. Molecular epidemiology of simian immunodeficiency virus SIVsm in U.S. primate centers unravels the origin of SIVmac and SIVstm. *J Virol* 2005; **79**: 8991-9005.
22. Henrickson RV, Maul DH, Osborn KG et al. Epidemic of acquired immunodeficiency in rhesus monkeys. *Lancet* 1983; **1**: 388-90.
23. Daniel MD, Letvin NL, King NW et al. Isolation of T-cell tropic HTLV-III-like retrovirus from macaques. *Science* 1985; **228**: 1201-4.
24. Barin F, M'Boup S, Denis F et al. Serological evidence for virus related to simian T-lymphotropic retrovirus III in residents of west Africa. *Lancet* 1985; **2**: 1387-9.
25. Peeters M, Chaix ML, Delaporte E. [Genetic diversity and phylogeographic distribution of SIV: how to understand the origin of HIV]. *Med Sci (Paris)* 2008; **24**: 621-8.
26. Castro-Nallar E, Pérez-Losada M, Burton GF et al. The evolution of HIV: inferences using phylogenetics. *Mol Phylogenet Evol* 2012; **62**: 777-92.
27. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med* 2011; **1**: a006841.
28. Bell SM, Bedford T. Modern-day SIV viral diversity generated by extensive recombination and cross-species transmission. *PLoS Pathog* 2017; **13**: e1006466.
29. Hahn BH, Shaw GM, De Cock KM et al. AIDS as a zoonosis: scientific and public health implications. *Science* 2000; **287**: 607-14.
30. Ayouba A, Akoua-Koffi C, Calvignac-Spencer S et al. Evidence for continuing cross-species transmission of SIVsmm to humans: characterization of a new HIV-2 lineage in rural Côte d'Ivoire. *Aids* 2013; **27**: 2488-91.
31. Ayouba A, Njouom R, Chia JE et al. Molecular characterization of a new mosaic Simian Immunodeficiency Virus in a naturally infected tantalus monkey (Chlorocebus tantalus) from Cameroon: a challenge to the virus-host co-evolution of SIVagm in African green monkeys. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2015; **30**: 65-73.
32. Sauter D, Kirchhoff F. Key Viral Adaptations Preceding the AIDS Pandemic. *Cell Host Microbe* 2019; **25**: 27-38.

33. Van Heuverswyn F, Li Y, Bailes E et al. Genetic diversity and phylogeographic clustering of SIVcpzPtt in wild chimpanzees in Cameroon. *Virology* 2007; **368**: 155-71.

34. Keele BF, Van Heuverswyn F, Li Y et al. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* 2006; **313**: 523-6.

35. Plantier JC, Leoz M, Dickerson JE et al. A new human immunodeficiency virus derived from gorillas. *Nat Med* 2009; **15**: 871-2.

36. Van Heuverswyn F, Li Y, Neel C et al. Human immunodeficiency viruses: SIV infection in wild gorillas. *Nature* 2006; **444**: 164.

37. Mitani JC. Demographic influences on the behavior of chimpanzees. *Primates* 2006; **47**: 6-13.

38. Bailes E, Gao F, Bibollet-Ruche F et al. Hybrid origin of SIV in chimpanzees. *Science* 2003; **300**: 1713.

39. Takehisa J, Kraus MH, Ayouba A et al. Origin and biology of simian immunodeficiency virus in wild-living western gorillas. *J Virol* 2009; **83**: 1635-48.

40. Kalish ML, Wolfe ND, Ndongmo CB et al. Central African hunters exposed to simian immunodeficiency virus. *Emerg Infect Dis* 2005; **11**: 1928-30.

41. Mossoun A, Calvignac-Spencer S, Anoh AE et al. Bushmeat Hunting and Zoonotic Transmission of Simian T-Lymphotropic Virus 1 in Tropical West and Central Africa. *J Virol* 2017; **91**.

42. 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.

43. Worobey M, Gemmel M, Teuwen DE et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* 2008; **455**: 661-4.

44. Chitnis A, Rawls D, Moore J. Origin of HIV type 1 in colonial French Equatorial Africa? *AIDS Res Hum Retroviruses* 2000; **16**: 5-8.

45. Sharp PM, Bailes E, Chaudhuri RR et al. The origins of acquired immune deficiency syndrome viruses: where and when? *Philos Trans R Soc Lond B Biol Sci* 2001; **356**: 867-76.

46. de Sousa JD, Müller V, Lemey P et al. High GUD incidence in the early 20 century created a particularly permissive time window for the origin and initial spread of epidemic HIV strains. *PLoS One* 2010; **5**: e9936.

47. Cameron DW, Simonsen JN, D'Costa LJ et al. Female to male transmission of human immunodeficiency virus type 1: risk factors for seroconversion in men. *Lancet* 1989; **2**: 403-7.

48. Marx PA, Alcabes PG, Drucker E. Serial human passage of simian immunodeficiency virus by unsterile injections and the emergence of epidemic human immunodeficiency virus in Africa. *Philos Trans R Soc Lond B Biol Sci* 2001; **356**: 911-20.

49. Drucker E, Alcabes PG, Marx PA. The injection century: massive unsterile injections and the emergence of human pathogens. *Lancet* 2001; **358**: 1989-92.

50. Gisselquist D. Emergence of the HIV type 1 epidemic in the twentieth century: comparing hypotheses to evidence. *AIDS Res Hum Retroviruses* 2003; **19**: 1071-8.

51. Zhu T, Korber BT, Nahmias AJ et al. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* 1998; **391**: 594-7.

52. Sharp PM, Hahn BH. The evolution of HIV-1 and the origin of AIDS. *Philos Trans R Soc Lond B Biol Sci* 2010; **365**: 2487-94.

53. Lemey P, Pybus OG, Rambaut A et al. The molecular population genetics of HIV-1 group O. *Genetics* 2004; **167**: 1059-68.

54. Wertheim JO, Worobey M. Dating the age of the SIV lineages that gave rise to HIV-1 and HIV-2. *PLoS Comput Biol* 2009; **5**: e1000377.

55. Korber B, Muldoon M, Theiler J et al. Timing the ancestor of the HIV-1 pandemic strains. *Science* 2000; **288**: 1789-96.

56. Clavel F, Guétard D, Brun-Vézinet F et al. Isolation of a new human retrovirus from West African patients with AIDS. *Science* 1986; **233**: 343-6.

57. Berzow D, Descamps D, Obermeier M et al. Human Immunodeficiency Virus-2 (HIV-2): A Summary of the Present Standard of Care and Treatment Options for Individuals Living with HIV-2 in Western Europe. *Clin Infect Dis* 2021; **72**: 503-9.

58. Gottlieb GS, Raugi DN, Smith RA. 90-90-90 for HIV-2? Ending the HIV-2 epidemic by enhancing care and clinical management of patients infected with HIV-2. *Lancet HIV* 2018; **5**: e390-e9.

59. Berry N, Jaffar S, Schim van der Loeff M et al. Low level viremia and high CD4% predict normal survival in a cohort of HIV type-2-infected villagers. *AIDS Res Hum Retroviruses* 2002; **18**: 1167-73.

60. da Silva ZJ, Oliveira I, Andersen A et al. Changes in prevalence and incidence of HIV-1, HIV-2 and dual infections in urban areas of Bissau, Guinea-Bissau: is HIV-2 disappearing? *Aids* 2008; **22**: 1195-202.

61. Ceccarelli G, Giovanetti M, Sagnelli C et al. Human Immunodeficiency Virus Type 2: The Neglected Threat. *Pathogens* 2021; **10**.

62. de Mendoza C, Cabezas T, Caballero E et al. HIV type 2 epidemic in Spain: challenges and missing opportunities. *Aids* 2017; **31**: 1353-64.

63. MMWR. HIV-2 Infection Surveillance--United States, 1987-2009. *MMWR Morb Mortal Wkly Rep* 2011; **60**: 985-8.

64. Evans BG, Gill ON, Gleave SR et al. HIV-2 in the United Kingdom--a review. *CDR (Lond Engl Rev)* 1991; **1**: R19-23.

65. Ibe S, Yokomaku Y, Shiino T et al. HIV-2 CRF01_AB: first circulating recombinant form of HIV-2. *J Acquir Immune Defic Syndr* 2010; **54**: 241-7.

66. UNAIDS. UNAIDS data 2021. (https://www.unaids.org/en/resources/documents/2021/2021_unaids_data) (Accessed 10 June 2022).

67. UNAIDS. Fact sheet - Latest global and regional statistics on the status of the AIDS epidemic. 2021. (https://www.unaids.org/en/resources/documents/2021/UNAIDS_FactSheet. (Accessed 10 June, 2022).

68. Gebregziabher M, Dai L, Vrana-Diaz C et al. Gender Disparities in Receipt of HIV Testing Results in Six Sub-Saharan African Countries. *Health Equity* 2018; **2**: 384-94.

69. Giovanetti M, Ciccozzi M, Parolin C et al. Molecular Epidemiology of HIV-1 in African Countries: A Comprehensive Overview. *Pathogens* 2020; **9**.

70. UNAIDS. UNAIDS. Global AIDS Strategy 2021-2026 — End Inequalities. End AIDS.https://www.unaids.org/en/resources/documents/2021/2021-2026-global-AIDS-strategy. (Accessed 12 June 2022).

71. UNAIDS. UN Joint Programme on HIV/AIDS (UNAIDS), UNAIDS data 2021. (https://www.unaids.org/en/resources/documents/2021/2021_unaids_data) Accessed 15 June 2022.

72. Lester FT, Ayehunie S, Zewdie D. Acquired immunodeficiency syndrome: seven cases in an Addis Ababa hospital. *Ethiopian medical journal* 1988; **26**: 139-45.

73. Tsega E, Mengesha B, Nordenfelt E et al. Serological survey of human immunodeficiency virus infection in Ethiopia. *Ethiopian medical journal* 1988; **26**: 179-84.

74. Mehret M KL, Zewdie D. Ayehunie S, Shanko B, Gizaw G, et al. HIV-1 infection and some related risk factors among female sex workers in Addis Ababa. *Ethiopian Journal of Health Development* 1990; **4: 171 –176**.

75. Kloos H, Mariam DH. HIV/AIDS in Ethiopia: An Overview. *Northeast African Studies* 2000; **7**: 13-40.

76. Assefa A, Rahlenbeck S, Molla K et al. Seroprevalence of HIV-1 and syphilis antibodies in blood donors in Gonder, Ethiopia, 1989-1993. *J Acquir Immune Defic Syndr (1988)* 1994; **7**: 1282-5.

77. Kebede D, Aklilu M, Sanders E. The HIV epidemic and the state of its surveillance in Ethiopia. *Ethiop Med J* 2000; **38**: 283-302.

78. Hladik W, Shabbir I, Jelaludin A et al. HIV/AIDS in Ethiopia: where is the epidemic heading? *Sex Transm Infect* 2006; **82 Suppl 1**: i32-5.

79. Desta S, Feleke W, Yusuf M et al. Prevalence of STD and STD related risk factors in sex workers of Addis Ababa. *The Ethiopian Journal of Health Development* 1990; **4**.

80. Mehret M KL, Shanko B, Belete F. Sexual behaviours and some social features off female sex workers in the city of Addis Ababa. *Ethiopian Journal of Health Development* 1990; **4 (2): 133-13**.

81. Negassa H, Kefene H, Khodakevich L et al. Profile of AIDS case in Ethiopia. *The Ethiopian Journal of Health Development* 1990; **4**.

82. Okubagzhi G, Singh S. Establishing an HIV/AIDS programme in developing countries: the Ethiopian experience. *Aids* 2002; **16**: 1575-86.

83. Hadgu T, G. Egziabher E, Gizaw G et al. Intersectoral collaboration in AIDS control in Ethiopia *The Ethiopian Journal of Health Development* 1990; **4**.

84. Zewdie D, Gizaw G, Khodakevich L et al. Development and management of the AIDS Control Programme in Ethiopia. . *The Ethiopian Journal of Health Development* 1990; **4**.

85. Mehret M, Mertens TE, Caraël M et al. Baseline for the evaluation of an AIDS programme using prevention indicators: a case study in Ethiopia. *Bull World Health Organ* 1996; **74**: 509-16.

86. Tsegaye A, Rinke De Wit TF, Mekonnen Y et al. Decline in prevalence of HIV-1 infection and syphilis among young women attending antenatal care clinics in Addis Ababa, Ethiopia: results from sentinel surveillance, 1995-2001. *J Acquir Immune Defic Syndr* 2002; **30**: 359-62.

87. Wolday D, Meles H, Hailu E et al. Temporal trends in the incidence of HIV infection in antenatal clinic attendees in Addis Ababa, Ethiopia, 1995-2003. *J Intern Med* 2007; **261**: 132-7.

88. Office FHAPaC. HIV Prevention in Ethiopia National Road Map 2018 - 2020.

89. Central Statistical Agency - CSA/Ethiopia, ICF. Ethiopia Demographic and Health Survey 2016. Addis Ababa, Ethiopia: CSA and ICF, 2017.

90. Institute EPH. HIV Related Estimats and Projections in Ethiopia for the Year-2020. 2021.

91. Institute EPH. HIV Related Estimats and Projections in Ethiopia for the Year-2019. 2020.

92. Seitz R. Human Immunodeficiency Virus (HIV). *Transfus Med Hemother* 2016; **43**: 203-22.

93. van Heuvel Y, Schatz S, Rosengarten JF et al. Infectious RNA: Human Immunodeficiency Virus (HIV) Biology, Therapeutic Intervention, and the Quest for a Vaccine. *Toxins (Basel)* 2022; **14**.

94. Guatelli JC, Siliciano RF, Kuritzkes DR et al. Human Immunodeficiency Virus. *Clinical Virology*, 2016; 795-840.

95. Holmes EC, Zhang LQ, Simmonds P et al. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. *Proc Natl Acad Sci U S A* 1992; **89**: 4835-9.

96. Zanini F, Puller V, Brodin J et al. In vivo mutation rates and the landscape of fitness costs of HIV-1. *Virus Evol* 2017; **3**: vex003.

97. Yeo JY, Goh GR, Su CT et al. The Determination of HIV-1 RT Mutation Rate, Its Possible Allosteric Effects, and Its Implications on Drug Resistance. *Viruses* 2020; **12**.

98. Abram ME, Ferris AL, Shao W et al. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol* 2010; **84**: 9864-78.

99. Zanini F, Brodin J, Albert J et al. Error rates, PCR recombination, and sampling depth in HIV-1 whole genome deep sequencing. *Virus Res* 2017; **239**: 106-14.

100. Preston BD, Poiesz BJ, Loeb LA. Fidelity of HIV-1 reverse transcriptase. *Science* 1988; **242**: 1168-71.

101. Drake JW. Rates of spontaneous mutation among RNA viruses. *Proc Natl Acad Sci U S A* 1993; **90**: 4171-5.

102. Perelson AS, Neumann AU, Markowitz M et al. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* 1996; **271**: 1582-6.

103. Ribeiro RM, Qin L, Chavez LL et al. Estimation of the initial viral growth rate and basic reproductive number during acute HIV-1 infection. *J Virol* 2010; **84**: 6096-102.

104. Levy DN, Aldrovandi GM, Kutsch O et al. Dynamics of HIV-1 recombination in its natural target cells. *Proc Natl Acad Sci U S A* 2004; **101**: 4204-9.

105. Hemelaar J. The origin and diversity of the HIV-1 pandemic. *Trends Mol Med* 2012; **18**: 182-92.

106. Immonen TT, Conway JM, Romero-Severson EO et al. Recombination Enhances HIV-1 Envelope Diversity by Facilitating the Survival of Latent Genomic Fragments in the Plasma Virus Population. *PLoS Comput Biol* 2015; **11**: e1004625.

107. Shankarappa R, Margolick JB, Gange SJ et al. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol* 1999; **73**: 10489-502.

108. Anderson JA, Teufel RJ, 2nd, Yin PD et al. Correlated template-switching events during minus-strand DNA synthesis: a mechanism for high negative interference during retroviral recombination. *J Virol* 1998; **72**: 1186-94.

109. Smyth RP, Davenport MP, Mak J. The origin of genetic diversity in HIV-1. *Virus Res* 2012; **169**: 415-29.

110. Lam TT, Hon CC, Tang JW. Use of phylogenetics in the molecular epidemiology and evolutionary studies of viral infections. *Crit Rev Clin Lab Sci* 2010; **47**: 5-49.

111. Zhuang J, Jetzt AE, Sun G et al. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J Virol* 2002; **76**: 11273-82.

112. Robertson DL, Anderson JP, Bradac JA et al. HIV-1 nomenclature proposal. *Science* 2000; **288**: 55-6.

113. Foley BT, Leitner T, Paraskevis D et al. Primate immunodeficiency virus classification and nomenclature: Review. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2016; **46**: 150-8.

114. Feng Y, Zhang C, Zhang M et al. First report of a novel HIV-1 recombinant form (CRF100_01C) comprising CRF01_AE and C among heterosexuals in Yunnan, China. *J Infect* 2018; **77**: 561-71.

115. 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.

116. Hemelaar J, Elangovan R, Yun J et al. Global and regional epidemiology of HIV-1 recombinants in 1990-2015: a systematic review and global survey. *Lancet HIV* 2020; **7**: e772-e81.

117. Esbjornsson J, Mild M, Mansson F et al. HIV-1 molecular epidemiology in Guinea-Bissau, West Africa: origin, demography and migrations. *PLoS One* 2011; **6**: e17025.

118. Vuilleumier S, Bonhoeffer S. Contribution of recombination to the evolutionary history of HIV. *Curr Opin HIV AIDS* 2015; **10**: 84-9.

119. Weissman DB, Hallatschek O. The rate of adaptation in large sexual populations with linear chromosomes. *Genetics* 2014; **196**: 1167-83.

120. Song H, Giorgi EE, Ganusov VV et al. Tracking HIV-1 recombination to resolve its contribution to HIV-1 evolution in natural infection. *Nat Commun* 2018; **9**: 1928.

121. Simon-Loriere E, Holmes EC. Why do RNA viruses recombine? *Nat Rev Microbiol* 2011; **9**: 617-26.

122. Brown RJ, Peters PJ, Caron C et al. Intercompartmental recombination of HIV-1 contributes to env intrahost diversity and modulates viral tropism and sensitivity to entry inhibitors. *J Virol* 2011; **85**: 6024-37.

123. Charpentier C, Nora T, Tenaillon O et al. Extensive recombination among human immunodeficiency virus type 1 quasispecies makes an important contribution to viral diversity in individual patients. *J Virol* 2006; **80**: 2472-82.

124. Nishimura Y, Shingai M, Lee WR et al. Recombination-mediated changes in coreceptor usage confer an augmented pathogenic phenotype in a nonhuman primate model of HIV-1-induced AIDS. *J Virol* 2011; **85**: 10617-26.

125. Kouri V, Khouri R, Alemán Y et al. CRF19_cpx is an Evolutionary fit HIV-1 Variant Strongly Associated With Rapid Progression to AIDS in Cuba. *EBioMedicine* 2015; **2**: 244-54.

126. Lau KA, Wang B, Miranda-Saksena M et al. Evidence for possible biological advantages of the newly emerging HIV-1 circulating recombinant form from Malaysia - CRF33_01B in comparison to its progenitors - CRF01_AE and subtype B. *Curr HIV Res* 2010; **8**: 259-71.

127. Nora T, Charpentier C, Tenaillon O et al. Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. *J Virol* 2007; **81**: 7620-8.

128. Moutouh L, Corbeil J, Richman DD. Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure. *Proc Natl Acad Sci U S A* 1996; **93**: 6106-11.

129. Rawson JMO, Nikolaitchik OA, Keele BF et al. Recombination is required for efficient HIV-1 replication and the maintenance of viral genome integrity. *Nucleic Acids Res* 2018; **46**: 10535-45.

130. Ritchie AJ, Cai F, Smith NM et al. Recombination-mediated escape from primary CD8+ T cells in acute HIV-1 infection. *Retrovirology* 2014; **11**: 69.

131. Streeck H, Li B, Poon AF et al. Immune-driven recombination and loss of control after HIV superinfection. *J Exp Med* 2008; **205**: 1789-96.

132. Liu SL, Mittler JE, Nickle DC et al. Selection for human immunodeficiency virus type 1 recombinants in a patient with rapid progression to AIDS. *J Virol* 2002; **76**: 10674-84.

133. Palm AA, Esbjörnsson J, Månsson F et al. Faster progression to AIDS and AIDS-related death among seroincident individuals infected with recombinant HIV-1 A3/CRF02_AG compared with sub-subtype A3. *J Infect Dis* 2014; **209**: 721-8.

134. Mild M, Esbjörnsson J, Fenyö EM et al. Frequent intrapatient recombination between human immunodeficiency virus type 1 R5 and X4 envelopes: implications for coreceptor switch. *J Virol* 2007; **81**: 3369-76.

135. Esbjörnsson J, Mild M, Audelin A et al. HIV-1 transmission between MSM and heterosexuals, and increasing proportions of circulating recombinant forms in the Nordic Countries. *Virus Evol* 2016; **2**: vew010.

136. Palm AA, Esbjörnsson J, Månsson F et al. Cocirculation of several similar but unique HIV-1 recombinant forms in Guinea-Bissau revealed by near full-length genomic sequencing. *AIDS Res Hum Retroviruses* 2015; **31**: 938-45.

137. Sallam M, Şahin G, Ingman M et al. Genetic characterization of human immunodeficiency virus type 1 transmission in the Middle East and North Africa. *Heliyon* 2017; **3**: e00352.

138. Nazziwa J, Faria NR, Chaplin B et al. Characterisation of HIV-1 Molecular Epidemiology in Nigeria: Origin, Diversity, Demography and Geographic Spread. *Scientific reports* 2020; **10**: 3468.

139. D'Arc M, Ayouba A, Esteban A et al. Origin of the HIV-1 group O epidemic in western lowland gorillas. *Proc Natl Acad Sci U S A* 2015; **112**: E1343-52.

140. 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.

141. Santos AF, Soares MA. HIV Genetic Diversity and Drug Resistance. *Viruses* 2010; **2**: 503-31.

142. Gartner MJ, Roche M, Churchill MJ et al. Understanding the mechanisms driving the spread of subtype C HIV-1. *EBioMedicine* 2020; **53**: 102682.

143. Salazar-Gonzalez JF, Salazar MG, Keele BF et al. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J Exp Med* 2009; **206**: 1273-89.

144. Dalgleish AG, Beverley PC, Clapham PR et al. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* 1984; **312**: 763-7.

145. Maddon PJ, Dalgleish AG, McDougal JS et al. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* 1986; **47**: 333-48.

146. McDougal JS, Kennedy MS, Sligh JM et al. Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. *Science* 1986; **231**: 382-5.

147. Wu L, Gerard NP, Wyatt R et al. CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature* 1996; **384**: 179-83.

148. Alkhatib G, Combadiere C, Broder CC et al. CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* 1996; **272**: 1955-8.

149. Choe H, Farzan M, Sun Y et al. The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell* 1996; **85**: 1135-48.

150. Feng Y, Broder CC, Kennedy PE et al. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* 1996; **272**: 872-7.

151. Philpott SM. HIV-1 coreceptor usage, transmission, and disease progression. *Curr HIV Res* 2003; **1**: 217-27.

152. Liu R, Paxton WA, Choe S et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 1996; **86**: 367-77.

153. Dean M, Carrington M, Winkler C et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science* 1996; **273**: 1856-62.

154. Samson M, Libert F, Doranz BJ et al. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* 1996; **382**: 722-5.

155. Wilkinson DA, Operskalski EA, Busch MP et al. A 32-bp deletion within the CCR5 locus protects against transmission of parenterally acquired human immunodeficiency virus but does not affect progression to AIDS-defining illness. *J Infect Dis* 1998; **178**: 1163-6.

156. Martinson JJ, Chapman NH, Rees DC et al. Global distribution of the CCR5 gene 32-basepair deletion. *Nat Genet* 1997; **16**: 100-3.

157. Salem AH, Batzer MA. Distribution of the HIV resistance CCR5-Delta32 allele among Egyptians and Syrians. *Mutat Res* 2007; **616**: 175-80.

158. Harrison SC. Viral membrane fusion. *Nat Struct Mol Biol* 2008; **15**: 690-8.

159. Schaller T, Ocwieja KE, Rasaiyaah J et al. HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. *PLoS Pathog* 2011; **7**: e1002439.

160. Cunha RF, Simões S, Carvalheiro M et al. Novel Antiretroviral Therapeutic Strategies for HIV. *Molecules* 2021; **26**.

161. Li C, Burdick RC, Nagashima K et al. HIV-1 cores retain their integrity until minutes before uncoating in the nucleus. *Proc Natl Acad Sci U S A* 2021; **118**.

162. Konvalinka J, Kräusslich HG, Müller B. Retroviral proteases and their roles in virion maturation. *Virology* 2015; **479-480**: 403-17.

163. Kharsany AB, Karim QA. HIV Infection and AIDS in Sub-Saharan Africa: Current Status, Challenges and Opportunities. *Open AIDS J* 2016; **10**: 34-48.

164. Røttingen JA, Cameron DW, Garnett GP. A systematic review of the epidemiologic interactions between classic sexually transmitted diseases and HIV: how much really is known? *Sex Transm Dis* 2001; **28**: 579-97.

165. Glynn JR, Biraro S, Weiss HA. Herpes simplex virus type 2: a key role in HIV incidence. *Aids* 2009; **23**: 1595-8.

166. Atashili J, Poole C, Ndumbe PM et al. Bacterial vaginosis and HIV acquisition: a meta-analysis of published studies. *Aids* 2008; **22**: 1493-501.

167. Cohen MS, Shaw GM, McMichael AJ et al. Acute HIV-1 Infection. *N Engl J Med* 2011; **364**: 1943-54.

168. Hladik F, McElrath MJ. Setting the stage: host invasion by HIV. *Nat Rev Immunol* 2008; **8**: 447-57.

169. Baeten JM, Kahle E, Lingappa JR et al. Genital HIV-1 RNA predicts risk of heterosexual HIV-1 transmission. *Sci Transl Med* 2011; **3**: 77ra29.

170. Baggaley RF, White RG, Boily MC. HIV transmission risk through anal intercourse: systematic review, meta-analysis and implications for HIV prevention. *Int J Epidemiol* 2010; **39**: 1048-63.

171. Maartens G, Celum C, Lewin SR. HIV infection: epidemiology, pathogenesis, treatment, and prevention. *Lancet* 2014; **384**: 258-71.

172. Quinn TC, Wawer MJ, Sewankambo N et al. Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. *N Engl J Med* 2000; **342**: 921-9.

173. Lingappa JR, Hughes JP, Wang RS et al. Estimating the impact of plasma HIV-1 RNA reductions on heterosexual HIV-1 transmission risk. *PLoS One* 2010; **5**: e12598.

174. Hogg RS, Rhone SA, Yip B et al. Antiviral effect of double and triple drug combinations amongst HIV-infected adults: lessons from the implementation of viral load-driven antiretroviral therapy. *Aids* 1998; **12**: 279-84.

175. Vernazza PL, Gilliam BL, Flepp M et al. Effect of antiviral treatment on the shedding of HIV-1 in semen. *Aids* 1997; **11**: 1249-54.

176. Cu-Uvin S, Caliendo AM, Reinert S et al. Effect of highly active antiretroviral therapy on cervicovaginal HIV-1 RNA. *Aids* 2000; **14**: 415-21.

177. Avettand-Fenoel V, Prazuck T, Hocqueloux L et al. HIV-DNA in rectal cells is well correlated with HIV-DNA in blood in different groups of patients, including long-term non-progressors. *Aids* 2008; **22**: 1880-2.

178. Castilla J, Del Romero J, Hernando V et al. Effectiveness of highly active antiretroviral therapy in reducing heterosexual transmission of HIV. *J Acquir Immune Defic Syndr* 2005; **40**: 96-101.

179. Cohen MS, Chen YQ, McCauley M et al. Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med* 2011; **365**: 493-505.

180. Cohen MS, Chen YQ, McCauley M et al. Antiretroviral Therapy for the Prevention of HIV-1 Transmission. *N Engl J Med* 2016; **375**: 830-9.

181. Rodger AJ, Cambiano V, Bruun T et al. Sexual Activity Without Condoms and Risk of HIV Transmission in Serodifferent Couples When the HIV-Positive Partner Is Using Suppressive Antiretroviral Therapy. *Jama* 2016; **316**: 171-81.

182. Rodger AJ, Cambiano V, Bruun T et al. Risk of HIV transmission through condomless sex in serodifferent gay couples with the HIV-positive partner taking suppressive antiretroviral therapy (PARTNER): final results of a multicentre, prospective, observational study. *Lancet* 2019; **393**: 2428-38.

183. De Cock KM, Fowler MG, Mercier E et al. Prevention of mother-to-child HIV transmission in resource-poor countries: translating research into policy and practice. *Jama* 2000; **283**: 1175-82.

184. Wood E, Milloy MJ, Montaner JS. HIV treatment as prevention among injection drug users. *Curr Opin HIV AIDS* 2012; **7**: 151-6.

185. Weiss HA, Quigley MA, Hayes RJ. Male circumcision and risk of HIV infection in sub-Saharan Africa: a systematic review and meta-analysis. *Aids* 2000; **14**: 2361-70.

186. Keele BF, Giorgi EE, Salazar-Gonzalez JF et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A* 2008; **105**: 7552-7.

187. Stacey AR, Norris PJ, Qin L et al. Induction of a striking systemic cytokine cascade prior to peak viremia in acute human immunodeficiency virus type 1 infection, in contrast to more modest and delayed responses in acute hepatitis B and C virus infections. *J Virol* 2009; **83**: 3719-33.

188. Daar ES, Moudgil T, Meyer RD et al. Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection. *N Engl J Med* 1991; **324**: 961-4.

189. Little SJ, McLean AR, Spina CA et al. Viral dynamics of acute HIV-1 infection. *J Exp Med* 1999; **190**: 841-50.

190. Mehandru S, Poles MA, Tenner-Racz K et al. Primary HIV-1 infection is associated with preferential depletion of CD4+ T lymphocytes from effector sites in the gastrointestinal tract. *J Exp Med* 2004; **200**: 761-70.

191. Guadalupe M, Reay E, Sankaran S et al. Severe CD4+ T-cell depletion in gut lymphoid tissue during primary human immunodeficiency virus type 1 infection and substantial delay in restoration following highly active antiretroviral therapy. *J Virol* 2003; **77**: 11708-17.

192. Mattapallil JJ, Douek DC, Hill B et al. Massive infection and loss of memory CD4+ T cells in multiple tissues during acute SIV infection. *Nature* 2005; **434**: 1093-7.

193. Smit-McBride Z, Mattapallil JJ, McChesney M et al. Gastrointestinal T lymphocytes retain high potential for cytokine responses but have severe CD4(+) T-cell depletion at all stages of simian immunodeficiency virus infection compared to peripheral lymphocytes. *J Virol* 1998; **72**: 6646-56.

194. Daar ES, Pilcher CD, Hecht FM. Clinical presentation and diagnosis of primary HIV-1 infection. *Curr Opin HIV AIDS* 2008; **3**: 10-5.

195. Brenner BG, Roger M, Routy JP et al. High rates of forward transmission events after acute/early HIV-1 infection. *J Infect Dis* 2007; **195**: 951-9.

196. Wawer MJ, Gray RH, Sewankambo NK et al. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *J Infect Dis* 2005; **191**: 1403-9.

197. Fisher M, Pao D, Brown AE et al. Determinants of HIV-1 transmission in men who have sex with men: a combined clinical, epidemiological and phylogenetic approach. *Aids* 2010; **24**: 1739-47.

198. Lundgren JD, Babiker AG, Gordin F et al. Initiation of Antiretroviral Therapy in Early Asymptomatic HIV Infection. *N Engl J Med* 2015; **373**: 795-807.

199. Blankson JN, Persaud D, Siliciano RF. The challenge of viral reservoirs in HIV-1 infection. *Annu Rev Med* 2002; **53**: 557-93.

200. Ananworanich J, Schuetz A, Vandergeeten C et al. Impact of multi-targeted antiretroviral treatment on gut T cell depletion and HIV reservoir seeding during acute HIV infection. *PLoS One* 2012; **7**: e33948.

201. Moore ZS, McCoy S, Kuruc J et al. Number of named partners and number of partners newly diagnosed with HIV infection identified by persons with acute versus established HIV infection. *J Acquir Immune Defic Syndr* 2009; **52**: 509-13.

202. Mellors JW, Muñoz A, Giorgi JV et al. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. *Ann Intern Med* 1997; **126**: 946-54.

203. Mellors JW, Rinaldo CR, Jr., Gupta P et al. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* 1996; **272**: 1167-70.

204. Walker BD. Elite control of HIV Infection: implications for vaccines and treatment. *Top HIV Med* 2007; **15**: 134-6.

205. Deacon NJ, Tsykin A, Solomon A et al. Genomic structure of an attenuated quasi species of HIV-1 from a blood transfusion donor and recipients. *Science* 1995; **270**: 988-91.

206. Migueles SA, Sabbaghian MS, Shupert WL et al. HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proc Natl Acad Sci U S A* 2000; **97**: 2709-14.

207. Delta: a randomised double-blind controlled trial comparing combinations of zidovudine plus didanosine or zalcitabine with zidovudine alone in HIV-infected individuals. Delta Coordinating Committee. *Lancet* 1996; **348**: 283-91.

208. WHO. Consolidated Guidelines on HIV Prevention, Testing, Treatment, Service Delivery and Monitoring: Recommendations for a Public HealthApproach. https://www.who.int/publications/i/item/9789240031593. 2021. (Accessed 18 June 2022).

209. Njom Nlend AE. Mother-to-Child Transmission of HIV Through Breastfeeding Improving Awareness and Education: A Short Narrative Review. *Int J Womens Health* 2022; **14**: 697-703.

210. UNAIDS. '15 by 15': a global target achieved. 2015. http://www.unaids.org/sites/default/files/media_asset/UNAIDS_15by15_ en.pdf (Accessed 01 May 2022).

211. WHO. Policy brief: update of recommendations on first- and second-line antiretroviral regimens. https://apps.who.int/iris/handle/10665/325892. 2019. (Accessed 15 December 2021).

212. Woollard SM, Kanmogne GD. Maraviroc: a review of its use in HIV infection and beyond. *Drug Des Devel Ther* 2015; **9**: 5447-68.

213. Zhang D, Li W, Jiang S. Peptide fusion inhibitors targeting the HIV-1 gp41: a patent review (2009 - 2014). *Expert Opin Ther Pat* 2015; **25**: 159-73.

214. Collier DA, Monit C, Gupta RK. The Impact of HIV-1 Drug Escape on the Global Treatment Landscape. *Cell Host Microbe* 2019; **26**: 48-60.

215. Tang MW, Shafer RW. HIV-1 antiretroviral resistance: scientific principles and clinical applications. *Drugs* 2012; **72**: e1-25.

216. Holec AD, Mandal S, Prathipati PK et al. Nucleotide Reverse Transcriptase Inhibitors: A Thorough Review, Present Status and Future Perspective as HIV Therapeutics. *Curr HIV Res* 2017; **15**: 411-21.

217. Clutter DS, Jordan MR, Bertagnolio S et al. HIV-1 drug resistance and resistance testing. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2016; **46**: 292-307.

218. Joly V, Yeni P. [Non-nucleoside reverse transcriptase inhibitors]. *Ann Med Interne (Paris)* 2000; **151**: 260-7.

219. de Béthune MP. Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: a review of the last 20 years (1989-2009). *Antiviral Res* 2010; **85**: 75-90.

220. Sluis-Cremer N, Tachedjian G. Mechanisms of inhibition of HIV replication by non-nucleoside reverse transcriptase inhibitors. *Virus Res* 2008; **134**: 147-56.

221. Lai MT, Munshi V, Lu M et al. Mechanistic Study of Common Non-Nucleoside Reverse Transcriptase Inhibitor-Resistant Mutations with K103N and Y181C Substitutions. *Viruses* 2016; **8**.

222. Lai MT, Feng M, Falgueyret JP et al. In vitro characterization of MK-1439, a novel HIV-1 nonnucleoside reverse transcriptase inhibitor. *Antimicrob Agents Chemother* 2014; **58**: 1652-63.

223. FDA. FDA-Approved HIV Medicines. https://hivinfo.nih.gov/understanding-hiv/fact-sheets/fda-approved-hiv-medicines. 2021. (Assessed13 December, 2021).

224. Mbhele N, Chimukangara B, Gordon M. HIV-1 integrase strand transfer inhibitors: a review of current drugs, recent advances and drug resistance. *Int J Antimicrob Agents* 2021; **57**: 106343.

225. Dow DE, Bartlett JA. Dolutegravir, the Second-Generation of Integrase Strand Transfer Inhibitors (INSTIs) for the Treatment of HIV. *Infect Dis Ther* 2014; **3**: 83-102.

226. Boffito M, Waters L, Cahn P et al. Perspectives on the Barrier to Resistance for Dolutegravir + Lamivudine, a Two-Drug Antiretroviral Therapy for HIV-1 Infection. *AIDS Res Hum Retroviruses* 2020; **36**: 13-8.

227. Hightower KE, Wang R, Deanda F et al. Dolutegravir (S/GSK1349572) exhibits significantly slower dissociation than raltegravir and elvitegravir from wild-type and integrase inhibitor-resistant HIV-1 integrase-DNA complexes. *Antimicrob Agents Chemother* 2011; **55**: 4552-9.

228. Kobayashi M, Yoshinaga T, Seki T et al. In Vitro antiretroviral properties of S/GSK1349572, a next-generation HIV integrase inhibitor. *Antimicrob Agents Chemother* 2011; **55**: 813-21.

229. Iwamoto M, Wenning LA, Petry AS et al. Safety, tolerability, and pharmacokinetics of raltegravir after single and multiple doses in healthy subjects. *Clin Pharmacol Ther* 2008; **83**: 293-9.

230. Yang LL, Li Q, Zhou LB et al. Meta-analysis and systematic review of the efficacy and resistance for human immunodeficiency virus type 1 integrase strand transfer inhibitors. *Int J Antimicrob Agents* 2019; **54**: 547-55.

231. Lepik KJ, Harrigan PR, Yip B et al. Emergent drug resistance with integrase strand transfer inhibitor-based regimens. *Aids* 2017; **31**: 1425-34.

232. McGee KS, Okeke NL, Hurt CB et al. Canary in the Coal Mine? Transmitted Mutations Conferring Resistance to All Integrase Strand Transfer Inhibitors in a Treatment-Naive Patient. *Open Forum Infect Dis* 2018; **5**: ofy294.
233. Oliveira M, Ibanescu RI, Anstett K et al. Selective resistance profiles emerging in patient-derived clinical isolates with cabotegravir, bictegravir, dolutegravir, and elvitegravir. *Retrovirology* 2018; **15**: 56.
234. Young B, Fransen S, Greenberg KS et al. Transmission of integrase strand-transfer inhibitor multidrug-resistant HIV-1: case report and response to raltegravir-containing antiretroviral therapy. *Antivir Ther* 2011; **16**: 253-6.
235. Arimide DA, Szojka ZI, Zealiyas K et al. Pre-Treatment Integrase Inhibitor Resistance and Natural Polymorphisms among HIV-1 Subtype C Infected Patients in Ethiopia. *Viruses* 2022; **14**: 729.
236. Dooley KE, Kaplan R, Mwelase N et al. Dolutegravir-based Antiretroviral Therapy for Patients Coinfected With Tuberculosis and Human Immunodeficiency Virus: A Multicenter, Noncomparative, Open-label, Randomized Trial. *Clin Infect Dis* 2020; **70**: 549-56.
237. Wensing AM, van Maarseveen NM, Nijhuis M. Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Antiviral Res* 2010; **85**: 59-74.
238. Lai Y-T. Small Molecule HIV-1 Attachment Inhibitors: Discovery, Mode of Action and Structural Basis of Inhibition. *Viruses* 2021; **13**: 843.
239. MOH. Guidelines for use of antiretroviral drugs in Ethiopia, Ministry of Health Disease Prevention and Control Department in collaboration with HIV/ AIDS Prevention and Control office (HAPCO) and Drug Administration and Control Authority (DACA), February 2003 Addis Ababa, Ethiopia. 2003.
240. HAPCO. Report on progress towards implementation of the UN Declaration of Commitment on HIV/AIDS, Federal Democratic Republic of Ethiopia, Federal HIV/AIDS Prevention and Control Office, Addis Ababa, Ethiopia, March 2010. 2010.
241. MOH. National Guideline for comprhensive HIV prevention, care and treatment. 2022.
242. MOH. National consolidated guidelines for comprehensive HIV prevention, care and treatment. 2018. https://www.afro.who.int/sites/default/files/2019-04/National%20Comprehensive%20HIV%20Care%20%20Guideline%202018.pdf. (Accessed 14 December 2021).
243. Abegaz WE, Grossman Z, Wolday D et al. Threshold survey evaluating transmitted HIV drug resistance among public antenatal clinic clients in Addis Ababa, Ethiopia. *Antivir Ther* 2008; **13 Suppl 2**: 89-94.
244. Kassu A, Fujino M, Matsuda M et al. Molecular epidemiology of HIV type 1 in treatment-naive patients in north Ethiopia. *AIDS Res Hum Retroviruses* 2007; **23**: 564-8.
245. Mulu A, Lange T, Liebert UG et al. Clade homogeneity and Pol gene polymorphisms in chronically HIV-1 infected antiretroviral treatment naive patients after the roll out of ART in Ethiopia. *BMC Infect Dis* 2014; **14**: 158.

246. Telele NF, Kalu AW, Gebre-Selassie S et al. Pretreatment drug resistance in a large countrywide Ethiopian HIV-1C cohort: a comparison of Sanger and high-throughput sequencing. *Scientific reports* 2018; **8**: 7556.

247. Tadesse BT, Tsai, O., Chala, A., Chaka, T. E., Eromo, T., Lapointe, H. R., Baraki, B., Shahid, A., Tadesse, S., Makonnen, E., Brumme, Z. L., Aklillu, E., & Brumme, C. J. Prevalence and Correlates of Pre-Treatment HIV Drug Resistance among HIV-Infected Children in Ethiopia. *Viruses* 2019; **11**: 877.

248. WHO. HIV Drug Resistance Report 2019. Geneva, Switzerland (WHO/CDS/HIV/19.21).2019. (Assessed Date April 2022).

249. Mulu A MM, Liebert UG. Upward trends of acquired drug resistances in Ethiopian HIV-1C isolates: A decade longitudinal study. *PLoS One* 2017: 12-22.

250. Reepalu A, Arimide DA, Balcha TT et al. Drug resistance in HIV-positive adults during the initial year of antiretroviral treatment at Ethiopian health centers. *Open Forum Infectious Diseases* 2021.

251. Telele NF, Kalu AW, Gebre-Selassie S et al. A viral genome wide association study and genotypic resistance testing in patients failing first line antiretroviral therapy in the first large countrywide Ethiopian HIV cohort. *BMC Infect Dis* 2019; **19**: 569.

252. Tufa TB, Fuchs A, Orth HM et al. Characterization of HIV-1 drug resistance among patients with failure of second-line combined antiretroviral therapy in central Ethiopia. *HIV medicine* 2022; **23**: 159-68.

253. Beyrer C, Pozniak A. HIV Drug Resistance - An Emerging Threat to Epidemic Control. *N Engl J Med* 2017; **377**: 1605-7.

254. Gupta RK, Jordan MR, Sultan BJ et al. Global trends in antiretroviral resistance in treatment-naive individuals with HIV after rollout of antiretroviral treatment in resource-limited settings: a global collaborative study and meta-regression analysis. *Lancet* 2012; **380**: 1250-8.

255. Cambiano V, Bertagnolio S, Jordan MR et al. Transmission of drug resistant HIV and its potential impact on mortality and treatment outcomes in resource-limited settings. *J Infect Dis* 2013; **207 Suppl 2**: S57-62.

256. Paredes R, Clotet B. Clinical management of HIV-1 resistance. *Antiviral Res* 2010; **85**: 245-65.

257. Quiñones-Mateu ME, Arts EJ. Fitness of drug resistant HIV-1: methodology and clinical implications. *Drug Resist Updat* 2002; **5**: 224-33.

258. Le T, Chiarella J, Simen BB et al. Low-abundance HIV drug-resistant viral variants in treatment-experienced persons correlate with historical antiretroviral use. *PLoS One* 2009; **4**: e6079.

259. Metzner KJ, Bonhoeffer S, Fischer M et al. Emergence of minor populations of human immunodeficiency virus type 1 carrying the M184V and L90M mutations in subjects undergoing structured treatment interruptions. *J Infect Dis* 2003; **188**: 1433-43.

260. Bertagnolio S, Hermans L, Jordan MR et al. Clinical Impact of Pretreatment Human Immunodeficiency Virus Drug Resistance in People Initiating Nonnucleoside Reverse Transcriptase Inhibitor–Containing Antiretroviral Therapy: A Systematic Review and Meta-analysis. *The Journal of Infectious Diseases* 2020; **224**: 377-88.

261. Gupta RK, Gregson J, Parkin N et al. HIV-1 drug resistance before initiation or re-initiation of first-line antiretroviral therapy in low-income and middle-income countries: a systematic review and meta-regression analysis. *The Lancet Infectious Diseases* 2018; **18**: 346-55.
262. Rhee SY, Blanco JL, Jordan MR et al. Geographic and temporal trends in the molecular epidemiology and genetic mechanisms of transmitted HIV-1 drug resistance: an individual-patient- and sequence-level meta-analysis. *PLoS Med* 2015; **12**: e1001810.
263. Bailey AJ, Rhee SY, Shafer RW. Integrase Strand Transfer Inhibitor Resistance in Integrase Strand Transfer Inhibitor-Naive Persons. *AIDS Res Hum Retroviruses* 2021.
264. Casadellà M, Santos JR, Noguera-Julian M et al. Primary resistance to integrase strand transfer inhibitors in Spain using ultrasensitive HIV-1 genotyping. *J Antimicrob Chemother* 2020; **75**: 3517-24.
265. Seatla KK, Maruapula D, Choga WT et al. HIV-1 Subtype C Drug Resistance Mutations in Heavily Treated Patients Failing Integrase Strand Transfer Inhibitor-Based Regimens in Botswana. *Viruses* 2021; **13**.
266. Blanco JL, Rojas J, Paredes R et al. Dolutegravir-based maintenance monotherapy versus dual therapy with lamivudine: a planned 24 week analysis of the DOLAM randomized clinical trial. *Journal of Antimicrobial Chemotherapy* 2018; **73**: 1965-71.
267. Lübke N, Jensen B, Hüttig F et al. Failure of Dolutegravir First-Line ART with Selection of Virus Carrying R263K and G118R. *N Engl J Med* 2019; **381**: 887-9.
268. Ndashimye E, Avino M, Olabode AS et al. Accumulation of integrase strand transfer inhibitor resistance mutations confers high-level resistance to dolutegravir in non-B subtype HIV-1 strains from patients failing raltegravir in Uganda. *J Antimicrob Chemother* 2020; **75**: 3525-33.
269. McCluskey SM, Siedner MJ, Marconi VC. Management of Virologic Failure and HIV Drug Resistance. *Infect Dis Clin North Am* 2019; **33**: 707-42.
270. Tang MW, Liu TF, Shafer RW. The HIVdb system for HIV-1 genotypic resistance interpretation. *Intervirology* 2012; **55**: 98-101.
271. Manyana S, Gounder L, Pillay M et al. HIV-1 Drug Resistance Genotyping in Resource Limited Settings: Current and Future Perspectives in Sequencing Technologies. *Viruses* 2021; **13**.
272. Chimukangara B, Samuel R, Naidoo K et al. Primary HIV-1 Drug Resistant Minority Variants. *AIDS Rev* 2017; **19**: 89-96.
273. Lee ER, Parkin N, Jennings C et al. Performance comparison of next generation sequencing analysis pipelines for HIV-1 drug resistance testing. *Scientific reports* 2020; **10**: 1634.
274. Chimukangara B, Giandhari J, Lessells R et al. Impact of pretreatment low-abundance HIV-1 drug-resistant variants on virological failure among HIV-1/TB-co-infected individuals. *J Antimicrob Chemother* 2020; **75**: 3319-26.
275. Mbunkah HA, Bertagnolio S, Hamers RL et al. Low-Abundance Drug-Resistant HIV-1 Variants in Antiretroviral Drug-Naive Individuals: A Systematic Review of

Detection Methods, Prevalence, and Clinical Impact. *J Infect Dis* 2020; **221**: 1584-97.

276. Capina R, Li K, Kearney L et al. Quality Control of Next-Generation Sequencing-Based HIV-1 Drug Resistance Data in Clinical Laboratory Information Systems Framework. *Viruses* 2020; **12**.
277. Gunthard HF, Calvez V, Paredes R et al. Human Immunodeficiency Virus Drug Resistance: 2018 Recommendations of the International Antiviral Society-USA Panel. *Clin Infect Dis* 2019; **68**: 177-87.
278. Ávila-Ríos S, Parkin N, Swanstrom R et al. Next-Generation Sequencing for HIV Drug Resistance Testing: Laboratory, Clinical, and Implementation Considerations. *Viruses* 2020; **12**.
279. Petropoulos CJ, Parkin NT, Limoli KL et al. A novel phenotypic drug susceptibility assay for human immunodeficiency virus type 1. *Antimicrob Agents Chemother* 2000; **44**: 920-8.
280. Nascimento FF, Reis MD, Yang Z. A biologist's guide to Bayesian phylogenetic analysis. *Nat Ecol Evol* 2017; **1**: 1446-54.
281. Baele G, Suchard MA, Rambaut A et al. Emerging Concepts of Data Integration in Pathogen Phylodynamics. *Syst Biol* 2017; **66**: e47-e65.
282. Vandamme PLMSA-M. The Phylogenetic Handbook: A practical approach to DNA and protein phylogeny. *Cambridge University Press Cambridge, United Kingdom* 2009.
283. Fitch WM, Margoliash E. Construction of phylogenetic trees. *Science* 1967; **155**: 279-84.
284. Sokal RR. A statistical method for evaluating systematic relationships. *Univ Kansas, Sci Bull* 1958; **38**: 1409-38.
285. Rzhetsky A, Nei M. Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Mol Biol Evol* 1993; **10**: 1073-95.
286. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987; **4**: 406-25.
287. Felsenstein J. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am J Hum Genet* 1973; **25**: 471-92.
288. Rannala B, Yang Z. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J Mol Evol* 1996; **43**: 304-11.
289. Mau B, Newton MA, Larget B. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Biometrics* 1999; **55**: 1-12.
290. Gascuel O, Steel M. Neighbor-joining revealed. *Mol Biol Evol* 2006; **23**: 1997-2000.
291. Kannan L, Wheeler WC. Maximum Parsimony on Phylogenetic networks. *Algorithms Mol Biol* 2012; **7**: 9.
292. Holder M, Lewis PO. Phylogeny estimation: traditional and Bayesian approaches. *Nat Rev Genet* 2003; **4**: 275-84.
293. Schulmeister S. Inconsistency of maximum parsimony revisited. *Syst Biol* 2004; **53**: 521-8.

294. Page RD, Holmes EC. *Molecular evolution: a phylogenetic approach*: John Wiley & Sons, 2009.

295. Huelsenbeck JP, Ronquist F, Nielsen R et al. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 2001; **294**: 2310-4.

296. Kuhner MK, Yamato J, Felsenstein J. Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics* 1995; **140**: 1421-30.

297. Donoghue PC, Yang Z. The evolution of methods for establishing evolutionary timescales. *Philos Trans R Soc Lond B Biol Sci* 2016; **371**.

298. Sallam M, Esbjörnsson J, Baldvinsdóttir G et al. Molecular epidemiology of HIV-1 in Iceland: Early introductions, transmission dynamics and recent outbreaks among injection drug users. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2017; **49**: 157-63.

299. Drummond AJ, Rambaut A, Shapiro B et al. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 2005; **22**: 1185-92.

300. Pybus OG, Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet* 2009; **10**: 540-50.

301. Lemey P, Rambaut A, Drummond AJ et al. Bayesian phylogeography finds its roots. *PLoS Comput Biol* 2009; **5**: e1000520.

302. Nduva GM, Otieno F, Kimani J et al. Phylogeographic Assessment Reveals Geographic Sources of HIV-1 Dissemination Among Men Who Have Sex With Men in Kenya. *Front Microbiol* 2022; **13**: 843330.

303. Nduva GM, Nazziwa J, Hassan AS et al. The Role of Phylogenetics in Discerning HIV-1 Mixing among Vulnerable Populations and Geographic Regions in Sub-Saharan Africa: A Systematic Review. *Viruses* 2021; **13**.

304. Nascimento FF, Gongora J, Charleston M et al. Evolution of endogenous retroviruses in the Suidae: evidence for different viral subpopulations in African and Eurasian host species. *BMC Evol Biol* 2011; **11**: 139.

305. Jarvis ED, Mirarab S, Aberer AJ et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 2014; **346**: 1320-31.

306. Misof B, Liu S, Meusemann K et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 2014; **346**: 763-7.

307. Raymann K, Brochier-Armanet C, Gribaldo S. The two-domain tree of life is linked to a new root for the Archaea. *Proc Natl Acad Sci U S A* 2015; **112**: 6670-5.

308. Bromham L, Duchêne S, Hua X et al. Bayesian molecular dating: opening up the black box. *Biol Rev Camb Philos Soc* 2018; **93**: 1165-91.

309. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 1970; **57**: 97-109.

310. Metropolis N, Rosenbluth AW, Rosenbluth MN et al. Equation of state calculations by fast computing machines. *The journal of chemical physics* 1953; **21**: 1087-92.

311. Tierney L. Markov chains for exploring posterior distributions. *the Annals of Statistics* 1994: 1701-28.

312. Esbjörnsson J. *HIV-1 evolution, disease progression and molecular epidemiology of HIV-1 single and HIV-1 and HIV-2 dual-infected individuals in Guinea-Bissau*: Lund University, 2010.

313. Patiño-Galindo J, González-Candelas F. Molecular evolution methods to study HIV-1 epidemics. *Future Virol* 2018; **13**: 399-404.

314. Som A. Causes, consequences and solutions of phylogenetic incongruence. *Brief Bioinform* 2015; **16**: 536-48.

315. Zwickl DJ, Hillis DM. Increased taxon sampling greatly reduces phylogenetic error. *Syst Biol* 2002; **51**: 588-98.

316. Hillis DM. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Syst Biol* 1998; **47**: 3-8.

317. Pick KS, Philippe H, Schreiber F et al. Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. *Mol Biol Evol* 2010; **27**: 1983-7.

318. Schierup MH, Hein J. Consequences of recombination on traditional phylogenetic analysis. *Genetics* 2000; **156**: 879-91.

319. Posada D, Crandall KA. The effect of recombination on the accuracy of phylogeny estimation. *J Mol Evol* 2002; **54**: 396-402.

320. Hué S, Clewley JP, Cane PA et al. HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. *Aids* 2004; **18**: 719-28.

321. Yebra G, Hodcroft EB, Ragonnet-Cronin ML et al. Using nearly full-genome HIV sequence data improves phylogeny reconstruction in a simulated epidemic. *Scientific reports* 2016; **6**: 39489.

322. Han X, Zhao B, An M et al. Molecular network-based intervention brings us closer to ending the HIV pandemic. *Front Med* 2020; **14**: 136-48.

323. Altschul SF, Gish W, Miller W et al. Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403-10.

324. Wong KM, Suchard MA, Huelsenbeck JP. Alignment uncertainty and genomic analysis. *Science* 2008; **319**: 473-6.

325. Morrison DA. Why would phylogeneticists ignore computerized sequence alignment? *Syst Biol* 2009; **58**: 150-8.

326. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 1999; **41 (1999), pp. 95-98**.

327. Thompson JD, Gibson TJ, Plewniak F et al. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997; **25**: 4876-82.

328. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004; **32**: 1792-7.

329. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013; **30**: 772-80.

330. Arenas M. Trends in substitution models of molecular evolution. *Front Genet* 2015; **6**: 319.

331. Vartanian JP, Henry M, Wain-Hobson S. Sustained G-->A hypermutation during reverse transcription of an entire human immunodeficiency virus type 1 strain Vau group O genome. *J Gen Virol* 2002; **83**: 801-5.

332. Rawson JM, Landman SR, Reilly CS et al. HIV-1 and HIV-2 exhibit similar mutation frequencies and spectra in the absence of G-to-A hypermutation. *Retrovirology* 2015; **12**: 60.

333. Jukes TH, Cantor CR. Evolution of protein molecules. *Mammalian protein metabolism* 1969; **3**: 21-132.

334. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 1980; **16**: 111-20.

335. Collins DW, Jukes TH. Rates of transition and transversion in coding sequences since the human-rodent divergence. *Genomics* 1994; **20**: 386-96.

336. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 1981; **17**: 368-76.

337. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 1985; **22**: 160-74.

338. Yang Z. Estimating the pattern of nucleotide substitution. *J Mol Evol* 1994; **39**: 105-11.

339. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on mathematics in the life sciences* 1986; **17**: 57-86.

340. Lanave C, Preparata G, Saccone C et al. A new method for calculating evolutionary substitution rates. *J Mol Evol* 1984; **20**: 86-93.

341. Geller R, Domingo-Calap P, Cuevas JM et al. The external domains of the HIV-1 envelope are a mutational cold spot. *Nat Commun* 2015; **6**: 8571.

342. Posada D, Crandall KA. Selecting the best-fit model of nucleotide substitution. *Syst Biol* 2001; **50**: 580-601.

343. Minin V, Abdo Z, Joyce P et al. Performance-based selection of likelihood models for phylogeny estimation. *Syst Biol* 2003; **52**: 674-83.

344. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 1998; **14**: 817-8.

345. Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A* 2001; **98**: 13757-62.

346. Posada D, Crandall KA. Selecting models of nucleotide substitution: an application to human immunodeficiency virus 1 (HIV-1). *Mol Biol Evol* 2001; **18**: 897-906.

347. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 1985; **39**: 783-91.

348. Hassan AS, Pybus OG, Sanders EJ et al. Defining HIV-1 transmission clusters based on sequence data. *Aids* 2017; **31**: 1211-22.

349. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* 2006; **55**: 539-52.

350. Nguyen L-T, Schmidt HA, von Haeseler A et al. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* 2014; **32**: 268-74.
351. Guindon S, Lethiec F, Duroux P et al. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* 2005; **33**: W557-9.
352. Grenfell BT, Pybus OG, Gog JR et al. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 2004; **303**: 327-32.
353. Holmes EC. Evolutionary history and phylogeography of human viruses. *Annu Rev Microbiol* 2008; **62**: 307-28.
354. Kitchen A, Miyamoto MM, Mulligan CJ. Utility of DNA viruses for studying human host history: case study of JC virus. *Mol Phylogenet Evol* 2008; **46**: 673-82.
355. Magiorkinis G, Magiorkinis E, Paraskevis D et al. The global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis. *PLoS Med* 2009; **6**: e1000198.
356. Nelson MI, Holmes EC. The evolution of epidemic influenza. *Nat Rev Genet* 2007; **8**: 196-205.
357. Kühnert D, Wu C-H, Drummond AJ. Phylogenetic and epidemic modeling of rapidly evolving infectious diseases. *Infection, Genetics and Evolution* 2011; **11**: 1825-41.
358. Nduva GM, Otieno F, Kimani J et al. Quantifying rates of HIV-1 flow between risk groups and geographic locations in Kenya: A country-wide phylogenetic study. *Virus Evol* 2022; **8**: veac016.
359. Gill MS, Lemey P, Faria NR et al. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol Biol Evol* 2013; **30**: 713-24.
360. Volz EM, Kosakovsky Pond SL, Ward MJ et al. Phylodynamics of infectious disease epidemics. *Genetics* 2009; **183**: 1421-30.
361. Lemey P, Rambaut A, Pybus OG. HIV evolutionary dynamics within and among hosts. *AIDS Rev* 2006; **8**: 125-40.
362. de Silva E, Ferguson NM, Fraser C. Inferring pandemic growth rates from sequence data. *J R Soc Interface* 2012; **9**: 1797-808.
363. Lewis F, Hughes GJ, Rambaut A et al. Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med* 2008; **5**: e50.
364. Wright S. Evolution in Mendelian Populations. *Genetics* 1931; **16**: 97-159.
365. Fisher RA. *The genetical theory of natural selection*: Рипол Классик, 1958.
366. Griffiths RC, Tavaré S. Sampling theory for neutral alleles in a varying environment. *Philos Trans R Soc Lond B Biol Sci* 1994; **344**: 403-10.
367. Dearlove B, Wilson DJ. Coalescent inference for infectious disease: meta-analysis of hepatitis C. *Philos Trans R Soc Lond B Biol Sci* 2013; **368**: 20120314.
368. Rasmussen DA, Volz EM, Koelle K. Phylodynamic inference for structured epidemiological models. *PLoS Comput Biol* 2014; **10**: e1003570.

369. Rouzine IM, Rodrigo A, Coffin JM. Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology. *Microbiol Mol Biol Rev* 2001; **65**: 151-85.

370. Stadler T, Kuhnert D, Bonhoeffer S et al. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences* 2013; **110**: 228-33.

371. Drummond A, Pybus OG, Rambaut A. Inference of viral evolutionary rates from molecular sequences. *Adv Parasitol* 2003; **54**: 331-58.

372. Stadler T, Kühnert D, Bonhoeffer S et al. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc Natl Acad Sci U S A* 2013; **110**: 228-33.

373. Gao F, Yue L, Robertson DL et al. Genetic diversity of human immunodeficiency virus type 2: evidence for distinct sequence subtypes with differences in virus biology. *J Virol* 1994; **68**: 7433-47.

374. Wertheim JO, Leigh Brown AJ, Hepler NL et al. The global transmission network of HIV-1. *J Infect Dis* 2014; **209**: 304-13.

375. Díez-Fuertes F, Cabello M, Thomson MM. Bayesian phylogeographic analyses clarify the origin of the HIV-1 subtype A variant circulating in former Soviet Union's countries. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2015; **33**: 197-205.

376. Cabello M, Mendoza Y, Bello G. Spatiotemporal dynamics of dissemination of non-pandemic HIV-1 subtype B clades in the Caribbean region. *PLoS One* 2014; **9**: e106045.

377. Delatorre E, Velasco-De-Castro CA, Pilotto JH et al. Short Communication: Reassessing the Origin of the HIV-1 CRF02_AG Lineages Circulating in Brazil. *AIDS Res Hum Retroviruses* 2015; **31**: 1230-7.

378. de Oliveira T, Pybus OG, Rambaut A et al. Molecular epidemiology: HIV-1 and HCV sequences from Libyan outbreak. *Nature* 2006; **444**: 836-7.

379. de Oliveira T, Kharsany AB, Graf T et al. Transmission networks and risk of HIV infection in KwaZulu-Natal, South Africa: a community-wide phylogenetic study. *Lancet HIV* 2017; **4**: e41-e50.

380. Bbosa N, Ssemwanga D, Nsubuga RN et al. Phylogeography of HIV-1 suggests that Ugandan fishing communities are a sink for, not a source of, virus from general populations. *Scientific reports* 2019; **9**: 1051.

381. Stecher M, Hoenigl M, Eis-Hübinger AM et al. Hotspots of Transmission Driving the Local Human Immunodeficiency Virus Epidemic in the Cologne-Bonn Region, Germany. *Clin Infect Dis* 2019; **68**: 1539-46.

382. Ragonnet-Cronin M, Hodcroft E, Hué S et al. Automated analysis of phylogenetic clusters. *BMC Bioinformatics* 2013; **14**: 317.

383. Vrbik I, Stephens DA, Roger M et al. The Gap Procedure: for the identification of phylogenetic clusters in HIV-1 sequence data. *BMC Bioinformatics* 2015; **16**: 355.

384. Kosakovsky Pond SL, Weaver S, Leigh Brown AJ et al. HIV-TRACE (TRAnsmission Cluster Engine): a Tool for Large Scale Molecular Epidemiology of HIV-1 and Other Rapidly Evolving Pathogens. *Mol Biol Evol* 2018; **35**: 1812-9.

385. Oster AM, Wertheim JO, Hernandez AL et al. Using Molecular HIV Surveillance Data to Understand Transmission Between Subpopulations in the United States. *J Acquir Immune Defic Syndr* 2015; **70**: 444-51.

386. Paraskevis D, Nikolopoulos GK, Magiorkinis G et al. The application of HIV molecular epidemiology to public health. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2016; **46**: 159-68.

387. Stadler T, Kouyos R, von Wyl V et al. Estimating the basic reproductive number from viral sequence data. *Mol Biol Evol* 2012; **29**: 347-57.

388. Novitsky V, Kuhnert D, Moyo S et al. Phylodynamic analysis of HIV sub-epidemics in Mochudi, Botswana. *Epidemics* 2015; **13**: 44-55.

389. Stadler T, Bonhoeffer S. Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. *Philos Trans R Soc Lond B Biol Sci* 2013; **368**: 20120198.

390. Eshleman SH, Hudelson SE, Redd AD et al. Analysis of genetic linkage of HIV from couples enrolled in the HIV Prevention Trials Network 052 trial. *J Infect Dis* 2011; **204**: 1918-26.

391. Rodger A, Bruun T, Cambiano V et al. HIV transmission risk through condomless sex if HIV+ partner on suppressive ART: PARTNER study. *Conference on Retroviruses and Opportunistic Infections*, p. 3-6.

392. Villabona-Arenas CJ, Hanage WP, Tully DC. Phylogenetic interpretation during outbreaks requires caution. *Nat Microbiol* 2020; **5**: 876-7.

393. Abidi SH, Nduva GM, Siddiqui D et al. Phylogenetic and Drug-Resistance Analysis of HIV-1 Sequences From an Extensive Paediatric HIV-1 Outbreak in Larkana, Pakistan. *Front Microbiol* 2021; **12**: 658186.

394. Hatzakis A, Sypsa V, Paraskevis D et al. Design and baseline findings of a large-scale rapid response to an HIV outbreak in people who inject drugs in Athens, Greece: the ARISTOTLE programme. *Addiction* 2015; **110**: 1453-67.

395. Jordan MR, Bennett DE, Bertagnolio S et al. World Health Organization surveys to monitor HIV drug resistance prevention and associated factors in sentinel antiretroviral treatment sites. *Antivir Ther* 2008; **13 Suppl 2**: 15-23.

396. Bennett DE, Bertagnolio S, Sutherland D et al. The World Health Organization's global strategy for prevention and assessment of HIV drug resistance. *Antivir Ther* 2008; **13 Suppl 2**: 1-13.

397. Balcha TT, Sturegård E, Winqvist N et al. Intensified tuberculosis case-finding in HIV-positive adults managed at Ethiopian health centers: diagnostic yield of Xpert MTB/RIF compared with smear microscopy and liquid culture. *PLoS One* 2014; **9**: e85478.

398. Rosemary A, Chika O, Jonathan O et al. Genotyping performance evaluation of commercially available HIV-1 drug resistance test. *PLoS One* 2018; **13**: e0198246.

399. Woods CK, Brumme CJ, Liu TF et al. Automating HIV drug resistance genotyping with RECall, a freely accessible sequence analysis tool. *J Clin Microbiol* 2012; **50**: 1936-42.

400. Van Laethem K, Schrooten Y, Covens K et al. A genotypic assay for the amplification and sequencing of integrase from diverse HIV-1 group M subtypes. *J Virol Methods* 2008; **153**: 176-81.

401. Pineda-Peña AC, Faria NR, Imbrechts S et al. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: performance evaluation of the new REGA version 3 and seven other tools. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2013; **19**: 337-48.

402. Struck D, Lawyer G, Ternes AM et al. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. *Nucleic Acids Res* 2014; **42**: e144.

403. Martin DP, Lemey P, Lott M et al. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 2010; **26**: 2462-3.

404. Schultz AK, Zhang M, Bulla I et al. jpHMM: improving the reliability of recombination prediction in HIV-1. *Nucleic Acids Res* 2009; **37**: W647-51.

405. Arimide DA, Abebe A, Kebede Y et al. HIV-genetic diversity and drug resistance transmission clusters in Gondar, Northern Ethiopia, 2003-2013. *PLoS One* 2018; **13**: e0205446.

406. Lole KS, Bollinger RC, Paranjape RS et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 1999; **73**: 152-60.

407. Abebe A, Kuiken CL, Goudsmit J et al. HIV type 1 subtype C in Addis Ababa, Ethiopia. *AIDS Res Hum Retroviruses* 1997; **13**: 1071-5.

408. Abebe A, Pollakis G, Fontanet AL et al. Identification of a genetic subcluster of HIV type 1 subtype C (C') widespread in Ethiopia. *AIDS Res Hum Retroviruses* 2000; **16**: 1909-14.

409. Pollakis G, Abebe A, Kliphuis A et al. Recombination of HIV type 1C (C'/C") in Ethiopia: possible link of EthHIV-1C' to subtype C sequences from the high-prevalence epidemics in India and Southern Africa. *AIDS Res Hum Retroviruses* 2003; **19**: 999-1008.

410. Delatorre EO, Bello G. Phylodynamics of HIV-1 subtype C epidemic in east Africa. *PLoS One* 2012; **7**: e41904.

411. Thomson MM, Fernandez-Garcia A. Phylogenetic structure in African HIV-1 subtype C revealed by selective sequential pruning. *Virology* 2011; **415**: 30-8.

412. Bruen TC, Philippe H, Bryant D. A simple and robust statistical test for detecting the presence of recombination. *Genetics* 2006; **172**: 2665-81.

413. Brauer MJ, Holder MT, Dries LA et al. Genetic algorithms and parallel processing in maximum-likelihood phylogeny inference. *Mol Biol Evol* 2002; **19**: 1717-26.

414. Guindon S, Dufayard JF, Lefort V et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010; **59**: 307-21.

415. Anisimova M, Gil M, Dufayard JF et al. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol* 2011; **60**: 685-99.

416. Schultz AK, Zhang M, Leitner T et al. A jumping profile Hidden Markov Model and applications to recombination sites in HIV and HCV genomes. *BMC Bioinformatics* 2006; **7**: 265.

417. Rambaut A, Lam TT, Max Carvalho L et al. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2016; **2**: vew007.

418. Bennett DE, Myatt M, Bertagnolio S et al. Recommendations for surveillance of transmitted HIV drug resistance in countries scaling up antiretroviral treatment. *Antivir Ther* 2008; **13 Suppl 2**: 25-36.

419. Rambaut A. FigTree v1.4.3: Tree Figure Drawing Tool. http://tree.bio.ed.ac.uk/software/figtree/. 2016. (Accessed 04 October 2019).

420. Parker J, Rambaut A, Pybus OG. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2008; **8**: 239-46.

421. Aldous JL, Pond SK, Poon A et al. Characterizing HIV transmission networks across the United States. *Clin Infect Dis* 2012; **55**: 1135-43.

422. Wertheim JO, Oster AM, Johnson JA et al. Transmission fitness of drug-resistant HIV revealed in a surveillance system transmission network. *Virus Evol* 2017; **3**: vex008.

423. Suchard MA, Lemey P, Baele G et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 2018; **4**: vey016.

424. Hill V, Baele G. Bayesian estimation of past population dynamics in BEAST 1.10 using the Skygrid coalescent model. *Mol Biol Evol* 2019.

425. Bouckaert R, Vaughan TG, Barido-Sottani J et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 2019; **15**: e1006650.

426. Vasylyeva TI, du Plessis L, Pineda-Pena AC et al. Tracing the Impact of Public Health Interventions on HIV-1 Transmission in Portugal Using Molecular Epidemiology. *J Infect Dis* 2019; **220**: 233-43.

427. Rambaut A, Drummond AJ, Xie D et al. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol* 2018; **67**: 901-4.

428. Drummond AJ, Suchard MA, Xie D et al. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 2012; **29**: 1969-73.

429. Li M, Chen X, Wang H et al. A Peptide Derived from Lens Epithelium-Derived Growth Factor Stimulates HIV-1 DNA Integration and Facilitates Intasome Structural Studies. *J Mol Biol* 2020; **432**: 2055-66.

430. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010; **31**: 455-61.

431. Eberhardt J, Santos-Martins D, Tillack AF et al. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model* 2021; **61**: 3891-8.

432. Nguyen HL, Ruxrungtham K, Delaugerre C. Genetic barrier to the development of resistance to integrase inhibitors in HIV-1 subtypes CRF01_AE and B. *Intervirology* 2012; **55**: 287-95.

433. Wilkinson E, Engelbrecht S, de Oliveira T. History and origin of the HIV-1 subtype C epidemic in South Africa and the greater southern African region. *Scientific reports* 2015; **5**: 16897.

434. Faria NR, Vidal N, Lourenco J et al. Distinct rates and patterns of spread of the major HIV-1 subtypes in Central and East Africa. *PLOS Pathogens* 2019; **15**: e1007976.

435. Gray RR, Tatem AJ, Lamers S et al. Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. *Aids* 2009; **23**: F9-f17.

436. Gilbert MT, Rambaut A, Wlasiuk G et al. The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci U S A* 2007; **104**: 18566-70.

437. Mir D, Graf T, Esteves de Matos Almeida S et al. Inferring population dynamics of HIV-1 subtype C epidemics in Eastern Africa and Southern Brazil applying different Bayesian phylodynamics approaches. *Scientific reports* 2018; **8**: 8778.

438. Khodakevich L, Mehret M, Negassa H et al. Progression of Human Immunodeficeiencyy Virus epidemic in Ethiopia. *The Ethiopian Journal of Health Development* 1990; **4**: 183-7.

439. Mehret M, Khodakevich L, Zewdie D et al. HIV-1 infection among employees of the Ethiopian Freightt Transport Corporation. *The Ethiopian Journal of Health Development* 1990; **4**: 177-82.

440. Martin JL. The impact of AIDS on gay male sexual behavior patterns in New York City. *Am J Public Health* 1987; **77**: 578-81.

441. Hessol NA, Lifson AR, O'Malley PM et al. Prevalence, incidence, and progression of human immunodeficiency virus infection in homosexual and bisexual men in hepatitis B vaccine trials, 1978-1988. *Am J Epidemiol* 1989; **130**: 1167-75.

442. Nelson KE, Celentano DD, Eiumtrakol S et al. Changes in sexual behavior and a decline in HIV infection among young men in Thailand. *N Engl J Med* 1996; **335**: 297-303.

443. Hué S, Pillay D, Clewley JP et al. Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proc Natl Acad Sci U S A* 2005; **102**: 4425-9.

444. Stoneburner RL, Low-Beer D. Population-level HIV declines and behavioral risk avoidance in Uganda. *Science* 2004; **304**: 714-8.

445. Halperin DT, Mugurungi O, Hallett TB et al. A surprising prevention success: why did the HIV epidemic decline in Zimbabwe? *PLoS Med* 2011; **8**: e1000414.

446. Hamers RL, Wallis CL, Kityo C et al. HIV-1 drug resistance in antiretroviral-naive individuals in sub-Saharan Africa after rollout of antiretroviral therapy: a multicentre observational study. *Lancet Infect Dis* 2011; **11**: 750-9.

447. Huruy K, Maier M, Mulu A et al. Limited increase in primary HIV-1C drug resistance mutations in treatment naïve individuals in Ethiopia. *Journal of medical virology* 2015; **87**: 978-84.

448. Yang WL, Kouyos RD, Böni J et al. Persistence of transmitted HIV-1 drug resistance mutations associated with fitness costs and viral genetic backgrounds. *PLoS Pathog* 2015; **11**: e1004722.

449. Castro H, Pillay D, Cane P et al. Persistence of HIV-1 transmitted drug resistance mutations. *J Infect Dis* 2013; **208**: 1459-63.

450. Organization WH. Operations manual for delivery of HIV prevention, care and treatment at primary health centres in high-prevalence, resource-constrained settings: Edition 1 for fieldtesting and country adaptation. 2008.

451. Ford N, Mills EJ, Egger M. Editorial commentary: immunodeficiency at start of antiretroviral therapy: the persistent problem of late presentation to care. *Clin Infect Dis* 2015; **60**: 1128-30.

452. Mulu A, Maier M, Liebert UG. Low Incidence of HIV-1C Acquired Drug Resistance 10 Years after Roll-Out of Antiretroviral Therapy in Ethiopia: A Prospective Cohort Study. *PLoS One* 2015; **10**: e0141318.

453. Mulu A, Maier M, Liebert UG. Upward trends of acquired drug resistances in Ethiopian HIV-1C isolates: A decade longitudinal study. *PLoS One* 2017; **12**: e0186619.

454. Aklilu M, Messele T, Tsegaye A et al. Factors associated with HIV-1 infection among sex workers of Addis Ababa, Ethiopia. *Aids* 2001; **15**: 87-96.

455. Kityo C, Thompson J, Nankya I et al. HIV Drug Resistance Mutations in Non-B Subtypes After Prolonged Virological Failure on NNRTI-Based First-Line Regimens in Sub-Saharan Africa. *J Acquir Immune Defic Syndr* 2017; **75**: e45-e54.

456. Doshi RH, Sande E, Ogwal M et al. Progress toward UNAIDS 90-90-90 targets: A respondent-driven survey among female sex workers in Kampala, Uganda. *PLoS One* 2018; **13**: e0201352.

457. Van Blerk L. AIDS, mobility and commercial sex in Ethiopia: Implications for policy. *AIDS Care* 2007; **19**: 79-86.

458. Prüss-Ustün A, Wolf J, Driscoll T et al. HIV due to female sex work: regional and global estimates. *PLoS One* 2013; **8**: e63476.

459. Lancaster KE, Powers KA, Lungu T et al. The HIV Care Continuum among Female Sex Workers: A Key Population in Lilongwe, Malawi. *PLoS One* 2016; **11**: e0147662.

460. Mountain E, Mishra S, Vickerman P et al. Antiretroviral therapy uptake, attrition, adherence and outcomes among HIV-infected female sex workers: a systematic review and meta-analysis. *PLoS One* 2014; **9**: e105645.

461. Holland CE, Papworth E, Billong SC et al. Antiretroviral treatment coverage for men who have sex with men and female sex workers living with HIV in Cameroon. *J Acquir Immune Defic Syndr* 2015; **68 Suppl 2**: S232-40.

462. Cowan FM, Mtetwa S, Davey C et al. Engagement with HIV prevention treatment and care among female sex workers in Zimbabwe: a respondent driven sampling survey. *PLoS One* 2013; **8**: e77080.

463. Lindman J, Djalo MA, Biai A et al. The HIV care continuum and HIV-1 drug resistance among female sex workers: a key population in Guinea-Bissau. *AIDS Res Ther* 2020; **17**: 33.

464. Delva W, Eaton JW, Meng F et al. HIV treatment as prevention: optimising the impact of expanded HIV treatment programmes. *PLoS Med* 2012; **9**: e1001258.

465. Alary M, Lowndes CM, Van de Perre P et al. Scale-up of combination prevention and antiretroviral therapy for female sex workers in West Africa: time for action. *Aids* 2013; **27**: 1369-74.

466. Moses S, Ramesh BM, Nagelkerke NJ et al. Impact of an intensive HIV prevention programme for female sex workers on HIV prevalence among antenatal clinic attenders in Karnataka state, south India: an ecological analysis. *Aids* 2008; **22 Suppl 5**: S101-8.

467. Namale G, Kamacooko O, Bagiire D et al. Sustained virological response and drug resistance among female sex workers living with HIV on antiretroviral therapy in Kampala, Uganda: a cross-sectional study. *Sex Transm Infect* 2019; **95**: 405-11.

468. Mountain E, Pickles M, Mishra S et al. The HIV care cascade and antiretroviral therapy in female sex workers: implications for HIV prevention. *Expert Rev Anti Infect Ther* 2014; **12**: 1203-19.

469. Cowan FM, Davey CB, Fearon E et al. The HIV Care Cascade Among Female Sex Workers in Zimbabwe: Results of a Population-Based Survey From the Sisters Antiretroviral Therapy Programme for Prevention of HIV, an Integrated Response (SAPPH-IRe) Trial. *J Acquir Immune Defic Syndr* 2017; **74**: 375-82.

470. Etta EM, Mavhandu L, Manhaeve C et al. High level of HIV-1 drug resistance mutations in patients with unsuppressed viral loads in rural northern South Africa. *AIDS Res Ther* 2017; **14**: 36.

471. Gupta RK, Hill A, Sawyer AW et al. Virological monitoring and resistance to first-line highly active antiretroviral therapy in adults infected with HIV-1 treated under WHO guidelines: a systematic review and meta-analysis. *Lancet Infect Dis* 2009; **9**: 409-17.

472. Abdissa A, Yilma D, Fonager J et al. Drug resistance in HIV patients with virological failure or slow virological response to antiretroviral therapy in Ethiopia. *BMC Infect Dis* 2014; **14**: 181.

473. Chen I, Connor MB, Clarke W et al. Antiretroviral Drug Use and HIV Drug Resistance Among HIV-Infected Black Men Who Have Sex With Men: HIV Prevention Trials Network 061. *J Acquir Immune Defic Syndr* 2015; **69**: 446-52.

474. Weinstock HS, Zaidi I, Heneine W et al. The epidemiology of antiretroviral drug resistance among drug-naive HIV-1-infected persons in 10 US cities. *J Infect Dis* 2004; **189**: 2174-80.

475. WHO. Dolutegravir (DTG) and the fixed dose combination (FDC) of tenofovir/lamivudine/dolutegravir (TLD). https://www.who.int/hiv/pub/arv/DTG-TLD-arv_briefing_2018.pdf. 2018. (Accessed 15 December 2021).

476. Theys K, Libin PJK, Van Laethem K et al. An Evolutionary Model-Based Approach To Quantify the Genetic Barrier to Drug Resistance in Fast-Evolving Viruses and Its Application to HIV-1 Subtypes and Integrase Inhibitors. *Antimicrob Agents Chemother* 2019; **63**.

477. Rogers L, Obasa AE, Jacobs GB et al. Structural Implications of Genotypic Variations in HIV-1 Integrase From Diverse Subtypes. *Front Microbiol* 2018; **9**: 1754.

478. Rhee SY, Liu TF, Kiuchi M et al. Natural variation of HIV-1 group M integrase: implications for a new class of antiretroviral inhibitors. *Retrovirology* 2008; **5**: 74.

479. Han YS, Mesplède T, Wainberg MA. Differences among HIV-1 subtypes in drug resistance against integrase inhibitors. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2016; **46**: 286-91.

480. Inzaule SC, Hamers RL, Noguera-Julian M et al. Primary resistance to integrase strand transfer inhibitors in patients infected with diverse HIV-1 subtypes in sub-Saharan Africa. *J Antimicrob Chemother* 2018; **73**: 1167-72.

481. Obasa AE, Mikasi SG, Brado D et al. Drug Resistance Mutations Against Protease, Reverse Transcriptase and Integrase Inhibitors in People Living With HIV-1 Receiving Boosted Protease Inhibitors in South Africa. *Front Microbiol* 2020; **11**: 438.

482. Fish MQ, Hewer R, Wallis CL et al. Natural polymorphisms of integrase among HIV type 1-infected South African patients. *AIDS Res Hum Retroviruses* 2010; **26**: 489-93.

483. Kim JY, Kim EJ, Choi JY et al. Genetic variation of the HIV-1 integrase region in newly diagnosed anti-retroviral drug-naïve patients with HIV/AIDS in Korea. *Clin Microbiol Infect* 2011; **17**: 1155-9.

484. Casadellà M, van Ham PM, Noguera-Julian M et al. Primary resistance to integrase strand-transfer inhibitors in Europe. *J Antimicrob Chemother* 2015; **70**: 2885-8.

485. Ndashimye E, Avino M, Kyeyune F et al. Absence of HIV-1 Drug Resistance Mutations Supports the Use of Dolutegravir in Uganda. *AIDS Res Hum Retroviruses* 2018; **34**: 404-14.

486. Gupta RK, Pillay D. HIV resistance and the developing world. *Int J Antimicrob Agents* 2007; **29**: 510-7.

487. Salou M, Butel C, Comlan AS et al. Challenges of scale-up to dolutegravir-based regimens in sub-Saharan Africa. *Aids* 2020; **34**: 783-7.

488. Elvis T, Andreas J, Thokozani K et al. Prospective enhanced monitoring of dolutegravir-based first line in Malawi. *Conference on Retroviruses and Opportunistic Infections* 2020; **Boston, Massachusetts, USA**.

489. van Oosterhout JJ, Chipungu C, Nkhoma L et al. Dolutegravir Resistance in Malawi's National HIV Treatment Program. *Open Forum Infect Dis* 2022; **9**: ofac148.

490. Cevik M, Orkin C, Sax PE. Emergent Resistance to Dolutegravir Among INSTI-Naïve Patients on First-line or Second-line Antiretroviral Therapy: A Review of Published Cases. *Open Forum Infect Dis* 2020; **7**: ofaa202.

491. Rhee S-Y, Grant PM, Tzou PL et al. A systematic review of the genetic mechanisms of dolutegravir resistance. *Journal of Antimicrobial Chemotherapy* 2019; **74**: 3135-49.

492. Arimide DA, Amogne MD, Kebede Y et al. High level of HIV drug resistance and virological non-suppression among female sex workers in Ethiopia: a nation-wide cross-sectional study. *J Acquir Immune Defic Syndr* 2021.

493. Arimide DA, Esquivel-Gómez LR, Kebede Y et al. Molecular Epidemiology and Transmission Dynamics of the HIV-1 Epidemic in Ethiopia: Epidemic Decline Coincided With Behavioral Interventions Before ART Scale-Up. *Frontiers in Microbiology* 2022; **13**.

494. Nduva GM, Hassan AS, Nazziwa J et al. HIV-1 Transmission Patterns Within and Between Risk Groups in Coastal Kenya. *Scientific reports* 2020; **10**: 6775.

495. Rife BD, Mavian C, Chen X et al. Phylodynamic applications in 21(st) century global infectious disease research. *Glob Health Res Policy* 2017; **2**: 13.

496. Mutenherwa F, Wassenaar DR, de Oliveira T. Ethical issues associated with HIV phylogenetics in HIV transmission dynamics research: A review of the literature using the Emanuel Framework. *Dev World Bioeth* 2019; **19**: 25-35.

497. Oster AM, Lyss SB, McClung RP et al. HIV Cluster and Outbreak Detection and Response: The Science and Experience. *Am J Prev Med* 2021; **61**: S130-s42.

498. Pillay D, Herbeck J, Cohen MS et al. PANGEA-HIV: phylogenetics for generalised epidemics in Africa. *Lancet Infect Dis* 2015; **15**: 259-61.

499. EPHI. Ethiopian national key population HIV bio-behavioral surveillance Round I, 2013 Report: EPHI; 2014. 2014.

500. Amogne MD, Balcha TT, Agardh A. Prevalence and correlates of physical violence and rape among female sex workers in Ethiopia: a cross-sectional study with respondent-driven sampling from 11 major towns. *BMJ Open* 2019; **9**: e028247.

Paper I

frontiers in Microbiology

ORIGINAL RESEARCH
published: 25 February 2022
doi: 10.3389/fmicb.2022.821006

Molecular Epidemiology and Transmission Dynamics of the HIV-1 Epidemic in Ethiopia: Epidemic Decline Coincided With Behavioral Interventions Before ART Scale-Up

*Dawit Assefa Arimide[1,2], Luis Roger Esquivel-Gómez[3], Yenew Kebede[4], Sviataslau Sasinovich[1], Taye Balcha[1], Per Björkman[1], Denise Kühnert[3] and Patrik Medstrand[1]**

[1]Department of Translational Medicine, Lund University, Malmo, Sweden, [2]TB/HIV Department, Ethiopian Public Health Institute, Addis Ababa, Ethiopia, [3]Transmission, Infection, Diversification and Evolution Group, Max-Planck Institute for the Science of Human History, Jena, Germany, [4]Africa Centre for Disease Prevention and Control, Africa Union Commission, Addis Ababa, Ethiopia

OPEN ACCESS

Edited by:
Kok Keng Tee,
University of Malaya, Malaysia

Reviewed by:
Teiichiro Shiino,
National Center for Global Health and Medicine, Japan
Edson Delatorre,
Federal University of Espirito Santo, Brazil
Kledoaldo Lima,
Hospital das Clínicas, Universidade Federal de Pernambuco, Brazil

***Correspondence:**
Patrik Medstrand
patrik.medstrand@med.lu.se

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: 23 November 2021
Accepted: 24 January 2022
Published: 25 February 2022

Citation:
Arimide DA, Esquivel-Gómez LR, Kebede Y, Sasinovich S, Balcha T, Björkman P, Kühnert D and Medstrand P (2022) Molecular Epidemiology and Transmission Dynamics of the HIV-1 Epidemic in Ethiopia: Epidemic Decline Coincided With Behavioral Interventions Before ART Scale-Up. Front. Microbiol. 13:821006. doi: 10.3389/fmicb.2022.821006

Background: Ethiopia is one of the sub-Saharan countries hit hard by the HIV epidemic. Previous studies have shown that subtype C dominates the Ethiopian HIV-1 epidemic, but the evolutionary and temporal dynamics of HIV-1 in Ethiopia have not been closely scrutinized. Understanding the evolutionary and epidemiological pattern of HIV is vital to monitor the spread, evaluate and implement HIV prevention strategies.

Methods: We analyzed 1,276 Ethiopian HIV-1 subtype C polymerase (*pol* sequences), including 144 newly generated sequences, collected from different parts of the country from 1986 to 2017. We employed state-of-art maximum likelihood and Bayesian phylodynamic analyses to comprehensively describe the evolutionary dynamics of the HIV-1 epidemic in Ethiopia. We used Bayesian phylodynamic models to estimate the dynamics of the effective population size (N_e) and reproductive numbers (R_e) through time for the HIV epidemic in Ethiopia.

Results: Our analysis revealed that the Ethiopian HIV-1 epidemic originated from two independent introductions at the beginning of the 1970s and 1980s from eastern and southern African countries, respectively, followed by epidemic growth reaching its maximum in the early 1990s. We identified three large clusters with a majority of Ethiopian sequences. Phylodynamic analyses revealed that all three clusters were characterized by high transmission rates during the early epidemic, followed by a decline in HIV-1 transmissions after 1990. R_e was high (4–6) during the earlier time of the epidemic but dropped significantly and remained low ($R_e < 1$) after the mid-1990. Similarly, with an expected shift in time, the effective population size (N_e) steadily increased until the beginning of 2000, followed by a decline and stabilization until recent years. The phylodynamic analyses corroborated the modeled UNAIDS incidence and prevalence estimates.

Conclusion: The rapid decline in the HIV epidemic took place a decade before introducing antiretroviral therapy in Ethiopia and coincided with early behavioral, preventive, and awareness interventions implemented in the country. Our findings highlight the importance of behavioral interventions and antiretroviral therapy scale-up to halt and maintain HIV transmissions at low levels ($R_e < 1$). The phylodynamic analyses provide epidemiological insights not directly available using standard surveillance and may inform the adjustment of public health strategies in HIV prevention in Ethiopia.

Keywords: effective reproductive number, effective population size, birth–death model, phylodynamic, HIV-1 epidemic, transmission cluster, behavioral intervention, Ethiopia

INTRODUCTION

The human immunodeficiency virus type 1 (HIV-1) is one of the most devastating infectious diseases in human history (UNAIDS, 2020). At the end of 2020, an estimated 38 million people were living with HIV/AIDS worldwide. Sub-Saharan Africa, the region where HIV-1 emerged during the 1920s, remains the most affected region, accounting for close to 70% of people living with HIV worldwide (Faria et al., 2014; UNAIDS, 2020). Despite the large-scale roll-out of antiretroviral treatment (ART), HIV incidence remains high, mainly in sub-Saharan Africa (UNAIDS, 2020). Ethiopia is one of the many sub-Saharan countries that was severely affected by the HIV epidemic.

HIV-1 is classified into four phylogenetically distinct groups: M (main), N (non-M, non-O), O (outlier), and P (pending), each representing different zoonotic cross-species transmissions of simian immunodeficiency viruses from non-human primates to humans (Sharp and Hahn, 2011; Faria et al., 2014; Giovanetti et al., 2020). Group M is the most prevalent, accounts for more than 95% of all the HIV-1 infections and is divided into 10 subtypes (A–D, F–H, and J–L), more than 102 different circulating recombinant forms (CRFs), and numerous unique recombinant forms (URFs; Hemelaar et al., 2019; Giovanetti et al., 2020). Subtype C is currently the dominant HIV-1 subtype and is responsible for nearly half all HIV-1 infections globally (Hemelaar et al., 2019). Although found worldwide, no official assignment of subtype C strains into phylogenetic sub-subtypes has been made. However, several distinct genetic clades associated with geography have been defined, the southern African clades (C-SA) and the eastern African clade (C-EA). Strains of the C-EA clade and a sub-clade of C-SA, termed C′-ET, are most prevalent in Ethiopia (Thomson and Fernandez-Garcia, 2011; Arimide et al., 2018).

The first HIV-1 infection and AIDS case report in Ethiopia was in 1984 and 1986, respectively (Lester et al., 1988; Tsega et al., 1988). Initially, the epidemic was concentrated to urban areas and along major commercial routes. Serology surveys revealed high prevalence (17%–55%) among risk populations (e.g., female sex workers: FSWs, long-distance truck drivers: LDTD, and soldiers; Mehret, 1990; Mebret et al., 1990). However, after introduction of antiretroviral therapy (ART) in public health care in 2005, the prevalence among the general population decreased and stabilized at significantly lower levels while the prevalence remained high in risk populations (EPHI, 2014).

The HIV epidemic in Ethiopia is considered a generalized epidemic with heterosexual transmission being the dominant mode of transmission (Kebede et al., 2000). Since 1985, Ethiopia has implemented several community-based HIV prevention programs to improve knowledge about the infection and mode of transmission, and interventions to reduce engagement in risk behavior (Mebret et al., 1990; Okubagzhi and Singh, 2002). However, the epidemiological dynamics and their correlations with introduction of various HIV prevention and interventions programs have not been characterized.

Similar to many low-income countries, epidemiological data regarding HIV from Ethiopia are sparse and incomplete, making surveillance of the HIV epidemic challenging. The increased availability of HIV genetic sequencing data and the development of phylogenetic and phylodynamic tools has enabled the use of molecular epidemiology analysis to describe the transmission dynamics and evolutionary history of HIV (Yusim et al., 2001; Delatorre and Bello, 2012; Mir et al., 2018; Vasylyeva et al., 2019). Previous studies in Ethiopia have shown that subtype C dominates the Ethiopian HIV epidemic and have provided valuable insight into HIV genetic diversity, its origins, and epidemic dynamics, but are limited in study participant numbers and geographic and temporal representation (Abebe et al., 2001a,b; Pollakis et al., 2003; Tully and Wood, 2010; Delatorre and Bello, 2012; Mir et al., 2018). Here, we used HIV-1 subtype C *pol* gene sequences collected from different regions of Ethiopia between 1986 and 2017. We employed state-of-the-art phylogenetic and phylodynamic methods, including both Bayesian coalescent and birth–death modeling, to elucidate evolutionary trajectories and temporal dynamics of the HIV-1 epidemic in Ethiopia.

MATERIALS AND METHODS

Baseline HIV-1 Drug Resistance Survey

We conducted a prospective HIVDR survey among antiretroviral-naïve adults in St. Paul General Specialized Hospital located in Addis Ababa, Ethiopia, in 2011. We performed the study according to the WHO-recommended survey methodology (Jordan et al., 2008). Treatment-naïve adults (>18 years) eligible to start ART at the St. Paul Generalized Specialized Hospital were consecutively enrolled. Whole blood specimens were collected and transported to the Ethiopia

Public Health Institute (EPHI), the national HIV laboratory, and WHO-accredited laboratory for viral load testing and HIVDR genotyping.

HIV genotyping was done using an in-house assay as described previously (Arimide et al., 2018). Briefly, a 1,084 base-pair fragment of HIV-1 *pol* (corresponding to positions 2,243–3,326 of HXB2; GenBank Accession Number: K03455) comprising amino acids 6–99 of the protease and 1–251 of the reverse transcriptase was obtained by RT-PCR and nested PCR. The purified PCR fragments were then sequenced and analyzed on the ABI 3500xl Genetic Analyzer (Applied Biosystems, Foster City, CA, United States). Sequence assembly and editing were performed using the RECall V 2.0 HIV-1 sequencing analysis tool (University of British Columbia, Vancouver, Canada; Woods et al., 2012). All sequences reported in this study have been deposited in GenBank under Accession Numbers OL598713-OL598856.

Study Population and Sequence Dataset

We used the dataset of newly sequenced HIV-1 *pol* sequences and retrieved all publicly available Ethiopian HIV-1 subtype C *pol* sequences (matching pos. 2,243–3,326 relative of HXB2) from the Los Alamos National Laboratory (LANL) HIV Sequence database[1] (Date of access, December 2019). The quality of HIV-1 sequences was verified using the online Quality Control program of the LANL HIV sequence database (see Footnote 1) and sequences with stop codons, frameshifts, and poor quality were removed. We retained only one sequence per patient and selected the earliest sequence for patients with multiple sequences.

We removed duplicate sequences and sequences with potential contamination using the ElimDupes online tool from LANL. Moreover, to identify Ethiopian country-specific transmission clusters, we included a dataset of similar sequences from GenBank by identifying the 10 genetically closest GenBank sequences with BLAST for each Ethiopian HIV-1 subtype C sequence (Altschul et al., 1990; Mount, 2007). We only included sequences of 950 nucleotides or longer with known isolation dates and country of isolation in the analysis, since this 950-bp region has sufficient signal to reconstruct transmission links among infected individuals (Hué et al., 2004).

HIV-1 Subtyping

Initial explorative HIV-1 subtyping was performed using the online automated subtyping tools REGA v3.0 (Pineda-Peña et al., 2013), COMET v2.2 (Struck et al., 2014), and RIP (Martin et al., 2010). Putative intra-subtype recombinant sequences were detected using jpHMM (jumping profile Hidden Markov Model)[2] (Schultz et al., 2009; Arimide et al., 2018). Only non-recombinant sequences were used for the analysis. Final subtyping was determined by maximum likelihood (ML) phylogenetic tree analysis with subtype reference sequences (Arimide et al., 2018).

[1]http://www.hiv.lanl.gov

[2]http://jphmm.gobics.de/submission_hiv

Maximum Likelihood Phylogenetic Analyses

A multiple sequence alignment was obtained using MAFFT V. 7 (Katoh and Standley, 2013) and was then manually edited using BioEdit V7.0.9.0 (Hall, 1999) until a non-redundant codon alignment was obtained. To avoid the effect of drug-induced convergent evolution, positions of identified mutations causing or contributing to HIVDR were removed from the alignment, resulting in a final alignment of 909 bp (Wensing et al., 2016).

The initial ML phylogenetic tree was constructed using an online version of PhyML (Guindon et al., 2010) under the GTR + I + Γ4 (general time-reversible nucleotide substitution model using the estimated proportion of invariable sites and four gamma categories). Heuristic tree search was performed using the SPR branch-swapping algorithm. Branch support was determined with aLRT-SH (approximate likelihood ratio test Shimodaira–Hasegawa-like) implemented in PhyML (Guindon et al., 2010). A branch in the phylogeny with an aLRT-SH value ≥0.9 was considered significant (Guindon et al., 2010; Esbjörnsson et al., 2016). The ML trees were visualized using FigTree v1.4.3 (Rambaut, 2016).

Our initial ML phylogenetic trees were constructed using the combined dataset of all Ethiopian sequences and sequences from the BLAST search. To comprehensively describe the HIV-1 subtype C circulating in Ethiopia, the dataset was divided into two based on phylogenetic branch support, the C-EA and C-ET clades.

Analysis of Transmission Clusters

Separate transmission cluster analysis was performed for the two data sets using the ML phylogenetic analysis implemented by IQ-TREE under GTR + I + Γ4 as selected as the best fitting substitution model for the dataset using jModelTest v2.1.7 and with 1,000 replicates for the aLRT-SH test (Nguyen et al., 2015). Clusters with an aLRT-SH support ≥0.9 were considered significant (Guindon et al., 2010; Esbjörnsson et al., 2016). A transmission cluster was defined as a cluster in the ML phylogeny from root to tips (Esbjörnsson et al., 2016; Hassan et al., 2017; Sallam et al., 2017; Arimide et al., 2018). Clusters with an aLRT-SH-support of ≥0.9 that had a majority (at least 80%) of Ethiopia sequences were considered an Ethiopian transmission cluster. Transmission clusters were also defined based on their sizes (number of sequences/cluster), into dyads (two sequences), medium-sized clusters/networks (3–14 sequences), and large clusters (≥15 sequences; Aldous et al., 2012; Esbjörnsson et al., 2016).

To determine whether there was phylogenetic clustering by geographic region, viral sequences were grouped into six geographic regions (sequence collection location). The strength of association between the geographic location and the phylogeny was determined using two phylogeny–trait association statistics, the parsimony score (PS) and the association index (AI) tests, both of which were implemented in the Bayesian Tip-association Significance testing (BaTS) program (Parker et al., 2008). A significance level of $p < 0.05$ was used in both statistics.

Estimating Temporal Signal

For each cluster, we assessed the temporal signal of the data sets by performing root-to-tip genetic distance using TempEst (Rambaut et al., 2016). Clusters that had a positive correlation between genetic diversity and time were considered for further analysis.

Estimating Viral Phylodynamic History

The birth–death skyline model (BDSKY; Stadler et al., 2012, 2013) implemented in BEAST2 v 2.6.2 was used to quantify epidemic growth through time described by changes in the effective reproductive number (R_e) which is the average number of secondary infections from an infected individual at any given time during the epidemic (Bouckaert et al., 2019; Vasylyeva et al., 2019). We used a lognormal distribution prior, LogNorm (0,1), for the effective reproductive number with the upper bound of 10, and a LogNorm (0,1) prior for the become uninfectious rate (δ) in units per year (i.e., the inverse of the time duration of being infectious in a unit of years). We used $\delta = 0.2$, corresponding to a 5-year duration of the infectious period, as the mean of the distribution. In order to account for the uneven number of sequences per year, we employed a different sampling probability (*rho*) prior for each year, using a *beta* distribution with a mean equal to the number of samples divided by the reported number of HIV cases in the country for that year. We estimated the change in R_e for six equally spaced intervals between the time to most recent common ancestor (tMRCA) and the most recent sampling year.

Phylodynamic analyses were also performed using the Bayesian Skygrid coalescent tree prior, implemented in BEAST 1.10,4 (Gill et al., 2013; Suchard et al., 2018; Hill and Baele, 2019), to estimate changes in effective population size (N_e) through time and estimate the population growth rates (r, years^{-1}) by using a logistic growth coalescent tree prior. Analyses were performed using the GTR + I + Γ4 nucleotide substitution model. The temporal scale of the evolutionary process was estimated using a relaxed uncorrelated molecular clock model with an underlying lognormal distribution with normal priors. This allowed the estimation of the evolutionary rate (μ, nucleotide substitutions per site per year, s/s/y), the age of the most recent common ancestor (tMRCA, years), and the phylodynamic parameters.

For each of the two phylodynamic approaches, we ran three independent Markov Chain Monte Carlo (MCMC) chains until all associated parameters converged to ensure good mixing (ESS > 200) after discarding the first 10% of the MCMC chains. The convergence of the MCMC was inspected visually and by calculating the ESS for each parameter using Tracer v 1.7.5 (Rambaut et al., 2018). We used LogCombiner to combine the different independent results (log and corresponding tree file) from the multiple chains (Drummond et al., 2012). We used the bdskytools package[3] in R to plot the results of the BDSKY analysis.

[3]https://github.com/laduplessis/bdskytools

Ethical Approval

We obtained ethical approval from the Research and Ethical Clearance Committee of EPHI and the National Health Research Ethics Review Committee of the Ministry of Science and Technology of Ethiopia. All participants for the baseline HIV drug resistance survey provided written informed consent to participate in the study.

RESULTS

Study Population and Initial Phylogenetic Analysis

We retrieved 1,132 Ethiopian *pol* sequences from LANL, collected from different parts of the country from 1986 to 2017. Additionally, we included 144 HIV-1 *pol* sequences from the baseline HIVDR survey. The combined dataset ($n = 1{,}276$ sequences) contained 399 putative recombinant sequences, which were removed from further analysis. We further included a dataset of similar sequences from GenBank by identifying the 10 genetically closest GenBank sequences with BLAST for each of the 877 non-recombinant Ethiopian HIV-1 subtype C sequences in the study. The final combined dataset contained 1,333 non-recombinant HIV-1 subtype C *pol* sequences (877 Ethiopian and 456 global), which were used for phylogenetic analysis. The ML phylogenetic tree identified two distinct and well-supported clades, the C-EA and C′-ET clades (**Figure 1**). Among the 877 Ethiopian sequences included in the analysis, the C-EA clade represented 567 (65.0%) of the sequences, while 310 (35.0%) belonged to the C′-ET clade.

Most of the Ethiopian C-EA sequences were found in one large cluster (aLRT = 0.87), and only 33 Ethiopian C-EA sequences fell outside this cluster. Sequences of the global dataset intermixed with the Ethiopian sequences and represented sequences obtained most frequently ($N = 67$, 33.3%) in other East African countries, North America, and Europe. Sequences from Burundi dominated the basally located sequences. In the case of the second major clade, the majority of the Ethiopian C′-ET sequences (95.5%) formed a well-supported sub-clade (aLRT = 0.92), branching off from the basally located sequences. Southern African countries' sequences were intermixed ($N = 121$, 47.1%) with the Ethiopian C′-ET sequences, but they were most prominent at the base of the clade.

Transmission Cluster Analysis

We inferred transmission clusters by separate ML phylogenetic tree analyses of the two clades (C-EA and C′-ET). For C-EA, the ML phylogenetic tree contained a total of 810 sequences (with reference sequences) and identified two large well-supported clusters of 259 (C-EA-259) and 148 sequences (C-EA-148). The C-EA-259 cluster (aLRT = 0.93) contained 213 Ethiopian (82.2% of the sequences of the cluster) and 46 non-Ethiopian sequences, collected 1988–2017, and the C-EA-148 cluster (aLRT = 0.95) contained 124 Ethiopian (83.8%) sequences and 24 non-Ethiopian sequences, collected 1996–2017 (**Figure 2**). Moreover, we identified six networks (medium-sized clusters) and nine dyads.

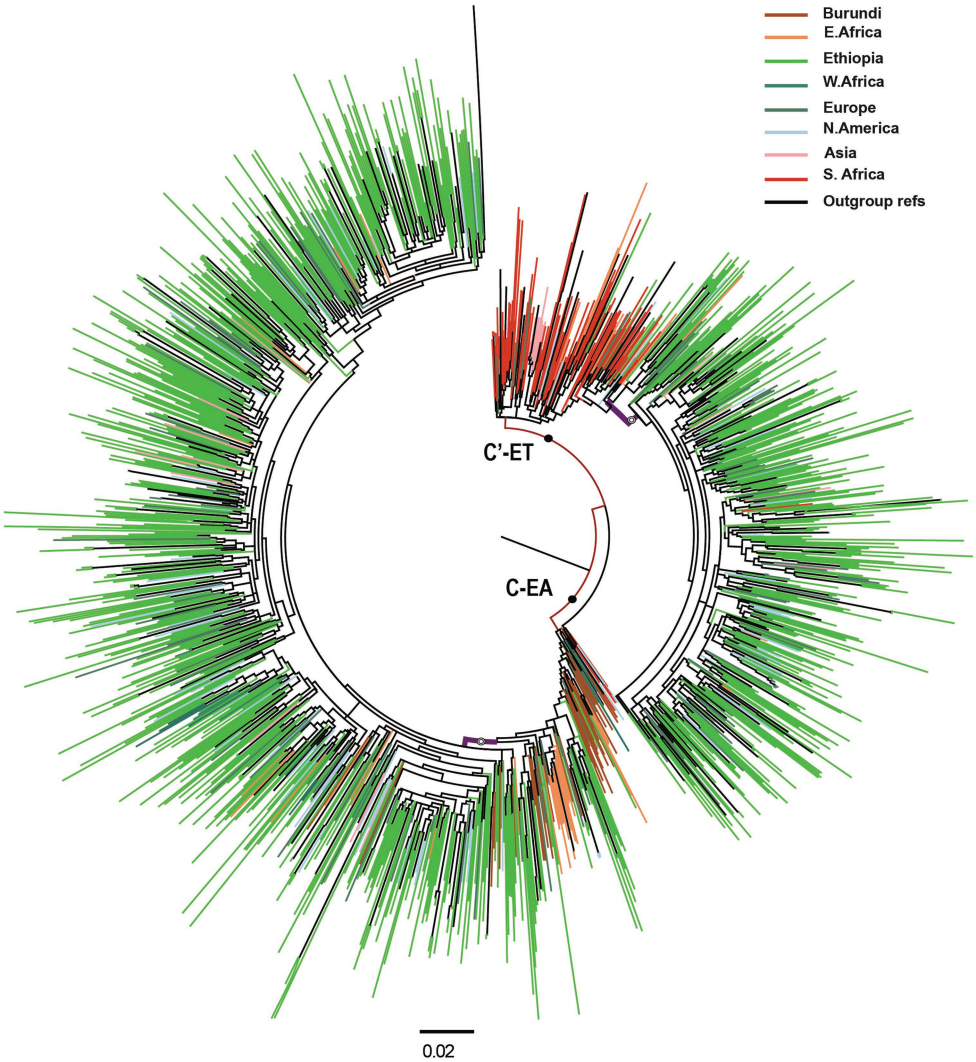


FIGURE 1 | Maximum likelihood (ML) phylogenetic tree of HIV-1 subtype C pol sequences ($n = 1,333$). Maximum likelihood phylogenetic tree constructed using 877 Ethiopian subtype C *pol* sequences collected 1986–2017 and 456 global subtype C *pol* sequences. Colored tips are according to the geographic origin of sequences, as indicated in the legend in the top right corner. Branches defining the major clades (C-EA and C′-ET) are those indicated with a filled circle, having branch support (aLRT-SH) >0.9. The Ethiopian clusters are indicated by open circles, having branch support (aLRT-SH) >0.9. The scale bar represents 0.02 substitutions/site.

Among the 580 C′-ET sequences, we identified one well-supported (aLRT = 0.96) large cluster containing 153 sequences [C′-ET-153; 124 Ethiopian (81.0% of the sequences in the cluster)], and 29 non-Ethiopian sequences collected 1995–2017 (**Figure 3**). Since we had no information about the associated risk behavior of the respective individuals, we could not associate cluster

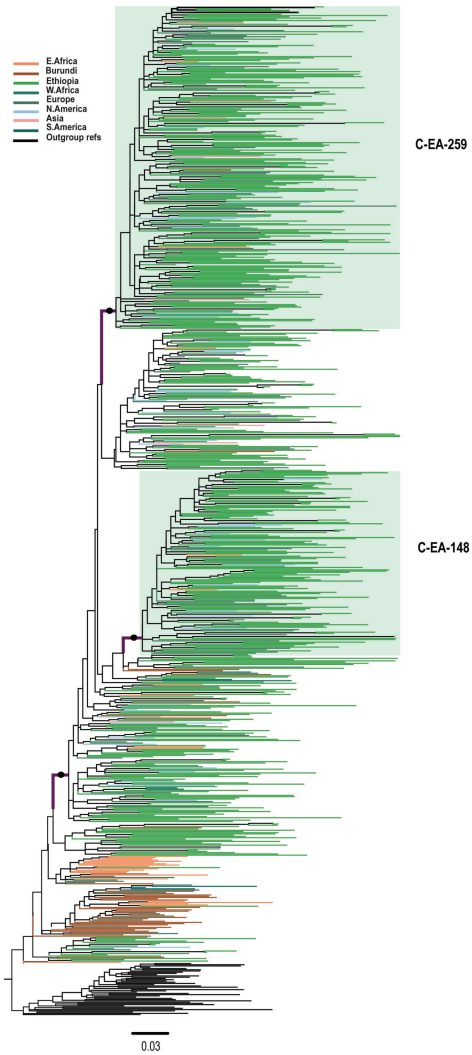


FIGURE 2 | Maximum likelihood phylogenetic tree of HIV-1 C-EA clade *pol* sequences ($n = 810$). Maximum likelihood phylogenetic tree constructed using 567 Ethiopian subtype C *pol* sequences collected between 1986 and 2017, 201 global subtype C *pol* sequences, and 42 reference sequences. Colored tips are according to the geographic origin of sequences, as indicated in the legend in the top left corner. The C-EA-259 and C-EA-148 clusters are highlighted in green, corresponding to a branch support aLRT-SH >0.9 and >80% Ethiopian sequence. The filled circle defining the C′-ET clade represents an aLRT-SH >0.9. The tree was rooted using the C′-ET reference sequences. The scale bar represents 0.03 nucleotide substitutions per site.

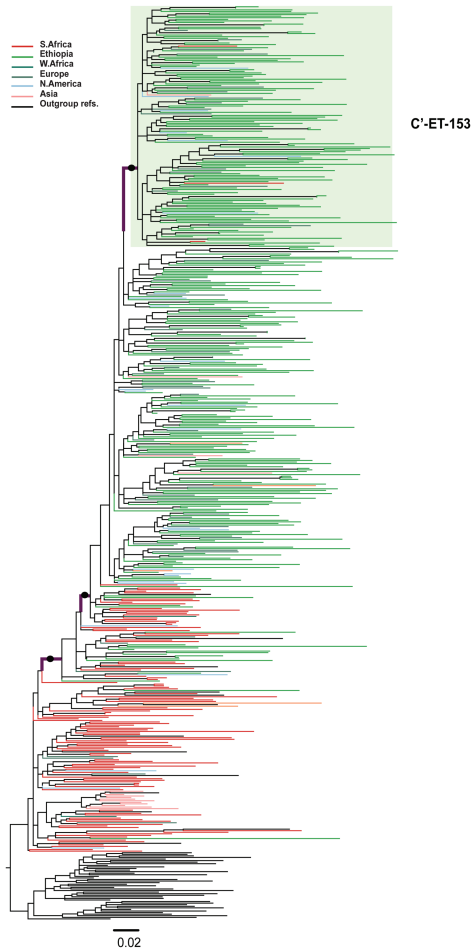


FIGURE 3 | Maximum likelihood phylogenetic tree of HIV-1 C′-ET clade *pol* sequences ($n = 580$). Maximum likelihood phylogenetic tree constructed using 310 Ethiopian subtype C *pol* sequences collected 1986–2017, 227 global subtype C *pol* sequences, and 43 reference sequences. Colored tips are according to the geographic origin of sequences, as indicated in the legend in the top left corner. The C′-ET-153 cluster, highlighted in a green shade, is defined by branch support (aLRT-SH) of >0.9 and >80% Ethiopian sequences. The branches with filled circles possibly defining the C′-ET clade have branch support (aLRT-SH) >0.9. The tree was rooted using the C-EA reference sequences. The scale bar represents 0.02 nucleotide substitutions per site.

formation with risk behavior. However, geographic location was available for analysis but was not associated with cluster formation.

Evolutionary Rates and Dates of HIV-1 Subtype C in Ethiopia

To comprehensively describe the Ethiopian HIV-1 epidemic, we further performed analyses on the three large clusters (C-EA-259, C-EA-148, and C′-ET-153). Root-to-tip analysis indicated a temporal signal in the three data sets (correlation coefficient of 0.53, 0.43, 0.47 for C-EA-259, C-EA-148, and C′-ET-153, respectively).

First, we estimated the tMRCA of the transmission clusters and the C-EA clade. The clade tMRCA represents the date of the origin of the circulating subtype C clade in the region. In contrast, the estimated tMRCA of the Ethiopian transmission clusters should approximate the introductions and local spread of the viral strains in the country (Dalai et al., 2009; Esbjornsson et al., 2011). Based on the inferred tMRCA, the posterior median estimates of C-EA-259 (1975, 95% HPD: 1970–1979) and C-EA-148 (1976, 95% HPD: 1963–1985) were older than the estimate for the C′-ET transmission cluster (1983, 95% HPD: 1975–1988; **Table 1**). The median root tMRCAs for C-EA clade was estimated to be 1971 (95% HPD: 1966–1976).

TABLE 1 | Population dynamics and evolutionary estimates for subtype C cluster in Ethiopia.

	Subtype C clade/Cluster		
	C-EA-259	**C-EA-148**	**C′-ET-153**
Sequences from Ethiopia (n)	213	124	124
Range of collection (year)	1988–2017	1996–2017	1995–2017
Mean coefficient of variation	0.24	0.23	0.33
Median evolutionary substitution rate (95% HPD)[1]	1.76 (1.49–2.00)	1.74 (1.11–2.40)	1.83 (1.33–2.32)
Median year of tMRCA (95% HPD)	1975 (1970–1979)	1976 (1963–1985)	1983 (1975–1988)
Median rate of population growth (95% HPD)[2]	0.66 (0.51–0.81)	0.61 (0.38–0.86)	0.80 (0.53–1.10)
Median epidemic doubling time (years) (95% HPD)[3]	1.05 (0.86–1.36)	1.12 (0.81–1.82)	0.86 (0.63–1.31)
Maximum effective reproductive number (R_e)[4]	6.13 (95% HPD, 3.53–10.14)	3.93 (95% HPD, 1.88–7.07)	4.88 (95% HPD, 2.57–8.57)
Basic reproductive number (R_0)[5]	4.30 (95% HPD: 3.55–4.05)	4.05 (95% HPD: 2.90–5.25)	5.00 (95% HPD: 3.65–6.50)
Median become uninfectious rate (95% HPD)[6]	0.13 (0.06, 0.20)	0.19 (0.08, 0.29)	0.21 (0.09, 0.33)

[1]*Median number of substitutions/site/year × 10^{-3}.*
[2]*Median population growth rate (r) per year, determined in BEAST v1.10.4 using a logistic tree prior.*
[3]*The time (years) required to double the effective number of infections (λ), calculated as $\lambda = ln(2)/r$, where r is the population growth rate.*
[4]*R_e (effective reproductive number) which reflect the average number of secondary infections from an infected individual at any given time during the epidemic.*
[5]*Basic reproductive number (R_0) which reflects the average number of infections generated by an infected individual in a population where all individuals are susceptible to infection calculated by using the formula $R_0 = rD + 1(56)$ (where r is the population growth rate and D is the average duration of infectiousness period).*
[6]*Become uninfectious rate, which reflects the inverse of the time duration of being infectious, in the unit of years.*

The median estimated evolutionary rate was in the range $1.76–1.82 \times 10^{-3}$ substitutions/site/year for the three clusters, with overlapping 95% HPD intervals (**Table 1**), indicating no significant difference of evolutionary rates among the three clusters.

Temporal Dynamics of Viral Transmission

Direct estimation of the temporal dynamics of the effective reproductive number, R_e, was performed using the BDSKY model. The BDSKY analysis assumed a piecewise constant R_e, changing over six equidistance intervals between the tMRCA and the most recent sampling. The R_e showed similar dynamics for the three clusters. From the start until the beginning of the 1990s, R_e remained consistently high ($R_e > 1$) and dropped below the epidemiological threshold ($R_e < 1$) at the mid-1990s and remained below one till recent years (**Figures 4A–C**). In all three clusters, we observed the maximum R_e values (4-6) during the early period (before 1990) of the epidemic (**Table 1**).

The posterior median estimates of the become non-infectious rate obtained for each cluster were of 0.13 (95% HPD: 0.06–0.20) for C-EA-259, 0.19 (95% HPD: 0.08–0.29) for C-EA-148, and 0.21 (95% HPD: 0.09–0.33) for C′-ET-153, which translates to an infectious period of ~8 years for C-EA-259 and ~5 years for the other two clusters. Despite the longer infectious period estimated for C-EA-259, the overlapping HPD indicates no significant differences among clusters.

We performed different sensitivity analyses to explore the robustness of our BDSKY estimates. We used different values for the mean of the become non-infectious rate prior (δ) going from 6 months to 10 years (*δ*: 2, 1, 0.5, 0.2, 0.125, and 0.1). We also performed the sensitivity analysis by estimating the effective reproductive number for six and ten equally spaced intervals between tMRCA and the most recent sample. We obtained similar results for all analyses. R_e was consistently >1 for the early period until the early 1990s, followed by a decline to $R_e < 1$ after the mid-1990s.

We further performed phylodynamic analysis using the Bayesian Skygrid model to estimate the temporal characteristics of the HIV-1 epidemic in Ethiopia. We analyzed the three Ethiopian clusters and estimated the change in the effective population size (N_e) through time, representing the change in the total number of infections contributing to new cases. The Bayesian skygrid inference revealed a rapid increase in N_e for all the three clusters from the initial introduction period until shortly before the year 2000, followed by a decline and stabilization in N_e until recent years (**Figures 4D–F**). We also determined the population growth rate (*r*), the rate of increase in the effective population size with time, using the logistic growth model of the coalescent parametric model. The median growth rate was 0.66, 0.61, and 0.80 $year^{-1}$ for clusters C-EA-259, C-EA-148, and C′-ET-153, respectively, with overlapping HPD intervals (**Table 1**).

We also estimated the mean coalescent-based basic reproductive number (R_0) values for each cluster from the logistic growth model using the formula $R_0 = rD + 1$ (Pybus et al., 2001; where r is the population growth rate and D is the average duration of infectiousness period). Assuming an average infectious period

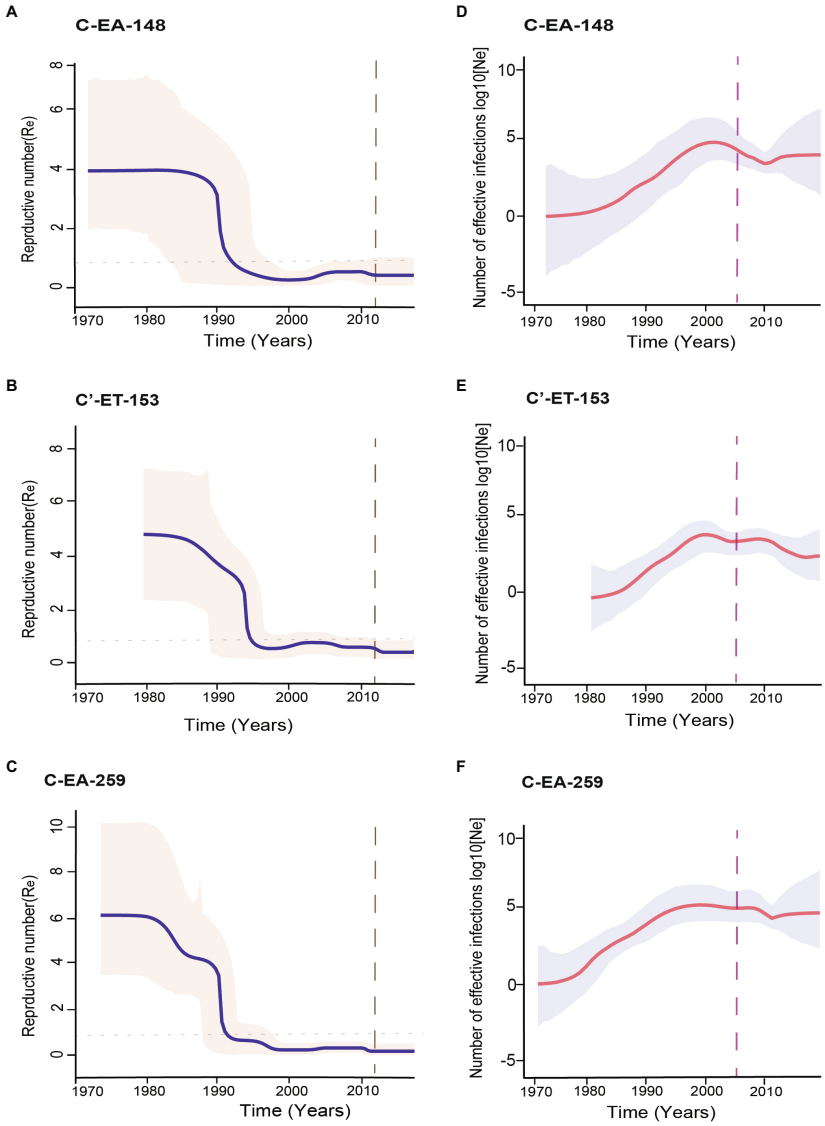


FIGURE 4 | Population dynamics of the HIV-1 epidemic in Ethiopia using the three major clusters (C-EA-259, C-EA-148, and C'ET-153). **(A–C)** The temporal dynamics of effective reproductive number (R_e) using the Bayesian birth–death model, the median R_e are shown by the continuous blue line, and indicated in a pink shade is the 95% highest probability density (HPD) intervals. The gray dashed line indicates the last coalescent event reported by the lineage through time (LTT) analysis. The horizontally dotted line represents the epidemiological threshold (Re = 1). **(D–F)** The median estimates of the effective population size (N_e) over time using the Bayesian skygrid model. The red line shows the median logarithmic effective population size (N_e) over viral generation time (*t*), representing effective transmissions, and the gray shade indicates the 95% highest probability density (HPD) intervals. The pink dashed line represents the time of antiretroviral therapy (ART) introduction in Ethiopia.

of 5 years, R_0 was in the range 4.0–5.0 for the three clusters, all with overlapping 95% HPD intervals (**Table 1**).

DISCUSSION

In this study, we analyzed a large dataset of HIV-1 *pol* sequences collected from different regions of Ethiopia during more than 30 years. We used both Bayesian coalescent and birth–death models to characterize the dynamics of the HIV-1 epidemic in the country. Overall, our analysis confirms that strains of two subtype C clades are circulating in Ethiopia, supporting the hypothesis that the HIV-1 epidemic in Ethiopia is the result of at least two independent HIV-1 introductions from eastern and southern African countries (Delatorre and Bello, 2012). Moreover, the phylodynamic analyses revealed that the epidemic dynamics in Ethiopia were characterized by an expanding epidemic growth from the start of the epidemic until the mid-1990s, followed by a sharp decline in HIV-1 transmissions. The decline in R_e occurred many years before introducing ART and coincided with early behavioral, preventive interventions, and public health awareness campaigns implemented in Ethiopia.

R_e is a proxy for HIV incidence and describes the transmission dynamics; $R_e > 1$ means that the epidemic is growing, $R_e < 1$ shows the epidemic is declining, while $R_e = 1$ shows that the epidemic is stabilizing (Stadler et al., 2013). Our phylodynamic analysis showed that the three clusters followed similar epidemic trends. The BDSKY model indicated epidemic growth ($R_e > 1$) from the 1970s to the early 1990s. The basic reproductive number (R_0) and mean initial R_e were comparably high for the clusters, indicating an early exponential epidemic growth. Similarly, a high epidemic growth rate was estimated for each cluster (0.61–0.80 $year^{-1}$) and a steady increase in N_e until the beginning of 2000, highlighting the upward trend of HIV transmissions in Ethiopia during the period.

The exponential epidemic growth observed in our analyses is consistent with retrospective serological data, which showed a massive increase of HIV infections among risk populations in Addis Ababa and cities along the main trading routes in Ethiopia during this early period. An extensive survey on FSWs operating in the main trading routes of Ethiopia in 1988 reported an HIV-1 prevalence between 5.3% and 38.1% (Mehret et al., 1990c). Studies performed in the capital Addis Ababa, 1988–90, showed an increase in prevalence from 25% to 54%, and 13% to 18% among FSWs and LDTDs, respectively (Khodakevich et al., 1990; Mehret et al., 1990a,c; Kebede et al., 2000), and 12%–18% among soldiers 1990–1993. Similarly, an increase in HIV prevalence among pregnant women attending antenatal care clinics (ANC) in Addis Ababa (4.6%–10.5%, 1989–90; Kebede et al., 2000) indicated extensive spread in the population.

During the early years, the rapid epidemic increase was most likely due to lack of awareness of HIV, high mobility among FSWs, high-risk sexual behavior, high STI prevalence among the general population (Desta et al., 1990; Mehret et al., 1990b; Negassa et al., 1990), while no prevention interventions were in place. The increased population movement following considerable urbanization and political instability in the country during this early period might also have contributed to the high HIV prevalence and epidemic spread (Hladik et al., 2006; Esbjornsson et al., 2011).

Although there is a lack of data that can describe the HIV epidemic on a national scale, different studies have shown a decline in new infections since the mid-1990s, corroborating our results (Kebede et al., 2000; Tsegaye et al., 2002; Wolday et al., 2007). The decline in HIV prevalence among young adult women (15–24 years) represents a well-established indicator of epidemic decline. It measures the frequency of relatively recent infections and is less influenced by death (Tsegaye et al., 2002). The HIV prevalence trend among young women (15–24 years) attending ANCs in Addis Ababa between 1995 and 2003 declined significantly from 24.2% to 12.9% (Tsegaye et al., 2002; Wolday et al., 2007). Moreover, there was a sharp decline in HIV prevalence among young blood donors in Addis Ababa and nine other towns during this period (Kebede et al., 2000).

Due to a lack of comprehensive data, it has not been easy to obtain estimates of the national incidence trend in Ethiopia. However, a study done to assess the temporal trend among pregnant women who attended the ANCs in the capital Addis Ababa, assessing >7,000 serum specimens collected 1995–2003, showed a significant decline in the HIV-1 incidence rate (from 7.7% to 2.0%, 1995–2003; Wolday et al., 2007). The reduction was substantial among young ANC attendees (aged 15–19 years), indicating an epidemic decline (7.8% to 0.0%, 1995–2003; Wolday et al., 2007). A mathematical modeling study also demonstrated a substantial reduction in the HIV incidence in Ethiopia after 1995 with an estimated annual decline of 6.3% per year, resulting in a total decrease of 77% between 1990 and 2016 (Deribew et al., 2019).

The early decline in the HIV transmissions observed in our study and documented in serological surveys coincide with the change of sexual behavior, prevention, and better control of other sexually transmitted infections (STIs) achieved through the sustained public education and mobilization campaigns. Ethiopia was one of the first countries in sub-Saharan Africa to introduce a task force to prevent and control HIV/AIDS and STI infections, including a national plan for the HIV epidemic response intervention (Zewdie et al., 1990; Kebede et al., 2000; Kloos and Mariam, 2000; Okubagzhi and Singh, 2002). During the early 1990s, Ethiopia had implemented a wide range of HIV prevention and information programs. Implementation of several behavioral interventions and awareness programs took place using the national media, schools, and public gatherings (Hadgu et al., 1990; Zewdie et al., 1990). These programs mainly focused on sustained health education, risk reduction, condom promotion, and prevention and control of STIs (Zewdie et al., 1990; Okubagzhi and Singh, 2002).

The national survey data on behavioral risk factors in Ethiopia are limited. However, different program reviews (1989–1991) and two nationwide surveys on condom use (1987–1993) revealed that these interventions led to changes in sexual risk behavior and increased knowledge about HIV/AIDS. Moreover, the intervention increased condom use and substantially reduced both non-regular partner and STI (Mehret et al., 1996). Similarly, another study showed condom use increased, and non-regular

partners decreased among high school students in Addis Ababa and Gondar in the period after 1990 (Kebede et al., 2000). Moreover, a study among male factory workers in Ethiopia showed a change in sexual risk behavior (Mekonnen et al., 2003). Although it is difficult to quantify the impact of the different interventions on HIV incidence, it is reasonable to assume that the various prevention programs impacted HIV transmissions.

Several other studies outside Ethiopia have reported a significant decline in HIV prevalence after behavioral interventions (Martin, 1987; Hessol et al., 1989; Nelson et al., 1996; Stoneburner and Low-Beer, 2004; Halperin et al., 2011). A study in Uganda and Zimbabwe showed a significant decline in HIV prevalence after 1990, resulting from public health intervention on reduced sexual risk behavior (Stoneburner and Low-Beer, 2004; Halperin et al., 2011). Similarly, behavioral interventions resulted in a substantial decrease in HIV transmissions among MSM in Europe and North America in the mid-1980s and heterosexuals in Thailand in the early 1990s (Martin, 1987; Hessol et al., 1989; Nelson et al., 1996; Hué et al., 2005). In line with our results, a comprehensive review of empirical and modeled HIV incidence trends across 20 countries in Sub-Saharan Africa, 1990–2012, revealed a decline in incidence commenced before introducing ART programs, highlighting the significance of behavioral intervention in reducing HIV transmissions (Taaffe et al., 2014).

The trends of the phylodynamic analyses (**Figure 4**) are in concordance with the UNAIDS HIV incidence and prevalence modeled estimates (**Figure 5**),[4] showing a high incidence and prevalence during the years before 1990–1995, followed by a decline in incidence and stabilization in prevalence.

Thus, our results align well with published serological and epidemiological trends in Ethiopia. The epidemic decline coincides with the timing of behavioral interventions in Ethiopia, suggesting a link between the early decline of HIV spread and behavioral interventions many years before the implementation of ART in the country. However, the introduction of ART, which has proved to successfully suppress HIV replication and reduce the risk of onward transmissions, has significantly contributed to reducing HIV transmission, mortality, and maintaining the epidemic decline (Cohen et al., 2011).

The phylogenetic analysis confirms that strains of two HIV-1 subtype C clades (C′-ET and C-EA) are circulating in Ethiopia, suggesting that the HIV epidemic in Ethiopia arose by at least two independent introductions of founder strains from the eastern and southern African countries, respectively (Pollakis et al., 2003; Delatorre and Bello, 2012). Previous studies have defined several distinct subtype C clades that, in most cases, are associated with geographical regions (Thomson and Fernandez-Garcia, 2011). In the case of the Ethiopian lineages, they represent a southern African clade (where the Ethiopian C-SA sub-clade named C′-ET is more or less confined to Ethiopia) and an eastern African clade (C-EA). Our findings align with previous studies showing a distinct phylogeographic subdivision of the HIV-1 subtype C circulating in east, central, and southern African countries (Abebe et al., 2000; Pollakis et al., 2003; Thomson and Fernandez-Garcia, 2011; Delatorre and Bello, 2012).

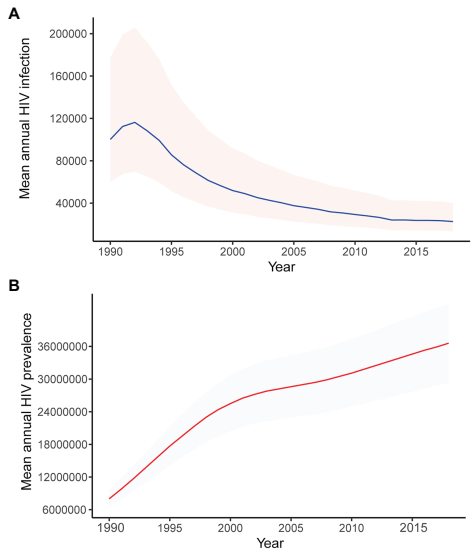


FIGURE 5 | Modeled mean annual HIV **(A)** incidence and **(B)** prevalence, 1990–2018. Data obtained from UNAIDS.

The transmission cluster analysis indicated that the Ethiopian sequences formed large clusters, indicative of a few major introductions or expansions in the country. The three Ethiopian clusters described here were mixed regarding collection sites, suggesting intermixing of the HIV epidemic in Ethiopia. Different socio-cultural and behavioral factors might also contribute to cluster formation, and assessing these factors are essential for designing HIV-1 transmission preventive strategies. However, the sequences obtained from the public database do not contain information on risk factors, sociodemographic, and other clinical information. Hence, further analysis on factors associated with cluster formation was not possible to discern in this study.

Sequences of Burundi dominated the basal root of the large monophyletic C-EA clade incorporating more than 90% of the Ethiopian sequences, which is in line with a previous study showing that the C-EA clade likely had its origin in Burundi (Delatorre and Bello, 2012). Moreover, the basal root of the monophyletic clade defining the C′-ET clade was dominated by sequences from southern African countries, possibly reflecting the origin of this clade from southern African countries. However, our analysis could not identify the exact countries. Interconnectivity between populations due to geographic proximity has been an essential factor for the spread of HIV across African countries (Wilkinson et al., 2015, 2016; Faria et al., 2019). However, the large distances and cultural interconnectivity between Ethiopia, Burundi, and southern African countries suggest that other factors were in play. Population

[4]https://aidsinfo.unaids.org

movements (due to unknown reasons) could have played a role in the introduction of HIV-1 subtype C to Ethiopia, similar to those observed in other parts of Africa (Gray et al., 2009).

Estimating the date of origin and timing of transmissions of HIV is essential to understanding the dynamics of HIV spread. Here, integral to our analysis of HIV transmission dynamics, we also obtained the tMRCA of the C-EA and C′-ET clades in Ethiopia. The molecular dating analysis suggested that the introduction of the C-EA clade took place more than a decade before the first reported AIDS case in Ethiopia. The dating is plausible considering that AIDS symptoms typically arise 6–10 years after infection. Moreover, our tMRCA estimates coincide with estimates of the introductions of the C-EA clade in other Eastern Africa countries, including Kenya, Tanzania, and Uganda (Delatorre and Bello, 2012), and are consistent with previous estimates for subtype C introduction in Ethiopia (Delatorre and Bello, 2012; Mir et al., 2018). Notably, this period also coincided with a large population migration from Burundi, which could have played a crucial role in disseminating the C-EA clade to Eastern Africa countries (Delatorre and Bello, 2012).

In contrast, the tMRCA of C′-ET was estimated at the beginning of the 1980s and is likely the result of a single introduction. This period coincided with the years of socio-political changes in the southern African countries and is associated with a steep growth of the HIV epidemic and viral migrations within southern African countries (Wilkinson et al., 2015, 2016).

To our knowledge, this study represents the most comprehensive study concerning the HIV epidemic in Ethiopia to date. It employs a large number of HIV-1 *pol* sequences collected during more than 30 years (1986–2017) from different geographical locations in Ethiopia. Moreover, we used state-of-art phylogenetic and phylodynamic methods to investigate the dynamics of the epidemic. Like many other molecular epidemiology studies, we incorporated HIV-1 *pol* gene sequences deposited in public databases in our analysis. As new HIV infections are recorded, more sequencing will allow to keep track of the ongoing transmission dynamics. The total sampling density was low, mainly due to a generally low sequencing coverage in Ethiopia, compared to the country's total number of infected individuals. Thus, the transmission clusters identified here cannot fully represent Ethiopia's entire HIV-1 transmission networks. Moreover, we based our analysis on HIV-1 *pol* sequences, representing the most sequenced HIV region due to the numerous published HIVDR studies. Although the HIV-1 *pol* fragment has sufficient phylogenetic signal for phylogenetic analysis of HIV (Hué et al., 2004), longer sequences, including whole genome sequences, may have provided a more informative inference of the HIV-1 molecular epidemiology and transmission history. Finally, the sequences used in this analysis lacked associated information, such as clinical, demographic, risk population assignment, or socio-economic data and, hence, we could not perform a detailed analysis of associated risk factors for HIV transmissions in our study.

In summary, we have employed state-of-art phylogenetic and phylodynamic approaches to describe the molecular epidemiology of HIV in Ethiopia. Our findings indicate that two distinct HIV subtype C strains were introduced in Ethiopia at the beginning of the 1970s and 1980s, followed by rapid epidemic growth until it started to decline in the mid-1990s, a decade before ART roll-out in Ethiopia. The sharp decline coincided with several behavioral prevention interventions and awareness campaigns. Our finding highlights the significance of scaling up behavioral and risk reduction interventions in addition to ART scale-up in the HIV/AIDS control strategy.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

DA, PB, and PM conceived and designed the study. DA, YK, PM, and TB coordinated the laboratory tests. DA, LE-G, SS, DK, and PM conducted the phylogenetic and phylodynamic analysis and interpreted the results. DA and PM wrote the manuscript. All authors reviewed the draft and contributed important intellectual content to the final version. All authors agreed and approved to the published version of the manuscript.

FUNDING

PM was supported by the Swedish Research Council (grant numbers 2019-05235 and 2020-02344), Österlund Foundation, and a donation to PM and PB through the Medical Faculty, Lund University, supported the research. LRE-G and DK were supported by the Max Planck Society. The computations were enabled by resources in project SNIC 2021-5-69 provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX, partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

ACKNOWLEDGMENTS

We express our deep appreciation to all patients from whom data were collected and analyzed and all staff involved in the HIV-1 drug resistance survey at the St. Paul General Specialized Hospital in Addis Ababa, Ethiopia. We also acknowledge the support given by the EPHI research laboratory team and St. Paul General Specialized Hospital and antiretroviral clinics staff.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.821006/full#supplementary-material

Supplementary Table S1 | The GenBank accession numbers and the year of sampling of sequences used for this study.

REFERENCES

Abebe, A., Lukashov, V. V., Pollakis, G., Kliphuis, A., Fontanet, A. L., Goudsmit, J., et al. (2001a). Timing of the HIV-1 subtype C epidemic in Ethiopia based on early virus strains and subsequent virus diversification. *AIDS* 15, 1555–1561. doi: 10.1097/00002030-200108170-00013

Abebe, A., Lukashov, V. V., Rinke De Wit, T. F., Fisseha, B., Tegbaru, B., Kliphuis, A., et al. (2001b). Timing of the introduction into Ethiopia of subcluster C' of HIV type 1 subtype C. *AIDS Res. Hum. Retrovir.* 17, 657–661. doi: 10.1089/088922201300119770

Abebe, A., Pollakis, G., Fontanet, A. L., Fisseha, B., Tegbaru, B., Kliphuis, A., et al. (2000). Identification of a genetic subcluster of HIV type 1 subtype C (C') widespread in Ethiopia. *AIDS Res. Hum. Retrovir.* 16, 1909–1914. doi: 10.1089/088922200050195865

Aldous, J. L., Pond, S. K., Poon, A., Jain, S., Qin, H., Kahn, J. S., et al. (2012). Characterizing HIV transmission networks across the United States. *Clin. Infect. Dis.* 55, 1135–1143. doi: 10.1093/cid/cis612

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Arimide, D. A., Abebe, A., Kebede, Y., Adugna, F., Tilahun, T., Kassa, D., et al. (2018). HIV-genetic diversity and drug resistance transmission clusters in Gondar, northern Ethiopia, 2003-2013. *PLoS One* 13:e0205446. doi: 10.1371/journal.pone.0205446

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., et al. (2019). BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15:e1006650. doi: 10.1371/journal.pcbi.1006650

Cohen, M. S., Chen, Y. Q., McCauley, M., Gamble, T., Hosseinipour, M. C., Kumarasamy, N., et al. (2011). Prevention of HIV-1 infection with early antiretroviral therapy. *N. Engl. J. Med.* 365, 493–505. doi: 10.1056/NEJMoa1105243

Dalai, S. C., de Oliveira, T., Harkins, G. W., Kassaye, S. G., Lint, J., Manasa, J., et al. (2009). Evolution and molecular epidemiology of subtype C HIV-1 in Zimbabwe. *AIDS* 23, 2523–2532. doi: 10.1097/QAD.0b013e3283320ef3

Delatorre, E. O., and Bello, G. (2012). Phylodynamics of HIV-1 subtype C epidemic in East Africa. *PLoS One 7*:e41904. doi: 10.1371/journal.pone.0041904

Deribew, A., Biadgilign, S., Deribe, K., Dejene, T., Tessema, G. A., Melaku, Y. A., et al. (2019). The burden of HIV/AIDS in Ethiopia from 1990 to 2016: evidence from the global burden of diseases 2016 study. *Ethiop. J. Health Sci.* 29, 859–868. doi: 10.4314/ejhs.v29i1.7

Desta, S., Feleke, W., Yusuf, M., Mehiret, M., Geyid, A., Ghidinelli, M., et al. (1990). Prevalence of STD and STD related risk factors in sex workers of Addis Ababa. *Ethiop. J. Heal. Dev.* 4, 149–153.

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

EPHI (2014). Ethiopian National Key Population HIV Bio-Behavioral Surveillance Round I, 2013 Report: EPHI; 2014.

Esbjörnsson, J., Mild, M., Audelin, A., Fonager, J., Skar, H., Bruun Jørgensen, L., et al. (2016). HIV-1 transmission between MSM and heterosexuals, and increasing proportions of circulating recombinant forms in the Nordic countries. *Virus Evol.* 2:vew010. doi: 10.1093/ve/vew010

Esbjornsson, J., Mild, M., Mansson, F., Norrgren, H., and Medstrand, P. (2011). HIV-1 molecular epidemiology in Guinea-Bissau, West Africa: origin, demography and migrations. *PLoS One* 6:e17025. doi: 10.1371/journal.pone.0017025

Faria, N. R., Rambaut, A., Suchard, M. A., Baele, G., Bedford, T., Ward, M. J., et al. (2014). HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* 346, 56–61. doi: 10.1126/science.1256739

Faria, N. R., Vidal, N., Lourenco, J., Raghwani, J., Sigaloff, K. C. E., Tatem, A. J., et al. (2019). Distinct rates and patterns of spread of the major HIV-1 subtypes in central and East Africa. *PLoS Pathog.* 15:e1007976. doi: 10.1371/journal.ppat.1007976

Gill, M. S., Lemey, P., Faria, N. R., Rambaut, A., Shapiro, B., and Suchard, M. A. (2013). Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol. Biol. Evol.* 30, 713–724. doi: 10.1093/molbev/mss265

Giovanetti, M., Ciccozzi, M., Parolin, C., and Borsetti, A. (2020). Molecular epidemiology of HIV-1 in African countries: a comprehensive overview. *Pathogens* 9:1072. doi: 10.3390/pathogens9121072

Gray, R. R., Tatem, A. J., Lamers, S., Hou, W., Laeyendecker, O., Serwadda, D., et al. (2009). Spatial phylodynamics of HIV-1 epidemic emergence in East Africa. *AIDS* 23, F9–F17. doi: 10.1097/QAD.0b013e32832faf61

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Hadgu, T. G., Egziabher, E., Gizaw, G., Yilma, A., Khodakevich, L., Zewdie, D., et al. (1990). Intersectoral collaboration in AIDS control in Ethiopia. *Ethiop. J. Heal. Dev.* 4, 7–99.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.

Halperin, D. T., Mugurungi, O., Hallett, T. B., Muchini, B., Campbell, B., Magure, T., et al. (2011). A surprising prevention success: why did the HIV epidemic decline in Zimbabwe? *PLoS Med.* 8:e1000414. doi: 10.1371/journal.pmed.1000414

Hassan, A. S., Pybus, O. G., Sanders, E. J., Albert, J., and Esbjörnsson, J. (2017). Defining HIV-1 transmission clusters based on sequence data. *AIDS* 31, 1211–1222. doi: 10.1097/QAD.0000000000001470

Hemelaar, J., Elangovan, R., Yun, J., Dickson-Tetteh, L., Fleminger, I., Kirtley, S., et al. (2019). Global and regional molecular epidemiology of HIV-1, 1990-2015: a systematic review, global survey, and trend analysis. *Lancet Infect. Dis.* 19, 143–155. doi: 10.1016/S1473-3099(18)30647-9

Hessol, N. A., Lifson, A. R., O'Malley, P. M., Doll, L. S., Jaffe, H. W., and Rutherford, G. W. (1989). Prevalence, incidence, and progression of human immunodeficiency virus infection in homosexual and bisexual men in hepatitis B vaccine trials, 1978-1988. *Am. J. Epidemiol.* 130, 1167–1175. doi: 10.1093/oxfordjournals.aje.a115445

Hill, V., and Baele, G. (2019). Bayesian estimation of past population dynamics in BEAST 1.10 using the Skygrid coalescent model. *Mol. Biol. Evol.* 36, 2620–2628. doi: 10.1093/molbev/msz172

Hladik, W., Shabbir, I., Jelaludin, A., Woldu, A., Tsehaynesh, M., and Tadesse, W. (2006). HIV/AIDS in Ethiopia: where is the epidemic heading? *Sex. Transm. Infect.* 82(Suppl. 1), i32–i35. doi: 10.1136/sti.2005.016592

Hué, S., Clewley, J. P., Cane, P. A., and Pillay, D. (2004). HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. *AIDS* 18, 719–728. doi: 10.1097/00002030-200403260-00002

Hué, S., Pillay, D., Clewley, J. P., and Pybus, O. G. (2005). Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proc. Natl. Acad. Sci. U. S. A.* 102, 4425–4429. doi: 10.1073/pnas.0407534102

Jordan, M. R., Bennett, D. E., Bertagnolio, S., Gilks, C. F., and Sutherland, D. (2008). World Health Organization surveys to monitor HIV drug resistance prevention and associated factors in sentinel antiretroviral treatment sites. *Antivir. Ther.* 13(Suppl. 2), 15–23. doi: 10.1186/s13104-016-2101-8

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kebede, D., Aklilu, M., and Sanders, E. (2000). The HIV epidemic and the state of its surveillance in Ethiopia. *Ethiop. Med. J.* 38, 283–302.

Khodakevich, L., Mehret, M., Negassa, H., and Shanko, B. (1990). Progression of human immunodeficeiencyy virus epidemic in Ethiopia. *Ethiop. J. Heal. Dev.* 4, 183–187.

Kloos, H., and Mariam, D. H. (2000). HIV/AIDS in Ethiopia: an overview. *Northeast. Afr. Stud.* 7, 13–40. doi: 10.1353/nas.2004.0006

Lester, F. T., Ayehunie, S., and Zewdie, D. (1988). Acquired immunodeficiency syndrome: seven cases in an Addis Ababa hospital. *Ethiop. Med. J.* 26, 139–145.

Martin, J. L. (1987). The impact of AIDS on gay male sexual behavior patterns in New York City. *Am. J. Public Health* 77, 578–581. doi: 10.2105/AJPH.77.5.578

Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D., and Lefeuvre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26, 2462–2463. doi: 10.1093/bioinformatics/btq467

Mebret, M., Kbodakcvicb, L., Zewdie, D., Ayebunic, S., Gizaw, G., Shanko, B., et al. (1990). HIV-1 infection and related risk factors among female sex workers in urban areas of Ethiopia. *Ethiop. J. Heal. Dev.* 4, 163–170.

Mehret, M. (1990). HIV-1 infection and related risk factors among female sex workers in urban areas in Ethiopia. *Ethiop. J. Health Dev.* 4(Suppl. 2), 163–170.

Mehret, M., Khodakevich, L., Zewdie, D., Gizaw, G., Ayehune, S., Shanko, B., et al. (1990a). HIV-1 infection among employees of the Ethiopian Freightt transport corporation. *Ethiop. J. Heal. Dev.* 4, 177–182.

Mehret, M., Mertens, T. E., Caraël, M., Negassa, H., Feleke, W., Yitbarek, N., et al. (1996). Baseline for the evaluation of an AIDS programme using prevention indicators: a case study in Ethiopia. *Bull. World Health Organ.* 74, 509–516.

Mehret, M. K. L., Shanko, B., and Belete, F. (1990b). Sexual behaviours and some social features off female sex workers in the city of Addis Ababa. *Ethiop. J. Health Dev.* 4, 133–113.

Mehret, M. K. L., Zewdie, D., Ayehunie, S., Shanko, B., Gizaw, G., et al. (1990c). HIV-1 infection and some related risk factors among female sex workers in Addis Ababa. *Ethiop. J. Health Dev.* 4, 171–176.

Mekonnen, Y., Sanders, E., Aklilu, M., Tsegaye, A., Rinke de Wit, T. F., Schaap, A., et al. (2003). Evidence of changes in sexual behaviours among male factory workers in Ethiopia. *AIDS* 17, 223–231. doi: 10.1097/00002030-200301240-00013

Mir, D., Graf, T., de Matos, E., Almeida, S., Pinto, A. R., Delatorre, E., et al. (2018). Inferring population dynamics of HIV-1 subtype C epidemics in eastern Africa and southern Brazil applying different Bayesian phylodynamics approaches. *Sci. Rep.* 8:8778. doi: 10.1038/s41598-018-26824-4

Mount, D. W. (2007). Using the basic local alignment search tool (BLAST). *Cold Spring Harb. Protoc.* 2007:pdb.top17. doi: 10.1101/pdb.top17

Negassa, H., Kefene, H., Khodakevich, L., Zewdie, D., and Shanko, B. (1990). Profile of AIDS case in Ethiopia. *Ethiop. J. Heal. Dev.* 4, 213–217.

Nelson, K. E., Celentano, D. D., Eiumtrakol, S., Hoover, D. R., Beyrer, C., Suprasert, S., et al. (1996). Changes in sexual behavior and a decline in HIV infection among young men in Thailand. *N. Engl. J. Med.* 335, 297–303. doi: 10.1056/NEJM199608013350501

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Okubagzhi, G., and Singh, S. (2002). Establishing an HIV/AIDS programme in developing countries: the Ethiopian experience. *AIDS* 16, 1575–1586. doi: 10.1097/00002030-200208160-00002

Parker, J., Rambaut, A., and Pybus, O. G. (2008). Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect. Genet. Evol.* 8, 239–246. doi: 10.1016/j.meegid.2007.08.001

Pineda-Peña, A. C., Faria, N. R., Imbrechts, S., Libin, P., Abecasis, A. B., Deforche, K., et al. (2013). Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: performance evaluation of the new REGA version 3 and seven other tools. *Infect. Genet. Evol.* 19, 337–348. doi: 10.1016/j.meegid.2013.04.032

Pollakis, G., Abebe, A., Kliphuis, A., De Wit, T. F., Fisseha, B., Tegbaru, B., et al. (2003). Recombination of HIV type 1C (C′/C″) in Ethiopia: possible link of EthHIV-1C′ to subtype C sequences from the high-prevalence epidemics in India and Southern Africa. *AIDS Res. Hum. Retrovir.* 19, 999–1008. doi: 10.1089/088922203322588350

Pybus, O. G., Charleston, M. A., Gupta, S., Rambaut, A., Holmes, E. C., and Harvey, P. H. (2001). The epidemic behavior of the hepatitis C virus. *Science* 292, 2323–2325. doi: 10.1126/science.1058321

Rambaut, A. (2016). FigTree v1.4.3: Tree Figure Drawing Tool. Available at: http://tree.bio.ed.ac.uk/software/figtree/ (Accessed October 4, 2019).

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in Bayesian Phylogenetics using tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032

Rambaut, A., Lam, T. T., Max Carvalho, L., and Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly path-O-gen). *Virus Evol.* 2:vew007. doi: 10.1093/ve/vew007

Sallam, M., Esbjornsson, J., Baldvinsdottir, G., Indriethason, H., Bjornsdottir, T. B., Widell, A., et al. (2017). Molecular epidemiology of HIV-1 in Iceland: early introductions, transmission dynamics and recent outbreaks among injection drug users. *Infect. Genet. Evol.* 49, 157–163. doi: 10.1016/j.meegid.2017.01.004

Schultz, A. K., Zhang, M., Bulla, I., Leitner, T., Korber, B., Morgenstern, B., et al. (2009). jpHMM: improving the reliability of recombination prediction in HIV-1. *Nucleic Acids Res.* 37, W647–W651. doi: 10.1093/nar/gkp371

Sharp, P. M., and Hahn, B. H. (2011). Origins of HIV and the AIDS pandemic. *Cold Spring Harb. Perspect. Med.* 1:a006841. doi: 10.1101/cshperspect.a006841

Stadler, T., Kouyos, R., von Wyl, V., Yerly, S., Boni, J., Burgisser, P., et al. (2012). Estimating the basic reproductive number from viral sequence data. *Mol. Biol. Evol.* 29, 347–357. doi: 10.1093/molbev/msr217

Stadler, T., Kuhnert, D., Bonhoeffer, S., and Drummond, A. J. (2013). Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl. Acad. Sci.* 110, 228–233. doi: 10.1073/pnas.1207965110

Stoneburner, R. L., and Low-Beer, D. (2004). Population-level HIV declines and behavioral risk avoidance in Uganda. *Science* 304, 714–718. doi: 10.1126/science.1093166

Struck, D., Lawyer, G., Ternes, A. M., Schmit, J. C., and Bercoff, D. P. (2014). COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. *Nucleic Acids Res.* 42:e144. doi: 10.1093/nar/gku739

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 4:vey016. doi: 10.1093/ve/vey016

Taaffe, J., Fraser-Hurt, N., Gorgens, M., and Harimurti, P. (2014). A Comprehensive Review of Empirical and Modeled HIV Incidence Trends (1990–2012). Policy Research Working Paper; No. 7042. World Bank. Available at: http://hdl.handle.net/10986/20378 (Accessed March 10, 2021).

Thomson, M. M., and Fernandez-Garcia, A. (2011). Phylogenetic structure in African HIV-1 subtype C revealed by selective sequential pruning. *Virology* 415, 30–38. doi: 10.1016/j.virol.2011.03.021

Tsega, E., Mengesha, B., Nordenfelt, E., Hansson, B. G., and Lindberg, J. (1988). Serological survey of human immunodeficiency virus infection in Ethiopia. *Ethiop. Med. J.* 26, 179–184.

Tsegaye, A., Rinke De Wit, T. F., Mekonnen, Y., Beyene, A., Aklilu, M., Messele, T., et al. (2002). Decline in prevalence of HIV-1 infection and syphilis among young women attending antenatal care clinics in Addis Ababa, Ethiopia: results from sentinel surveillance, 1995-2001. *J. Acquir. Immune Defic. Syndr.* 30, 359–362. doi: 10.1097/00126334-200207010-00013

Tully, D. C., and Wood, C. (2010). Chronology and evolution of the HIV-1 subtype C epidemic in Ethiopia. *AIDS* 24, 1577–1582. doi: 10.1097/QAD.0b013e32833999e1

UNAIDS (2020). Global HIV & AIDS statistics—Fact Sheet. Available at: https://www.unaids.org/en/resources/fact-sheet (Accessed March 10, 2021).

Vasylyeva, T. I., du Plessis, L., Pineda-Pena, A. C., Kuhnert, D., Lemey, P., Vandamme, A. M., et al. (2019). Tracing the impact of public health interventions on HIV-1 transmission in Portugal using molecular epidemiology. *J. Infect. Dis.* 220, 233–243. doi: 10.1093/infdis/jiz085

Wensing, A. M., Calvez, V., Günthard, H. F., Johnson, V. A., Paredes, R., Pillay, D., et al. (2016). 2017 update of the drug resistance mutations in HIV-1. *Top. Antivir. Med.* 24, 132–133.

Wilkinson, E., Engelbrecht, S., and de Oliveira, T. (2015). History and origin of the HIV-1 subtype C epidemic in South Africa and the greater southern African region. *Sci. Rep.* 5:16897. doi: 10.1038/srep16897

Wilkinson, E., Rasmussen, D., Ratmann, O., Stadler, T., Engelbrecht, S., and de Oliveira, T. (2016). Origin, imports and exports of HIV-1 subtype C in South Africa: a historical perspective. *Infect. Genet. Evol.* 46, 200–208. doi: 10.1016/j.meegid.2016.07.008

Wolday, D., Meles, H., Hailu, E., Messele, T., Mengistu, Y., Fekadu, M., et al. (2007). Temporal trends in the incidence of HIV infection in antenatal clinic attendees in Addis Ababa, Ethiopia, 1995-2003. *J. Intern. Med.* 261, 132–137. doi: 10.1111/j.1365-2796.2006.01740.x

Woods, C. K., Brumme, C. J., Liu, T. F., Chui, C. K., Chu, A. L., Wynhoven, B., et al. (2012). Automating HIV drug resistance genotyping with RECall, a freely accessible sequence analysis tool. *J. Clin. Microbiol.* 50, 1936–1942. doi: 10.1128/JCM.06689-11

Yusim, K., Peeters, M., Pybus, O. G., Bhattacharya, T., Delaporte, E., Mulanga, C., et al. (2001). Using human immunodeficiency virus type 1 sequences to

infer historical features of the acquired immune deficiency syndrome epidemic and human immunodeficiency virus evolution. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 356, 855–866. doi: 10.1098/rstb.2001.0859

Zewdie, D., Gizaw, G., Khodakevich, L., Degifie, G., and Wemeue, M. (1990). Development and management of the AIDS control programme in Ethiopia. *Ethiop. J. Heal. Dev.* 4, 87–96.

Author Disclaimer: The findings and conclusions in this paper are those of the authors and do not necessarily represent the official position of the funding agencies.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Arimide, Esquivel-Gómez, Kebede, Sasinovich, Balcha, Björkman, Kühnert and Medstrand. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Paper II

PLOS ONE

RESEARCH ARTICLE

HIV-genetic diversity and drug resistance transmission clusters in Gondar, Northern Ethiopia, 2003-2013

Dawit Assefa Arimide[1,2], Almaz Abebe[2], Yenew Kebede[3], Fekadu Adugna[4], Tesfaye Tilahun[2], Desta Kassa[2], Yibeltal Assefa[5], Taye Tolera Balcha[6], Per Björkman[1], Patrik Medstrand[1]*

1 Department of Translational Medicine, Lund University, Malmö, Sweden, **2** The Ethiopian Public Health Institute, Addis Ababa, Ethiopia, **3** Center for Disease Control and Prevention (CDC)-Ethiopia, Addis Ababa, Ethiopia, **4** World Health Organization (WHO)-Ethiopia, Addis Ababa, Ethiopia, **5** School of Public Health, University of Queensland, Brisbane, Australia, **6** Arman Hansen research Institute (AHRI), Addis Ababa, Ethiopia

* patrik.medstrand@med.lu.se

Check for updates

OPEN ACCESS

Citation: Arimide DA, Abebe A, Kebede Y, Adugna F, Tilahun T, Kassa D, et al. (2018) HIV-genetic diversity and drug resistance transmission clusters in Gondar, Northern Ethiopia, 2003-2013. PLoS ONE 13(10): e0205446. https://doi.org/10.1371/journal.pone.0205446

Editor: Jason Blackard, University of Cincinnati College of Medicine, UNITED STATES

Received: June 7, 2018

Accepted: September 25, 2018

Published: October 10, 2018

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. All sequences reported in this study have been deposited in Genbank under Accession Numbers MH324937-MH325003.

Funding: This work received support from Grant number: NA; CDC-Ethiopia: https://www.cdc.gov/globalhealth/countries/ethiopia/; Grant number: NA; WHO-Ethiopia: http://www.afro.who.int/countries/ethiopia; Grant number: 2015-05842; Vetenskapsrådet, Sweden: https://www.vr.se/; and

Abstract

Background

The HIV-1 epidemic in Ethiopia has been shown to be dominated by two phylogenetically distinct subtype C clades, the Ethiopian (C'-ET) and East African (C-EA) clades, however, little is known about the temporal dynamics of the HIV epidemic with respect to subtypes and distinct clades. Moreover, there is only limited information concerning transmission of HIV-1 drug resistance (TDR) in the country.

Methods

A cross-sectional survey was conducted among young antiretroviral therapy (ART)-naïve individuals recently diagnosed with HIV infection, in Gondar, Ethiopia, 2011–2013 using the WHO recommended threshold survey. A total of 84 study participants with a median age of 22 years were enrolled. HIV-1 genotyping was performed and investigated for drug resistance in 67 individuals. Phylogenetic analyses were performed on all available HIV sequences obtained from Gondar (n = 301) which were used to define subtype C clades, temporal trends and local transmission clusters. Dating of transmission clusters was performed using BEAST.

Result

Four of 67 individuals (6.0%) carried a HIV drug resistance mutation strain, all associated with non-nucleoside reverse transcriptase inhibitors (NNRTI). Strains of the C-EA clade were most prevalent as we found no evidence of temporal changes during this time period. However, strains of the C-SA clade, prevalent in Southern Africa, have been introduced in Ethiopia, and became more abundant during the study period. The oldest Gondar transmission clusters dated back to 1980 (C-EA), 1983 (C-SA) and 1990 (C'-ET) indicating the

Grant number: NA; Medical Faculty, Lund University, Sweden: https://www.med.lu.se/. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

presence of strains of different subtype C clades at about the same time point in Gondar. Moreover, some of the larger clusters dated back to the 1980s but transmissions within clusters have been ongoing up till end of the study period. Besides being associated with more sequences and larger clusters, the C-EA clade sequences were also associated with clustering of HIVDR sequences. One cluster was associated with the G190A mutation and showed onward transmissions at high rate.

Conclusion

TDR was detected in 6.0% of the sequenced samples and confirmed pervious reports that the two subtype C clades, C-EA and C'-ET, are common in Ethiopia. Moreover, the findings indicated an increased diversity in the epidemic as well as differences in transmission clusters sizes of the different clades and association with resistance mutations. These findings provide epidemiological insights not directly available using standard surveillance and may inform the adjustment of public health strategies in HIV prevention in Ethiopia.

Introduction

The global scale up of antiretroviral therapy (ART) has resulted in decline in HIV related morbidity, mortality and HIV transmission. In most low and middle income countries (LMIC) standardized first line antiretroviral regimens are used, consisting of two nucleoside reverse transcriptase inhibitor (NRTI) and one non-nucleoside reverse transcriptase inhibitor (NNRTI)] [1–3]. The emergence of HIV drug resistance (HIVDR), particularly towards drugs with low genetic barriers, eg. NNRTIs, has been shown to increase with time after introduction of ART programs [4]. One explanation for this is likely to be the lack of routine monitoring of plasma viral load, which has not been scaled up at the same rate as ART service expansion (ART roll-out) [2–9]. This leads to delayed identification of patients with treatment failure, with risk of accumulation of drug resistance mutations (DRM) in such individuals. Furthermore, individuals with unrecognized virological treatment failure are potential reservoirs for onwards transmission of viruses with DRM (commonly referred to as transmitted drug resistance; TDR) [2, 9–11]. If this occurs in populations with high incidence and at high risk of onward transmissions, the prevalence of drug resistant HIV strains may increase and be further amplified in the population [12]. Thus, the emergence of drug resistant viral strains constitutes a threat to the outcome of ART programs [13, 14]. Virological monitoring and resistance surveillance is therefore a priority [2].

The first HIV positive sera and AIDS case was diagnosed in Ethiopia in 1984 and 1986, respectively [15, 16]. By the late 1980s a high prevalence of HIV-1 was detected among commercial sex workers and among long distance truck drivers [17]. ART was introduced in the public health sector in 2003, with free ART provision since 2005 [18]. In 2017, an estimated ~740,000 individuals were living with HIV and ~426,000 had initiated ART [19]. With the scale up and decentralization of the service in Ethiopia, the emergence and transmission of resistance is expected, as has been evident in other LMIC [4–7]. Following the WHO recommendation, Ethiopia implemented a strategy for the prevention and monitoring of HIVDR to maximize the durable efficacy of affordable and potent first-line ART regimens [20].

A threshold survey was performed according to WHO guideline in the capital Addis Ababa, the city where ART was first started in Ethiopia, among treatment naïve women attending antenatal clinics in 2005. This survey revealed no major drug resistance for any

available HIV drug class [21]. A nationwide study 2009–2011 indicated a pretreatment DR (PDR) level of 3.9%[22]. Moreover, three studies evaluated PDR among patients in Gondar (located in Northern Ethiopia) [23–25]., Even though these studies were not performed according to the WHO threshold surveillance method, they indicated an increase of PDR after ART roll out in the country.

The Ethiopian HIV epidemic is dominated by HIV subtype C, similar to the countries of Southern Africa, and although Ethiopia shares borders with countries were subtypes A and D are common, these subtypes have rarely been identified in Ethiopia [26]. Previous studies have shown that two distinct subtype C strains are co-circulating in Ethiopia, designated C and C' [27], and recombinant forms of the C and C' strains [28]. Ten distinct subtype C clades (termed C1-C10) have been defined based on phylogenetic relationship. Clades C1-C9 were mainly represented by sequences obtained from countries in southern Africa while the C10 clade represented sequences from East and Central Africa[29]. Further molecular characterization revealed that the Ethiopian C strain were similar to strains circulating in other East African countries while the C' strain were shown to represented a distinct subclade of a southern African clade. Thus, based on previous phylogenetic studies, the African subtype C stains can be divided into three major groups: the southern African subtype C clades (C-SA), the Ethiopian C' clade (C'-ET), and the central and east African subtype C clade (C-EA) [29, 30]. Previous phylodynamic studies have indicated that HIV-1 was introduced in Ethiopia in the late 1960s to the early 1970s, more than 10 years before the first documented AIDS case [30, 31].

The aim of this study was to estimate the prevalence of TDR in young adults with assumed recent HIV-1 infection using the WHO threshold surveillance method, to describe the molecular epidemiology of HIV-1 in terms of genetic diversity, transmission clusters and DRM transmissions within clusters in Gondar, Ethiopia.

Methods

Study design and site selection

A cross-sectional survey was conducted between August 2011 and December 2013 among antiretroviral-naive adults to evaluate transmitted drug resistance, according to the World Health Organization (WHO)-recommended threshold survey methodology [13]. It was conducted in two of the major Voluntary Counselling and Testing (VCT) clinics in Gondar, located 700 km north of Addis Ababa. Gondar is the second largest town in Amhara region and, it was one of the areas in the country where public ART was first initiated (in year 2003). The HIV-prevalence 2003–2013 in Gondar was on a stable level (mean: 10.6%) while the national prevalence during the same time period showed a declining trend, from 12% to 4.4% [32]. Furthermore, three studies conducted at the hospital in 2003–2010 showed the presence of HIVDR in the area [23–25]. The sample size for the current survey followed the sequential sampling method selected by WHO for the surveillance of transmitted HIVDR in low-resource settings. According to the recommendation, it is advised to collect about 70 specimens of eligible individuals consecutively diagnosed with HIV in sites within a survey area [20].

Study participants

Individuals with new diagnosed HIV infection, among VCT clients at the two survey sites were asked to participate in the study. By adopting the WHO recommended inclusion criterion, participants 18–25 years old, who had no prior history of HIV/AIDS-related illness and no history of ART, resident of Gondar for more than a year, and had no history of previous pregnancy were consecutively enrolled. After obtaining written informed consent, 10 ml of blood was collected by venepuncture. Screening for HIV was done using point of care rapid

testing format employed for HIV diagnosis in Ethiopia. This algorithm uses HIV (1 + 2) Antibody Colloidal Gold (KHB, Shanghai Kehua Bio-engineering Co Ltd, China) as a screening test, followed by HIV 1/2 STAT-PAK (Chembio Diagnostics, USA) if positive. In cases with negative STAT-PAK results following a positive KHB test, a third test, Unigold HIV (Trinity Biotech, Ireland), was used as confirmation. Specimens were transported to Ethiopia Public Health Institute, National HIV laboratory, Addis Ababa (a WHO accredited laboratory for genotyping) on dry ice for long term storage at -80˚C until genotyping. At the time of blood sampling basic demographic and clinical information, including age, gender, and history of HIV test was collected using a standardized questionnaire.

HIV-1 genotyping

A 1084-bp fragment of HIV-1 *pol* (corresponding to the position: 2243–3326 of HXB2, Genbank Accession Number: K03455) comprising amino acids 6–99 of the protease (PR) and 1–251 of the reverse transcriptase (RT) was amplified using an in-house genotyping assay as described in [33], and S1 Text. PCR products were directly sequenced using six primers (three on each strand) on an ABI 3100 or an ABI 3500xl DNA Genetic Analyzer (Applied Biosystems). Sequence assembly and editing were performed using RECall V 2.0 HIV-1 sequencing analysis tool [34].

Identification of drug resistance mutations

Surveillance drug resistance mutations (SDRMs) were examined according to the Stanford Genotypic Resistance calibrated population resistance (CPR) tool version 6.0 based on the WHO surveillance transmitted drug resistance mutation list of 2009 [35, 36]. Classification of TDR level (low: $< 5\%$, moderate: 5–15%, or high: $>15\%$) was made based on the WHO threshold survey protocol [20].

HIV-1 subtyping and recombination analysis

Sequence quality control was performed using the online Quality Control program of the Los Alamos HIV sequence database (hiv.lanl.gov). The REGA and Comet online subtyping tools were used for initial classification into subtypes and inter-subtype recombinants [37, 38]. Putative intra-subtype recombinants were verified using Simplot ver. 3.5.1[39]. All sequences were also screened for recombination using RDP ver. 3.44 [40]. Final subtyping was performed through phylogenetic analysis using the reference HIV-1 data set from Los Alamos HIV sequence database (hiv.lanl.gov). Sequences were aligned using ClustalX2 [41] and then edited to a final length of 1044 bases using BioEdit v4.0.6 (http://www.mbio.ncsu.edu/bioedit/bioedit.html). A Maximum likelihood (ML) phylogenetic tree was constructed using the online version of PhyML with the GTR+I+Γ nucleotide substitution model (using estimated proportion of invariable sites and four gamma categories) and NNI plus SPR to estimate the tree topology. Branch support was determined with aLRT-SH (approximate likelihood ratio test Shimodaira-Hasegawa like) implemented in PhyML. A branch in the phylogeny with an aLRT-SH value $\geq$0.9 was considered significant [42, 43].

To further dissect the subtype C distribution in Gondar, sub-subtyping into subtype C clades and detection of putative intra-subtype C recombinants were performed using sequences obtained from this study population and all previously reported HIV-1 *pol* sequences from Gondar [23–25]. Briefly, reference sequences of the three subtype C clades were obtained as described in S1 Text. Putative intra-subtype recombinant subtype C sequences were identified by jpHMM [44] using parameters as outlined in S1 Text. The final subtyping ML tree is shown in S1 Fig.

Clade distribution and transmission cluster analysis

For the purpose of a comprehensive analysis of temporal changes of subtype C clades over time in Gondar, subtype C clades were identified using ML phylogenetic tree analysis. For the purpose of transmission cluster analysis, we followed a methodology described previously and included a data set of similar sequences from GenBank by identifying the ten best scoring GenBank sequences with BLAST for each of the Gondar sequences in the study [43, 45, 46]. The final data set contained 491 sequences (n = 299 from Gondar and n = 192 GenBank reference sequences; see S2 Table) which were used to define local (Gondar) transmission clusters. Transmission clusters were defined as described previously [43, 45–48]. Briefly, a transmission cluster was defined as a cluster in the ML phylogeny from root to tips. A cluster with an aLRT SH-support of ≥0.9 that had a majority (at least 80%) of sequences from Gondar was considered as a Gondar (local) transmission cluster. Transmission clusters were defined based on their sizes (number of sequences/cluster), into dyads (two sequences), medium sized clusters/networks (3–14 sequences) and large clusters (≥15 sequences) [43, 49]. Clusters sharing >33% of the same DRM were defined as a putative drug resistance transmission clusters [50].

Evolutionary and phylodynamic analysis

To estimate an evolutionary rate for each of the three clades which were needed to perform dated cluster analysis using phylodynamic analyses, three data sets were assembled, containing both global and Ethiopian sequences which were considered to be representative for the different clades. This approach has been used previously for HIV-1 subtype B [51, 52]. The analysis was performed using the sequences shown in S1 Fig and S2 Table, by randomly selecting a subset of sequences from each clade (N = 86, N = 65 and N = 70 for C-EA, C'-ET and C-SA, respectively). A maximum of one sequence of each transmission cluster was allowed in the final data set. For the data set containing the southern African clade (C-SA) sequences, the temporal signal was weak (R^2 = 0.03), as assessed by root-to-tip analysis using TempEst, indicating that the data set was not optimal for estimating reliable substitution rates, while the R^2 was higher for the C-EA and C'-ET clades (R^2 = 0.35 and 0.22, respectively), indicating a better temporal signal in the datasets [53]. We therefore estimated the substitution rates for the C-EA and C'-ET sequences only and relied on previous estimates for the C-SA clades (see below). For these analyses we centered the prior mean rate at 0.001 substitutions/site/year and specified the standard deviation to 0.33 on a lognormal distribution such that the 2.5 and 97.5 percentiles of the distribution contained the rates obtained previously of HIV-1 *pol* of subtype C [54, 55]. The evolutionary rate was estimated by employing the Bayesian Markov Chains Monte Carlo (MCMC) clock method implemented in the BEAST software package v1.8.4 [56]. We estimated the evolutionary rate using both a strict and relaxed Bayesian MCMC clock, in both cases with a flexible demographic model (the Bayesian skyline plot) as a tree prior. This analysis estimated posterior distribution of median rates for the C-EA and C'-ET clades to 1.26 x 10^{-3} (95% HPD: $8.80x10^{-4}$–$1.64x10^{-3}$) and $1.44x10^{-3}$ (95% HPD: $9.70x10^{-4}$–$1.91x10^{-3}$) substitutions/site/year, respectively (S1 Table). Using these rates as priors, subsequent BEAST analysis was performed on the C-EA and C'-ET sequences obtained from Gondar only while a previous evolutionary rate of 2.15x10-3 ($1.79x10^{-3}$–$2.60x10^{-3}$, 95% HPD) substitutions/site/year was used as a prior for the C-SA sequences of Gondar [54]. In all cases, a lognormal distribution on the rate prior was employed, as described above, such that the prior mean rate was centered at each clade-specific mean rate and that the 2.5 and 97.5 percentiles of the distribution contained the rates obtained from the above analyses (S1 Table). For all analyses we used both strict and relaxed clock models and specified a codon position partitioning model (the SRD06 model) for all data sets [57]. We also employed different demographic tree priors on

each data set, the skyride, the logistic and exponential demographic models implemented in BEAST (S1 Table). Model comparisons were done by Bayes factor (BF) analysis of marginal likelihoods [58]. BF values larger than 3 and 5 were considered to represent strong and very strong evidence respectively, against H_0 [59].

Statistical analysis

Statistical tests were performed using SPSS 24 (IBM Corp., Armonk, NY, USA). DR prevalence was determined with a confidence interval (CI) of 95% using the Wilson method. Categorical variables were compared using 2-tailed Fisher's exact test, while continuous variables were compared using Mann-Whitney 2-tailed U test. Trends over time were analysed using linear-by-linear test for association.

Ethical approval

Scientific and ethical approval was granted by the Research and Ethical Clearance Committee of the Ethiopian Public Health Institute, and the National Health Research Ethics Review Committee of Ministry of Science and Technology of Ethiopia. All participants provided written informed consent.

Availability of data

Nucleotide sequences reported in this study have been deposited in the Genbank repository (Accession Numbers: KM390990—KM391398).

Results

Population characteristics

A total of 84 individuals attending the VCT clinics at Gondar University hospital (n = 64) and Gondar health centre (n = 20) met the inclusion criteria and were enrolled in the study.

All samples were confirmed as HIV-1 positive. In total, 78% (67/84) of the samples were successfully amplified and sequenced. The median age among the 84 individuals was 22 years (IQR: 20–23) and 83% were females, which was similar to the age (22 years [IQR: 20–24]) and gender distribution (87% females) of the 67 participants whose samples were successfully genotyped (p = 0.992 and p = 0.653, M-W and Fisher exact test, respectively).

Levels of transmitted drug resistance

TDR levels were determined from the 67 sequenced samples. Four specimens (6.0%) carried major DRMs, corresponding to a moderate TDR level [60]. Two had K103N, one G190S and one Y181C, all representing resistance to non-nucleoside reverse transcriptase inhibitors (NNRTI). No NRTI or protease inhibitor (PI) associated DRMs were identified among the sequenced samples (Table 1). Although there were more females enrolled in the survey than men there were no association between gender and DRMs (2 men and 2 females carried HIVDR virus; p = 0.084, FET).

Level of HIV-1 drug resistance in Gondar 2003–2013

Analysis of HIV DRMs of two previous studies from Gondar after ART roll out in 2009 and 2010 as described in Methods, revealed levels comparable to our threshold study (4.4–8.2% versus 6.0%; Table 1) [23, 25]. These estimates were higher than the levels reported in the study performed before ART roll-out in Gondar 2003 (3.3%) [24]. However, the confidence

Table 1. Prevalence of HIV drug resistance mutations in Gondar 2003–2013.

Drug class and DRM[1]	Populations[2]			
	Before ART, 2003 (n = 92)	**General, 2009 (n = 158)**	**General, 2010 (n = 61)**	**Young (present study), 2011–2013 (n = 67)**
All classes	3 (**3.3**, 1.1–9.2)	8 (**5.1**, 2.6–9.7)	5 (**8.2**, 3.6–17.8)	4 (**6.0**, 2.4–14.4)
PI	1 (1.0, 0.-5.9)	2 (1.3, 0.4–4.5)	1 (1.6, 0.3–8.7)	0 (0.0, 0.0–5.4)
M46I	-	1	1	-
F53L	-	1	-	-
I85V	1	-	-	-
NRTI	0 (0.0, 0.0–4.0)	3 (1.9, 0.7–5.4)	1 (1.6, 0.3–8.7)	0 (0.0, 0.0–5.4)
D67E	-	1	-	-
M184I	-	-	1	-
L210W	-	2	-	-
NNRTI	2 (2.2, 0.6–7.6)	3 (1.9, 0.7–5.4)	3 (4.9, 1.7–13.5)	4 (6.0, 2.4–14.4)
K101E	-	1	-	-
K103N	-	-	-	2
Y181I	-	-	1	-
Y181C	-	-	-	1
G190A	2	2	1	-
G190E	-	-	1	-
G190S	-	-	-	1

1: Drug class: PI: protease inhibitors; NRTI: nucleoside reverse transcriptase inhibitor; NNRTI: non-nucleoside reverse transcriptase inhibitors; DRM: drug resistance mutation. Mutations were defined by the Stanford Genotypic Resistance Interpretation Algorithm (http://hivdb.stanford.edu/pages/algs/HIVdb.html) using the calibrated population resistance (CPR) tool version 6.0 (http://cpr.stanford.edu/cpr/servlet/CPR), based on the WHO surveillance transmitted drug resistance mutation list of 2009.

2: Study Population, year of sample collection (number of samples genotyped). Number of drug resistance mutations for all or per class (**% DRM**, 95% CI) or number of specific drug resistance mutations. Nucleotide sequences were obtained from this and previous studies in Gondar [23–25]; 2003, Kassu et al. (2007); 2009, Mulu et al. (2009); 2010, Huruy et al. (2015); 2011–2013, this study.

https://doi.org/10.1371/journal.pone.0205446.t001

intervals of the levels of DRMs were largely overlapping between the studies and the proportion of DRMs before and after ART roll out were not different (p = 0.584, FET; Table 1). DRMs of the previous three studies were as in the present study, most often associated with NNRTIs. Taken together, the G190A/S/E mutations were most prominent (seven of the 12 NNRTI mutations), followed by the K103N, Y181I/C and K101E substitutions (two, two and one, respectively). Mutations conferring resistance to NRTIs and PIs were found in earlier studies at low levels (1.0–1.9%) but were not identified in the present study (Table 1).

Subtype C clade distribution and temporal changes 2003–2013

Phylogenetic subtyping revealed that 94% (63 of 67) of study participants enrolled 2011–2013 were infected with HIV-1 subtype C, while the remaining four study participants had subtype A, B or A/C recombinant sequences (S2 Table). Since previous observations have suggested more rapid expansion of the C'-ET strains than the C-EA strains in Ethiopia, we further investigated the distribution of our sequences into different subtype C clades [61]. Detailed analysis indicated that 59 of 63 sequences represented the three major subtype C clades: C-EA (n = 30), C'-ET (n = 16) and C-SA (n = 13), while four sequences were putative C-EA/C'-ET recombinants (S2 Table). For in-depth phylogenetic analysis, we next constructed a new data set that represented non-recombinant subtype C sequences of the current study (n = 59), the three

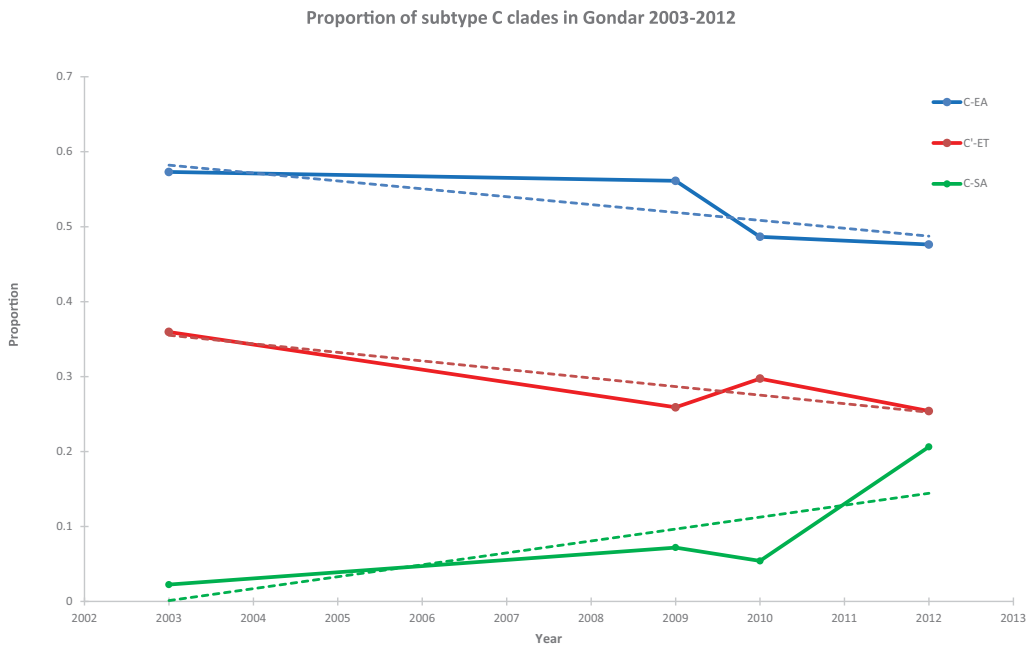


Fig 1. Trend analysis of subtype C clades 2003–2013. Proportion of different subtype C clades in Gondar 2003–2013. A dotted line indicates the trend over time. The 2012 time point included specimens obtained 2011–2013 (current study).

https://doi.org/10.1371/journal.pone.0205446.g001

previous studies from Gondar in 2003–2010 (n = 242) and a Genbank reference data set (n = 190), resulting in a final data set of 491 sequences collected 1986–2013 (n = 301 from Gondar; S2 Table). Phylogenetic subtyping using all 491 sequences revealed that sequences fell into one of the three subtype C clades (S1 Fig). Among the 301 sequences from Gondar, 59% (n = 177) were classified as C-EA, 32% (n = 97) as C'-ET and 9% (n = 27) as C-SA. Notably, sequences of the C-SA clade have not been described in Ethiopia previously and the proportion of this clade in Gondar increased from 2% in 2003 to 23% in 2012 ($p<0.001$, two-tailed linear-by-linear test for association) while there was a modest decline of the two major clades (C-EA: 60%-51%; C'-ET: 38%-26%; Fig 1) during the same time period. However, the proportion of C-EA remained higher than C'-ET and C-SA at each time point 2003–2013. Further analysis of the temporal changes revealed that there was no changes of the proportions of the C-EA and C'-ET clades in Gondar during the same time period ($p = 0.146$ and $p = 0.173$, respectively, two-tailed linear-by-linear test for association). Among the sequences with DRMs identified in Gondar (Table 1 and S2 Table), 13 were classified as belonging to the C-EA clade, two to the each of the C'-ET and C-SA clades, however, there were no association of DRMs to either of the clades (C-EA vs C'-ET, $p = 0.095$; C-EA vs C-SA, $p = 1.000$; and C'-ET vs C-SA, $p = 0.210$; FET for all comparisons).

Transmission cluster analysis and association with drug resistance mutations

Transmission cluster analysis was performed separately for the C-EA and C-SA/C'-ET sequences by constructing new ML trees of the two data sets (Fig 2). The C-EA phylogeny indicated that the sequences from Gondar as well as sequences from other cities in Ethiopia intermixed with the global dataset (mostly with sequences from other East African countries and Europe). The C-SA/C'-ET phylogeny showed a clear separation of the two clades similar to that seen for the subtyping phylogeny of the whole data set (S1 Fig). While the C-SA clade mostly contained sequences of southern African countries, sequences of Gondar were intermixed in the phylogeny. The C'-ET clade consisted mostly of Ethiopian, followed by sequences obtained in Europe and North America. Gondar transmission clusters were defined as statistically supported phylogenetic clusters dominated by Gondar sequences (see Methods, Table 2 and S3 Table). In total, we identified 28 clusters (which together contained 35% [105 of 301] of the sequences from Gondar) which was indicative for multiple introductions of HIV into Gondar followed by local spread (21 dyads, six medium sized clusters/networks and one large clusters; Fig 2, Table 2 and S3 Table). The Gondar C-EA sequences were found more frequently in transmission clusters compared to the C'-ET sequences (38% [68 of 177 sequences] in 15 local transmission clusters versus 23% [22 out of 97 sequences] in 11 clusters); $p = 0.010$, FET). Likewise, C-SA sequences were found more frequently in clusters (56% [15 of 27]) compared to C'-ET ($p = 0.002$, FET) while there were no difference between the number of C-EA and C-SA sequences in cluster ($p = 0.098$, FET). The mean cluster size for C-EA was five Gondar sequences/cluster (range: 2–28). Nine C-EA clusters had two sequences/cluster (dyads), five had $\geq$3–8 sequences/cluster while one cluster was large and contained 32 sequences (Fig 2). All C'-ET clusters were represented as dyads while the C-SA clusters was represented by one dyad and one cluster with 13 members. Among C-EA, six of 15 clusters contained >2 sequences which was higher than for C'-ET (none of the 11 clusters had more than two sequences; $p = 0.020$, FET) but this difference was not apparent between C-EA and C-SA (one cluster contained >2 sequences; $p = 1.000$, FET). Nine of the 15 C-EA clusters, seven of the 12 C'-ET clusters and one of the two C-SA clusters contained sequences of participants obtained in more than one of the surveys conducted in Gondar 2003–2013 indicated that clusters were populated over time in Gondar (Table 2). Moreover, eight of the 28 clusters (three C-EA clusters, three C'-ET and both of the C-SA clusters) had members from the last survey conducted 2011–2013, indicating recent transmissions within these clusters.

Dating of transmission clusters of different subtype C clades in Gondar

To study the timing of introductions and differences between the different subtype C clades clusters, we performed Bayesian coalescent analysis using BEAST with sequences obtained from Gondar only (n = 301). Since the Maximum likelihood phylogeny showed that the strains of the different clades in Gondar were dispersed among sequences of other origins and most likely represented multiple introductions, the estimated root tMRCAs of each subtype C clade represent the date of the origin of the circulating subtype C clades in the region (although only Gondar sequences were used in the date estimations) while the estimated tMRCA of the transmission clusters should approximate introductions and local spread of the viral strains in Gondar, cf references [54, 62]. The age of the transmission clusters ranged from estimated dates of 1980–2009. The oldest clusters were estimated to 1980, 1983 and 1990 for C-EA, C-SA and C'-ET, respectively (Table 2 and S1 Table). Fifteen of the 28 clusters had estimated tMRCAs before year 2000. The 15 C-EA and two C-SA transmission clusters had the oldest estimated median tMRCAs (median dates: 1986 and 1991, respectively), while the 11 C'-ET clusters were

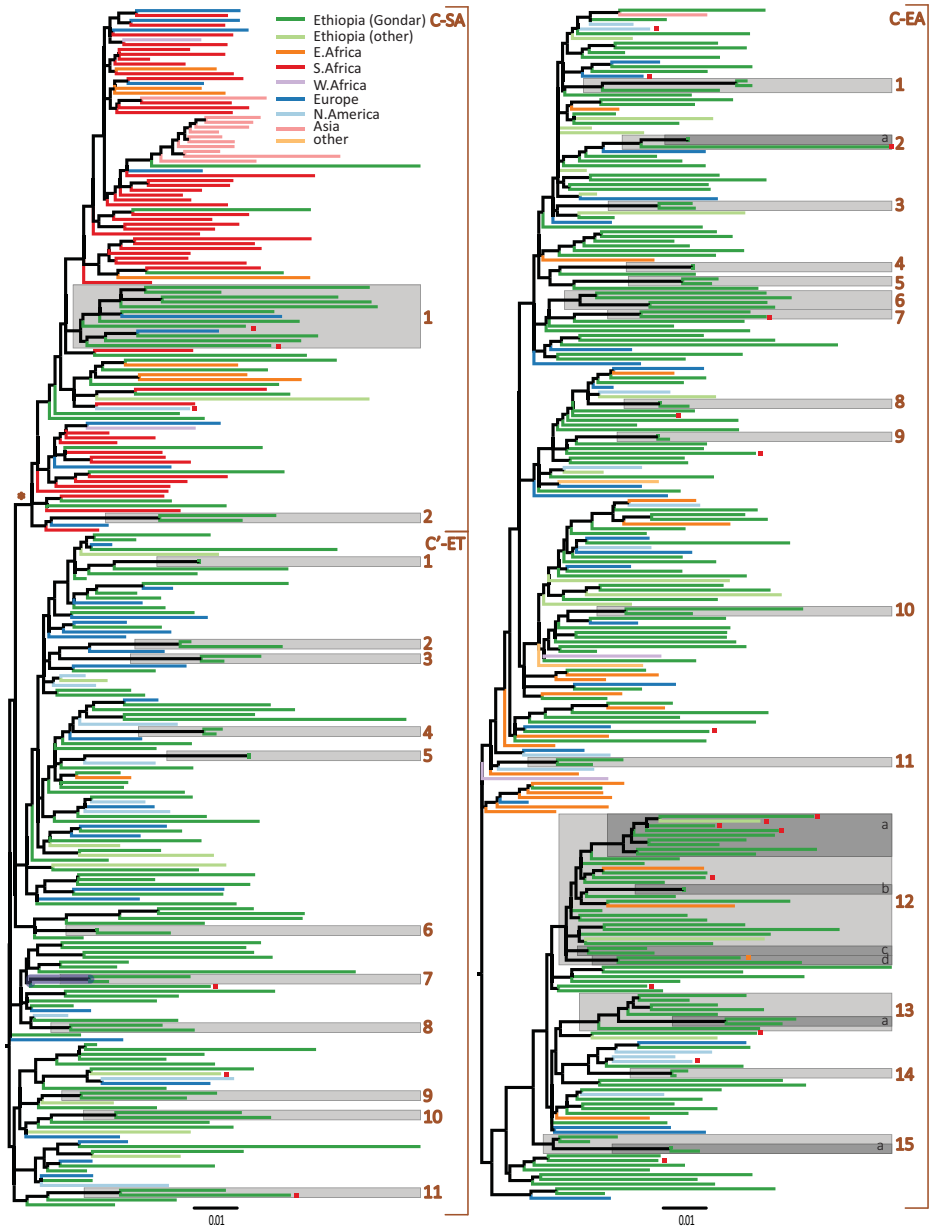

Ethiopia (Gondar)
Ethiopia (other)
E.Africa
S.Africa
W.Africa
Europe
N.America
Asia
other
C-SA
C'-ET
C-EA
0.01
0.01

Fig 2. Maximum likelihood phylogenetic tree. Monophyletic clusters are shown in greyed boxes, defined as clusters with a branch support (aLRT-SH) >0.9 and with >80% sequences obtained from Gondar. Tips are coloured according to collection place. The colour code is indicated in the top of the left panel. A red square at a tip indicates a DRM sequence. (Left panel) The C-SA and C'-ET clades. The branch separating the two clades are indicated with an asterisk and is represented by a branch support of >0.9. (Right panel) The C-EA clade phylogeny.

https://doi.org/10.1371/journal.pone.0205446.g002

younger (median date: 2001). Two C'-ET clusters were young (≤5 years) whereas the large C-EA and network C-SA clusters were estimated to be 28 and 30 years with estimated median tMRCAs of 1985 and 1983, respectively. Taken together, this indicated that the C-EA and C-SA was introduced earlier into Gondar compared to the C'-ET strain. The median population evolutionary rate of C-EA was higher than for C'-ET and C-SA (2.29 $x10^{-3}$ and $1.85x10^{-3}$ versus $1.70x10^{-3}$ substitutions/site/year, respectively), however the 95% credible intervals overlapped (S1 Table).

Transmission cluster associated with drug resistance mutations

Sequences with DRMs were found more frequently in clusters among the C-EA compared to the C'-ET sequences (57% versus 25%; $p<0.001$, FET). The C-EA Cluster 12 represented the largest cluster and contained several sub-clusters indicating a large transmission network in Gondar over an extended time period. In subcluster 12a, four of nine sequences carried the G190A mutation (Table 2 and Fig 2) which suggested cluster-associated transmission of this resistance mutation in Gondar. To verify the age estimate of cluster 12 we obtained dated phylogenies using the Bayesian MCMC method by using sequences of this cluster only in the analysis (S3 Table). The median estimated evolutionary rate of the cluster was 1.34×10^{-3} substitutions/site/year, slower than the median rate estimate of the entire C-EA cluster but with overlapping credible intervals (S1 Table). Moreover, the median tMRCA of cluster 12 was estimated to 1980 (95% HPD: 1969–1989) which was similar to the estimate (1985; 95% HPD: 1967–1985) obtained from the analysis of all Gondar sequences (Table 2 and S1 Table). Since the internode intervals in the MCMC tree represent the maximum transmission time intervals, we estimated the time between transmissions events. Analysis of the internode intervals (from the MCMC tree; S2 Fig) revealed that the maximum median transmission intervals for cluster 12 were 1.58 years (IQR: 1.24–2.31). The median parametric estimate of the population growth rate also indicated frequent transmissions associated with sequences of cluster 12 at approximately 0.5 new infections/individual/year which correspond to a median doubling time of infections every 1.34 years (16 months). The demographic plot analysis indicated that this cluster increased in size and the number of effective infections, N_e, reached its peak in 2003, but declined thereafter (from 4990 in 2003 to 2404 in 2013; S1 Table). However, the four sequences that had the G190A mutation were collected during three of the four surveillance studies in Gondar 2003–2013 (Table 2 and Fig 2) suggesting onward transmission of this DRM over many years in Gondar.

Discussion

This study represents the first threshold survey for DRM performed among young ART-naïve HIV-1 positive individuals in Gondar, Northern Ethiopia, using the WHO threshold methodology. We also used the sequence data obtained from drug resistance analysis together with sequence data obtained from previous studies for a detailed molecular epidemiological investigation of the HIV epidemic in Gondar. Detailed analysis stratified on the three subtype C clades identified transmission clusters in Gondar 2003–2013, which comprised 35% of all available sequences during this time. This finding indicated that HIV-1 has been introduced on multiple occasions, followed by local transmissions. Dated phylogenies revealed that about

Table 2. Characteristics of Gondar transmission clusters.

	Members in Clusters						
Cluster[1]	**Total**	**From Gondar**	**Outside Gondar**	**Collection Years**	**DRM[2]**	**tMRCA (age) [3]**	**tMRCA (calender year) [4]**
C-EA-1	3	3	0	2009		30 (21–40)	1983 (1973–1992)
C-EA-2	3	3	0	2003, 2009	F53L	27 (20–37)	1986 (1976–1993)
C-EA-2a	2	2	0	2003		10 (10–12)	2003 (2001–2003)
C-EA-3	2	2	0	2009		12 (7–18)	2001 (1995–2006)
C-EA-4	2	2	0	2003		10 (10–11)	2003 (2002–2003)
C-EA-5	2	2	0	2009		10 (7–16)	2003 (1997–2006)
C-EA-6	4	4	0	2009, 2010, 2011		33 (24–44)	1980 (1969–1989)
C-EA-7	2	2	0	2003, 2009	D67E (1)	27 (19–37)	1986 (1976–1994)
C-EA-8	2	2	0	2009		8 (5–12)	2005 (2001–2008)
C-EA-9	2	2	0	2003, 2009		11 (10–12)	2002 (2001–2003)
C-EA-10	2	2	0	2003, 2009		26 (19–36)	1987 (1977–1994)
C-EA-11	2	2	0	2003, 2009		16 (12–21)	1997 (1992–2001)
C-EA-12	32	28	4	2003, 2009, 2010, 2011, 2012	G190A (4), Y181C (1)	28 (23–34)	1985 (1979–1990)
C-EA-12a	9	8	1	2003, 2009, 2010, 2011	G190A (4)	23 (19–28)	1990 (1985–1994)
C-EA-12b	2	2	0	2003, 2009		10 (10–11)	2003 (2002–2003)
C-EA-12c	2	2	0	2009, 2012	Y181C (1)	21 (16–26)	1992 (1987–1997)
C-EA-12d	2	2	0	2009		23 (17–29)	1990 (1984–1996)
C-EA-13	8	8	0	2003, 2009, 2010, 2011		26 (21–32)	1987 (1981–1992)
C-EA-13a	2	2	0	2009		13 (10–18)	2000 (1995–2003)
C-EA-14	2	2	0	2009		6 (4–9)	2007 (2004–2009)
C-EA-15	4	4	0	2003, 2009		22 (17–27)	1991 (1986–1996)
C-EA-15a	2	2	0	2009		6 (4–9)	2007 (2004–2009)
C-SA-1	13	13	0	2003, 2009, 2010, 2011, 2012,	K103N (1), G190E (1)	30 (23–39)	1983 (1974–1990)
C-SA-2	2	2	0	2011, 2012		13 (8–19)	2000 (1994–2005)
C'-ET-1	2	2	0	2009		23 (15–30)	1990 (1983–1998)
C'-ET-2	2	2	0	2010, 2012		23 (16–29)	1990 (1984–1997)
C'-ET-3	2	2	0	2009, 2010		21 (15–27)	1992 (1986–1998)
C'-ET-4	2	2	0	2009		4 (4–6)	2009 (2007–2009)
C'-ET-5	2	2	0	2009		10 (5–15)	2003 (1998–2008)
C'-ET-6	2	2	0	2003, 2009		10 (7–15)	2003 (1998–2006)
C'-ET-7	2	2	0	2003, 2010		7 (5–10)	2006 (2003–2008)
C'-ET-8	2	2	0	2010, 2012		4 (4–5)	2009 (2008–2009)
C'-ET-9	2	2	0	2010, 2011		12 (10–16)	2001 (1997–2003)
C'-ET-10	2	2	0	2009		16 (11–24)	1997 (1989–2002)
C'-ET-11	2	2	0	2003, 2009	L210W+M46I (1)	22 (17–27)	1991 (1986–1996)

[1]Cluster refers to the subtype C clades (C-EA, C-SA or C'-ET) followed by cluster number as shown in Fig 2. Clusters were defined as having an aLRT-SH support of ≥0.9 and containing at least 80% of sequences collected in Gondar.

[2]DRM: Drug resistance mutations. Mutations were defined by the Stanford Genotypic Resistance Interpretation Algorithm (http://hivdb.stanford.edu/pages/algs/HIVdb.html) using the calibrated population resistance (CPR) tool version 6.0 (http://cpr.stanford.edu/cpr/servlet/CPR), based on the WHO surveillance transmitted drug resistance mutation list of 2009.

[3]tMRCA (age): time to the most recent common ancestor. Indicated is median ages, and in parenthesis, the 95% highest posterior density credible interval. Cluster tMRCAs was determined using BEAST v1.8.4 with a logistic tree prior. Priors and other model parameters are indicated in Methods and S1 Table.

[4]tMRCA (calender year): time to the most recent common ancestor. Calendar years were obtained by subtracting the tMRCA age estimate from the most recent sampled sequence (2013) used in the analysis.

https://doi.org/10.1371/journal.pone.0205446.t002

half of the local clusters originated before 2000. However, several of the clusters were long lasting and in some cases ongoing active transmission chains were detected. Importantly, we show that the G190A mutation has spread in Gondar by rapid transmission within local clusters. Even though DRM transmission within clusters has been described in other parts of the world, this is, to our knowledge, the first example of cluster-associated DRM transmission in sub Saharan Africa.

Eight to ten years after ART roll out in 2003 we found a moderate level of TDR in Gondar which is in agreement with observations of increased TDR prevalence after roll-out of ART from different world regions [5, 7, 63]. The overall prevalence of TDR (6%), found in this study is comparable to the 5.6% prevalence in sub-Saharan Africa 6–8 years after ART roll-out, and the 7.4% prevalence estimate in East Africa, eight years after ART roll-out [4, 6, 63]. Two previous studies conducted in Gondar in 2009–2010, and in the country 2009–2011, 6–8 years after ART roll out showed DRM levels of 4–6% [22, 23, 25]. Despite long term ART administration in the area there was not a significant difference in the DRM level. A direct comparison of the temporal DRM levels presented here is not possible, since the former studies targeted older age groups as opposed to our threshold study that included a younger population of HIV infected individuals. Thus, in Gondar the DRM prevalence among different age groups were not different although TDR differences between age groups and gender has been reported from other sub Saharan countries [19, 64]. However, even moderate levels of TDR among young sexually active individuals should raise concerns as it may result in less effective ART in a considerable proportion of recently HIV-1 infected individuals.

The DRMs detected in this study were all associated with the NNRTIs efavirenz (EFV) and nevirapine (NVP). This finding is not unexpected considering the low genetic barrier of these drugs to the development of resistance and their wide use as part of first line ART regimen [6]. The specific NNRTI associated mutations found in this study (K103N, G109S and Y181C) have been reported to account for the most common NNRTI-associated mutations in all world regions and HIV subtypes [65]. These mutations have also been found among patients failing treatment in Gondar, indicating a link between acquired and transmitted drug resistance [25, 66, 67]. Furthermore, strains with NNRTIs mutations may persist for a long time before being replaced by wild type making the likelihood of persistence after transmission higher [68]. In line with this, previous studies have reported significant global increases of NNRTI TDR over time since ART rollout, especially in East Africa, with an estimated 36% per year increases after ART roll out [6]. Compared to the previous studies conducted in Gondar we did not find DRMs associated with PIs and NRTIs, which could be attributed to a generally low sample size often seen in threshold studies. Thus, NNRTIs are associated with both acquired and transmitted resistance in Ethiopia, as is the case in many other LMICs, and represents an obstacle to the long-term success of ART and control of HIV transmission. As more individuals initiate ART, these observations advocate that more robust and durable first line regimens and/or improved ART monitoring are required [69].

Phylogenetic analysis revealed multiple introductions of HIV in Gondar, followed by local transmission, a phenomenon that has been described previously for other local HIV epidemics [46, 54, 70]. Our study confirms previous observations and argues against the hypothesis that a single initial lineage was introduced and became responsible for the subtype C epidemic in the country [31]. Our study also confirms pervious reports that the two subtype C clades, C-EA and C'-ET, are dominant in Ethiopia [27]. We found that the C-EA clade has been the most prevalent HIV clade in Gondar 2003–2013 since it represented the major circulating strain in Gondar at the end of the study period, as we found no evidence of significant temporal changes during this time period. However, we also show that the HIV strains of the C-SA clades, the most prevalent HIV subtype C strains prevalent in Southern Africa, have been

introduced in Ethiopia, and became increasingly prevalent in Gondar 2003–2013. Sequences belonging to the C-EA clade were found to be more prevalent in clusters and were also associated with larger clusters compared to the sequences of the C'-ET clade indicating a variable intensity of the C-EA and C'-ET epidemics in Gondar. The C-EA and C'-ET strains were introduced from East Africa and Southern Africa respectively, at approximately the same time into Ethiopia (1978–1981)[30]. Our evolutionary estimates were in the range of previous estimates of subtype C and the oldest Gondar transmission clusters dated back to 1980 (C-EA), 1983 (C-SA) and 1990 (C'-ET) supporting the presence of strains of different subtype C clades at about the same time point in Gondar [30, 54]. We were able to identify old transmission clusters, since our definition of clusters followed an approach, which omitted the use of a distance cut-off, and thereby not excluding long lasting transmission chains [43, 45, 48]. Our finding showed that some clusters were long-lived and could have played an important impact in fuelling the local HIV epidemic. The larger old clusters (C-SA cluster 1 and C-EA cluster 12) had been populated until recently since the initial introduction in Gondar in the 1980s. Long-lived, large clusters have been shown to be associated with transmission of HIV in groups with high risk behaviour [43, 47]. This finding was further supported by the observation that the majority of clusters contained sequences isolated in two or more of the surveys conducted in Gondar 2003–2013. The propensity of CEA sequences to be part of clusters compared to the C'-ET could be associated with sociodemographic factors and/or underling biological differences of the CEA and C'-ET viruses. Recent infection, age, gender, drug resistance and unawareness of infection status of sexually active individuals have been shown to represent risk factors associated with transmission potential and cluster size [43, 47, 71–74]. Transmission cluster of populations with high risk of HIV acquisition are larger compared to that populations with lower risk behaviour, for example between men who have sex with men (MSM) and heterosexual networks in Europe and USA, and among recently infected individuals [43, 47, 73, 75–78]. It is also interesting to note that all transmission clusters among the C'-ET sequences represented dyads, that is, a pattern associated with less onward transmission that may represent sporadic transmission among couples and/or individuals with lower number of partners [47]. Since we had no data regarding risk behaviours of the study participants, further studies are needed to establish any association between risk behaviours and gender as well as other sociodemographic factors with clustering of the different subtype C strains in Ethiopia. Moreover, it has previously been suggested that biological differences between the two Ethiopian strains may be coupled to transmissibility due to a generally higher viral load among patients infected with C'-ET which has been suggested as an explanation for an observed abundance of the C'-ET strain in Ethiopia in previous studies [61, 79]. However, our results indicated a higher prevalence of C-EA in Gondar 2003–2013 and further studies are needed to establish any association between biological properties, replicative capacity, disease progression, risk behaviours and transmission rates among the strains of the Ethiopian subtype C clades [79, 80]

Besides being associated with more sequences and larger clusters, the CEA clade was also associated with clustering of DRM sequences. The role of DRMs in clustered transmission has been addressed in a number of studies. Previous studies of the Swiss HIV cohort, indicated that TDR viruses were more associated with clustering compared to non-TDR viruses [73]. An association between clustering and increased transmission of viruses harbouring NNRTI DRMs, including the G190A DRM found here, has been described in Quebec, Canada. Transmission clustering were also associated with sexual behaviour, mainly that of MSM [19]. In the present study, we identified one cluster with the G190A mutation with onward transmissions of this DRM for at least 8 years (2003–2010). The G190A mutation represents a slowly reverting HIVDR mutation and when such mutations are present in a population with frequent transmission it might persist and expand [67, 81]. We estimated that the entire large cluster

and the G190A cluster had transmission intervals at least every year. Since not all members of a given transmission cluster have been sampled, the average time between transmissions is overestimated since the transmission chain is incomplete, suggesting that transmissions have taken place at a much higher rate [82]. In fact, many molecular epidemiology studies suffer from low sampling, representing only a fraction of all infected individuals at the time of the study. The 301 sequences from Gondar (population: ~207,000) 2003–2013 were estimated to roughly represent 1.3% of the infected population during the study period (considering a mean prevalence of 10.6% 2003–2014; [32] and Methods). Using a linear projection the cluster could be up to 75 times larger, i.e cluster 12 with 32 individual specimens may represent 2400 individual or more in the cluster (which could be compared with the estimated number of effective infections in 2013 of 2400 of the demographic plot analysis; S1 Table) and the G190A associated subcluster 12a of about 450 individuals. Although these numbers represent a rough estimate it illustrates the impact of fast onward transmission that is commonly seen among groups with recent infections and/or with high risk of HIV acquisitions. Fast transmission times may favour the accumulation of TDR mutations especially when transmission times are faster than the average time of reversions of the mutations.

In conclusion, our study showed a moderate prevalence of TDR in Gondar. We could also demonstrate a hitherto unrecognized diversity of the HIV-1 epidemic in the area, and confirm pervious reports that the two subtype C clades, C-EA and C'-ET, are dominant in Ethiopia but our study showed an increased diversity in the epidemic since we showed multiple introductions and that viral strains of the C-SA clade, prevalent in southern Africa, has been introduced and increased in prevalence in Gondar 2003–2013. We also show differences in transmission clusters sizes of the different subtype C clades and association with DRMs. These findings provide epidemiological insights not available using standard surveillance approaches and may inform the improvement of public health strategies in HIV prevention in Ethiopia and similar LMIC settings.

Supporting information

S1 Text. HIV-1 subtyping, identification of intra-subtype recombinants and construction of a reference data set.
(DOCX)

S1 Table. Population dynamics and evolutionary estimates of subtype C clades in Gondar.
(DOCX)

S2 Table. Taxa included in the phylogenetic analysis.
(DOCX)

S3 Table. Taxa associated with clusters.
(DOCX)

S1 Fig. Phylogenetic subtyping of sequences obtained from Gondar 2003–2013.
(DOCX)

S2 Fig. Maximum clade credibility tree of C-EA Cluster 12.
(DOCX)

Acknowledgments

We express our appreciation to the study participants, and the staffs of the VCT clinics. We also acknowledge the support given to us by Gondar university hospital and HIV Research laboratory team.

Author Contributions

Conceptualization: Dawit Assefa Arimide, Almaz Abebe, Yenew Kebede, Fekadu Adugna, Desta Kassa, Yibeltal Assefa, Patrik Medstrand.

Data curation: Dawit Assefa Arimide, Tesfaye Tilahun.

Formal analysis: Dawit Assefa Arimide, Patrik Medstrand.

Funding acquisition: Patrik Medstrand.

Investigation: Taye Tolera Balcha, Per Björkman, Patrik Medstrand.

Methodology: Dawit Assefa Arimide, Tesfaye Tilahun, Per Björkman, Patrik Medstrand.

Project administration: Tesfaye Tilahun, Yibeltal Assefa.

Supervision: Almaz Abebe, Yenew Kebede, Fekadu Adugna, Desta Kassa, Taye Tolera Balcha, Per Björkman, Patrik Medstrand.

Validation: Patrik Medstrand.

Writing – original draft: Dawit Assefa Arimide, Patrik Medstrand.

Writing – review & editing: Per Björkman, Patrik Medstrand.

References

1. WHO. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach. http://apps.who.int/iris/bitstream/10665/208825/1/9789241549684_eng.pdf?ua=1. 2016. PubMed PMID: 24716260.
2. Kityo C, Thompson J, Nankya I, Hoppe A, Ndashimye E, Warambwa C, et al. HIV Drug Resistance Mutations in Non-B Subtypes After Prolonged Virological Failure on NNRTI-Based First-Line Regimens in Sub-Saharan Africa. Journal of acquired immune deficiency syndromes. 2017; 75(2):e45–e54. Epub 2017/01/28. https://doi.org/10.1097/QAI.0000000000001285 PMID: 28129253; PubMed Central PMCID: PMCPMC5427983.
3. Gilks CF, Crowley S, Ekpini R, Gove S, Perriens J, Souteyrand Y, et al. The WHO public-health approach to antiretroviral treatment against HIV in resource-limited settings. Lancet. 2006; 368 (9534):505–10. Epub 2006/08/08. https://doi.org/10.1016/S0140-6736(06)69158-7 PMID: 16890837.
4. Hamers RL, Wallis CL, Kityo C, Siwale M, Mandaliya K, Conradie F, et al. HIV-1 drug resistance in antiretroviral-naive individuals in sub-Saharan Africa after rollout of antiretroviral therapy: a multicentre observational study. The Lancet Infectious diseases. 2011; 11(10):750–9. Epub 2011/08/02. https://doi.org/10.1016/S1473-3099(11)70149-9 PMID: 21802367.
5. Aghokeng AF, Kouanfack C, Laurent C, Ebong E, Atem-Tambe A, Butel C, et al. Scale-up of antiretroviral treatment in sub-Saharan Africa is accompanied by increasing HIV-1 drug resistance mutations in drug-naive patients. Aids. 2011; 25(17):2183–8. https://doi.org/10.1097/QAD.0b013e32834bbbe9 PMID: 21860346.
6. Gupta RK, Jordan MR, Sultan BJ, Hill A, Davis DH, Gregson J, et al. Global trends in antiretroviral resistance in treatment-naive individuals with HIV after rollout of antiretroviral treatment in resource-limited settings: a global collaborative study and meta-regression analysis. Lancet. 2012; 380(9849):1250–8. Epub 2012/07/26. https://doi.org/10.1016/S0140-6736(12)61038-1 PMID: 22828485; PubMed Central PMCID: PMCPMC3790969.
7. Hamers RL, Kityo C, Lange JM, de Wit TF, Mugyenyi P. Global threat from drug resistant HIV in sub-Saharan Africa. Bmj. 2012; 344:e4159. Epub 2012/06/20. https://doi.org/10.1136/bmj.e4159 PMID: 22709963.
8. Rawizza HE, Chaplin B, Meloni ST, Eisen G, Rao T, Sankale JL, et al. Immunologic criteria are poor predictors of virologic outcome: implications for HIV treatment monitoring in resource-limited settings. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America. 2011; 53(12):1283–90. https://doi.org/10.1093/cid/cir729 PMID: 22080121; PubMed Central PMCID: PMC3246873.
9. Sethi AK, Celentano DD, Gange SJ, Moore RD, Gallant JE. Association between adherence to antiretroviral therapy and human immunodeficiency virus drug resistance. Clinical infectious diseases: an

official publication of the Infectious Diseases Society of America. 2003; 37(8):1112–8. Epub 2003/10/03. https://doi.org/10.1086/378301 PMID: 14523777.

10. Sigaloff KC, Ramatsebe T, Viana R, de Wit TF, Wallis CL, Stevens WS. Accumulation of HIV drug resistance mutations in patients failing first-line antiretroviral treatment in South Africa. AIDS research and human retroviruses. 2012; 28(2):171–5. Epub 2011/08/09. https://doi.org/10.1089/aid.2011.0136 PMID: 21819219.
11. van Zyl GU, van der Merwe L, Claassen M, Zeier M, Preiser W. Antiretroviral resistance patterns and factors associated with resistance in adult patients failing NNRTI-based regimens in the Western Cape, South Africa. Journal of medical virology. 2011; 83(10):1764–9. Epub 2011/08/13. https://doi.org/10.1002/jmv.22189 PMID: 21837793.
12. Beyrer C, Pozniak A. HIV Drug Resistance—An Emerging Threat to Epidemic Control. N Engl J Med. 2017; 377(17):1605–7. https://doi.org/10.1056/NEJMp1710608 PMID: 29069566.
13. Bennett DE, Bertagnolio S, Sutherland D, Gilks CF. The World Health Organization's global strategy for prevention and assessment of HIV drug resistance. Antiviral therapy. 2008; 13 Suppl 2:1–13. Epub 2008/06/26. PMID: 18578063.
14. Cambiano V, Bertagnolio S, Jordan MR, Lundgren JD, Phillips A. Transmission of drug resistant HIV and its potential impact on mortality and treatment outcomes in resource-limited settings. The Journal of infectious diseases. 2013; 207 Suppl 2:S57–62. Epub 2013/05/25. https://doi.org/10.1093/infdis/jit111 PMID: 23687290.
15. Lester FT, Ayehunie S, Zewdie D. Acquired immunodeficiency syndrome: seven cases in an Addis Ababa hospital. Ethiop Med J. 1988; 26(3):139–45. Epub 1988/07/01. PMID: 3416846.
16. Tsega E, Mengesha B, Nordenfelt E, Hansson BG, Lindberg J. Serological survey of human immunodeficiency virus infection in Ethiopia. Ethiop Med J. 1988; 26(4):179–84. Epub 1988/10/01. PMID: 3215178.
17. Van Blerk L. AIDS, mobility and commercial sex in Ethiopia: Implications for policy. AIDS Care. 2007; 19(1):79–86. https://doi.org/10.1080/09540120600805091 PMID: 17129861.
18. Ministry of Health Ethiopia. Guideline for implementation of antiretroviral therapy in Ethiopia http://www.etharc.org/oromia/resources/publication/ethartguide.pdf. 2005.
19. Kasang C, Kalluvya S, Majinge C, Stich A, Bodem J, Kongola G, et al. HIV drug resistance (HIVDR) in antiretroviral therapy-naive patients in Tanzania not eligible for WHO threshold HIVDR survey is dramatically high. PloS one. 2011; 6(8):e23091. https://doi.org/10.1371/journal.pone.0023091 PMID: 21886779; PubMed Central PMCID: PMC3158766.
20. Bennett DE, Myatt M, Bertagnolio S, Sutherland D, Gilks CF. Recommendations for surveillance of transmitted HIV drug resistance in countries scaling up antiretroviral treatment. Antiviral therapy. 2008; 13 Suppl 2:25–36. Epub 2008/06/26. PMID: 18575189.
21. Abegaz WE, Grossman Z, Wolday D, Ram D, Kaplan J, Sibide K, et al. Threshold survey evaluating transmitted HIV drug resistance among public antenatal clinic clients in Addis Ababa, Ethiopia. Antiviral therapy. 2008; 13 Suppl 2:89–94. Epub 2008/06/26. PMID: 18575196.
22. Telele NF, Kalu AW, Gebre-Selassie S, Fekade D, Abdurahman S, Marrone G, et al. Pretreatment drug resistance in a large countrywide Ethiopian HIV-1C cohort: a comparison of Sanger and high-throughput sequencing. Scientific reports. 2018; 8(1):7556. https://doi.org/10.1038/s41598-018-25888-6 PMID: 29765082.
23. Huruy K, Maier M, Mulu A, Liebert UG. Limited increase in primary HIV-1C drug resistance mutations in treatment naive individuals in Ethiopia. Journal of medical virology. 2015; 87(6):978–84. https://doi.org/10.1002/jmv.24110 PMID: 25649964.
24. Kassu A, Fujino M, Matsuda M, Nishizawa M, Ota F, Sugiura W. Molecular epidemiology of HIV type 1 in treatment-naive patients in north Ethiopia. AIDS research and human retroviruses. 2007; 23(4):564–8. Epub 2007/04/25. https://doi.org/10.1089/aid.2006.0270 PMID: 17451346.
25. Mulu A, Lange T, Liebert UG, Maier M. Clade homogeneity and Pol gene polymorphisms in chronically HIV-1 infected antiretroviral treatment naive patients after the roll out of ART in Ethiopia. BMC infectious diseases. 2014; 14:158. Epub 2014/03/25. https://doi.org/10.1186/1471-2334-14-158 PMID: 24655349; PubMed Central PMCID: PMCPMC3976149.
26. Abebe A, Kuiken CL, Goudsmit J, Valk M, Messele T, Sahlu T, et al. HIV type 1 subtype C in Addis Ababa, Ethiopia. AIDS research and human retroviruses. 1997; 13(12):1071–5. Epub 1997/08/10. https://doi.org/10.1089/aid.1997.13.1071 PMID: 9264295.
27. Abebe A, Pollakis G, Fontanet AL, Fisseha B, Tegbaru B, Kliphuis A, et al. Identification of a genetic subcluster of HIV type 1 subtype C (C') widespread in Ethiopia. AIDS research and human retroviruses. 2000; 16(17):1909–14. Epub 2000/12/16. https://doi.org/10.1089/08892220050195865 PMID: 11118076.
28. Pollakis G, Abebe A, Kliphuis A, De Wit TF, Fisseha B, Tegbaru B, et al. Recombination of HIV type 1C (C'/C") in Ethiopia: possible link of EthHIV-1C' to subtype C sequences from the high-prevalence

epidemics in India and Southern Africa. AIDS research and human retroviruses. 2003; 19(11):999–1008. Epub 2003/12/18. https://doi.org/10.1089/088922203322588350 PMID: 14678607.

29. Thomson MM, Fernandez-Garcia A. Phylogenetic structure in African HIV-1 subtype C revealed by selective sequential pruning. Virology. 2011; 415(1):30–8. Epub 2011/04/22. https://doi.org/10.1016/j.virol.2011.03.021 PMID: 21507449.

30. Delatorre EO, Bello G. Phylodynamics of HIV-1 subtype C epidemic in east Africa. PloS one. 2012; 7(7):e41904. Epub 2012/08/01. https://doi.org/10.1371/journal.pone.0041904 PMID: 22848653; PubMed Central PMCID: PMCPMC3407063.

31. Tully DC, Wood C. Chronology and evolution of the HIV-1 subtype C epidemic in Ethiopia. Aids. 2010; 24(10):1577–82. https://doi.org/10.1097/QAD.0b013e32833999e1 PMID: 20539092; PubMed Central PMCID: PMC2898272.

32. The Ethiopian Public Health Institute. Report on the 2014 Round Antenatal Care based Sentinel HIV Surveillance in Ethiopia. https://www.ephi.gov.et/images/pictures/2014roundANCbasedHIVsurveillancereport.pdf. 2015.

33. Zhou Z, Wagar N, DeVos JR, Rottinghaus E, Diallo K, Nguyen DB, et al. Optimization of a low cost and broadly sensitive genotyping assay for HIV-1 drug resistance surveillance and monitoring in resource-limited settings. PloS one. 2011; 6(11):e28184. https://doi.org/10.1371/journal.pone.0028184 PMID: 22132237; PubMed Central PMCID: PMCPMC3223235.

34. Woods CK, Brumme CJ, Liu TF, Chui CK, Chu AL, Wynhoven B, et al. Automating HIV drug resistance genotyping with RECall, a freely accessible sequence analysis tool. Journal of clinical microbiology. 2012; 50(6):1936–42. https://doi.org/10.1128/JCM.06689-11 PMID: 22403431; PubMed Central PMCID: PMCPMC3372133.

35. Bennett DE, Camacho RJ, Otelea D, Kuritzkes DR, Fleury H, Kiuchi M, et al. Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. PloS one. 2009; 4(3):e4724. https://doi.org/10.1371/journal.pone.0004724 PMID: 19266092; PubMed Central PMCID: PMCPMC2648874.

36. Gifford RJ, Liu TF, Rhee SY, Kiuchi M, Hue S, Pillay D, et al. The calibrated population resistance tool: standardized genotypic estimation of transmitted HIV-1 drug resistance. Bioinformatics. 2009; 25(9):1197–8. https://doi.org/10.1093/bioinformatics/btp134 PMID: 19304876; PubMed Central PMCID: PMCPMC2672634.

37. Struck D, Lawyer G, Ternes AM, Schmit JC, Bercoff DP. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. Nucleic acids research. 2014; 42(18):e144. Epub 2014/08/15. https://doi.org/10.1093/nar/gku739 PMID: 25120265; PubMed Central PMCID: PMCPMC4191385.

38. Pineda-Pena AC, Faria NR, Imbrechts S, Libin P, Abecasis AB, Deforche K, et al. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: performance evaluation of the new REGA version 3 and seven other tools. Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2013; 19:337–48. Epub 2013/05/11. https://doi.org/10.1016/j.meegid.2013.04.032 PMID: 23660484.

39. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. Journal of virology. 1999; 73(1):152–60. Epub 1998/12/16. PMID: 9847317; PubMed Central PMCID: PMCPMC103818.

40. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics. 2010; 26(19):2462–3. Epub 2010/08/28. https://doi.org/10.1093/bioinformatics/btq467 PMID: 20798170; PubMed Central PMCID: PMCPMC2944210.

41. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ. Multiple sequence alignment with Clustal X. Trends in biochemical sciences. 1998; 23(10):403–5. Epub 1998/11/12. PMID: 9810230.

42. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology. 2010; 59(3):307–21. Epub 2010/06/09. https://doi.org/10.1093/sysbio/syq010 PMID: 20525638.

43. Esbjornsson J, Mild M, Audelin A, Fonager J, Skar H, Bruun Jorgensen L, et al. HIV-1 transmission between MSM and heterosexuals, and increasing proportions of circulating recombinant forms in the Nordic Countries. Virus Evol. 2016; 2(1):vew010. Epub 2016/10/25. https://doi.org/10.1093/ve/vew010 PMID: 27774303; PubMed Central PMCID: PMCPMC4989887.

44. Schultz AK, Zhang M, Bulla I, Leitner T, Korber B, Morgenstern B, et al. jpHMM: improving the reliability of recombination prediction in HIV-1. Nucleic acids research. 2009; 37(Web Server issue):W647–51. Epub 2009/05/16. https://doi.org/10.1093/nar/gkp371 PMID: 19443440; PubMed Central PMCID: PMCPMC2703979.

45. Hassan AS, Pybus OG, Sanders EJ, Albert J, Esbjornsson J. Defining HIV-1 transmission clusters based on sequence data. Aids. 2017; 31(9):1211–22. Epub 2017/03/30. https://doi.org/10.1097/QAD.0000000000001470 PMID: 28353537; PubMed Central PMCID: PMCPMC5482559.

46. Sallam M, Esbjornsson J, Baldvinsdottir G, Indriethason H, Bjornsdottir TB, Widell A, et al. Molecular epidemiology of HIV-1 in Iceland: Early introductions, transmission dynamics and recent outbreaks among injection drug users. Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2017; 49:157–63. https://doi.org/10.1016/j.meegid.2017.01.004 PMID: 28082188.

47. Kouyos RD, von Wyl V, Yerly S, Boni J, Taffe P, Shah C, et al. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. The Journal of infectious diseases. 2010; 201(10):1488–97. Epub 2010/04/14. https://doi.org/10.1086/651951 PMID: 20384495.

48. Patino-Galindo JA, Torres-Puente M, Bracho MA, Alastrue I, Juan A, Navarro D, et al. The molecular epidemiology of HIV-1 in the Comunidad Valenciana (Spain): analysis of transmission clusters. Scientific reports. 2017; 7(1):11584. https://doi.org/10.1038/s41598-017-10286-1 PMID: 28912478; PubMed Central PMCID: PMC5599654.

49. Aldous JL, Pond SK, Poon A, Jain S, Qin H, Kahn JS, et al. Characterizing HIV transmission networks across the United States. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America. 2012; 55(8):1135–43. Epub 2012/07/13. https://doi.org/10.1093/cid/cis612 PMID: 22784872; PubMed Central PMCID: PMCPMC3529609.

50. Wertheim JO, Oster AM, Johnson JA, Switzer WM, Saduvala N, Hernandez AL, et al. Transmission fitness of drug-resistant HIV revealed in a surveillance system transmission network. Virus Evol. 2017; 3(1):vex008. Epub 2017/05/02. https://doi.org/10.1093/ve/vex008 PMID: 28458918; PubMed Central PMCID: PMCPMC5399924.

51. Hue S, Pillay D, Clewley JP, Pybus OG. Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102(12):4425–9. Epub 2005/03/16. https://doi.org/10.1073/pnas.0407534102 PMID: 15767575; PubMed Central PMCID: PMCPMC555492.

52. Salemi M, de Oliveira T, Ciccozzi M, Rezza G, Goodenow MM. High-resolution molecular epidemiology and evolutionary history of HIV-1 subtypes in Albania. PloS one. 2008; 3(1):e1390. https://doi.org/10.1371/journal.pone.0001390 PMID: 18167549; PubMed Central PMCID: PMC2148102.

53. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016; 2(1):vew007. Epub 2016/10/25. https://doi.org/10.1093/ve/vew007 PMID: 27774300; PubMed Central PMCID: PMCPMC4989882.

54. Dalai SC, de Oliveira T, Harkins GW, Kassaye SG, Lint J, Manasa J, et al. Evolution and molecular epidemiology of subtype C HIV-1 in Zimbabwe. Aids. 2009; 23(18):2523–32. Epub 2009/09/23. https://doi.org/10.1097/QAD.0b013e3283320ef3 PMID: 19770693; PubMed Central PMCID: PMCPMC2923658.

55. Wilkinson E, Rasmussen D, Ratmann O, Stadler T, Engelbrecht S, de Oliveira T. Origin, imports and exports of HIV-1 subtype C in South Africa: A historical perspective. Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2016; 46:200–8. Epub 2016/07/17. https://doi.org/10.1016/j.meegid.2016.07.008 PMID: 27421210.

56. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution. 2012; 29(8):1969–73. Epub 2012/03/01. https://doi.org/10.1093/molbev/mss075 PMID: 22367748; PubMed Central PMCID: PMCPMC3408070.

57. Shapiro B, Rambaut A, Drummond AJ. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Molecular biology and evolution. 2006; 23(1):7–9. https://doi.org/10.1093/molbev/msj021 PMID: 16177232.

58. Suchard MA, Weiss RE, Sinsheimer JS. Bayesian selection of continuous-time Markov chain evolutionary models. Molecular biology and evolution. 2001; 18(6):1001–13. https://doi.org/10.1093/oxfordjournals.molbev.a003872 PMID: 11371589.

59. Kass RE, Raftery AE. Bayes Factors. Journal of the American Statistical Association. 1995; 90(430):773–95. https://doi.org/10.1080/01621459.1995.10476572

60. Myatt M, Bennett DE. A novel sequential sampling technique for the surveillance of transmitted HIV drug resistance by cross-sectional survey for use in low resource settings. Antiviral therapy. 2008; 13 Suppl 2:37–48. Epub 2008/06/26. PMID: 18575190.

61. Ayele W, Mekonnen Y, Messele T, Mengistu Y, Tsegaye A, Bakker M, et al. Differences in HIV type 1 RNA plasma load profile of closely related cocirculating Ethiopian subtype C strains: C and C'. AIDS research and human retroviruses. 2010; 26(7):805–13. Epub 2010/07/14. https://doi.org/10.1089/aid.2009.0152 PMID: 20624072.

62. Esbjornsson J, Mild M, Mansson F, Norrgren H, Medstrand P. HIV-1 molecular epidemiology in Guinea-Bissau, West Africa: origin, demography and migrations. PloS one. 2011; 6(2):e17025. https://doi.org/10.1371/journal.pone.0017025 PMID: 21365013; PubMed Central PMCID: PMC3041826.

63. Hamers RL, Sigaloff KC, Kityo C, Mugyenyi P, de Wit TF. Emerging HIV-1 drug resistance after roll-out of antiretroviral therapy in sub-Saharan Africa. Curr Opin HIV AIDS. 2013; 8(1):19–26. Epub 2012/11/13. https://doi.org/10.1097/COH.0b013e32835b7f94 PMID: 23143140.

64. Silverman RA, Beck IA, Kiptinness C, Levine M, Milne R, McGrath CJ, et al. Prevalence of Pre-antiretroviral-Treatment Drug Resistance by Gender, Age, and Other Factors in HIV-Infected Individuals Initiating Therapy in Kenya, 2013–2014. The Journal of infectious diseases. 2017; 216(12):1569–78. https://doi.org/10.1093/infdis/jix544 PMID: 29040633; PubMed Central PMCID: PMC5853791.

65. Rhee SY, Blanco JL, Jordan MR, Taylor J, Lemey P, Varghese V, et al. Geographic and temporal trends in the molecular epidemiology and genetic mechanisms of transmitted HIV-1 drug resistance: an individual-patient- and sequence-level meta-analysis. PLoS medicine. 2015; 12(4):e1001810. Epub 2015/04/08. https://doi.org/10.1371/journal.pmed.1001810 PMID: 25849352; PubMed Central PMCID: PMCPMC4388826.

66. Tilghman MW, Perez-Santiago J, Osorio G, Little SJ, Richman DD, Mathews WC, et al. Community HIV-1 drug resistance is associated with transmitted drug resistance. HIV medicine. 2014; 15(6):339–46. https://doi.org/10.1111/hiv.12122 PMID: 24417811; PubMed Central PMCID: PMC4055747.

67. Winand R, Theys K, Eusebio M, Aerts J, Camacho RJ, Gomes P, et al. Assessing transmissibility of HIV-1 drug resistance mutations from treated and from drug-naive individuals. Aids. 2015; 29(15):2045–52. https://doi.org/10.1097/QAD.0000000000000811 PMID: 26355575; PubMed Central PMCID: PMC4570689.

68. Castro H, Pillay D, Cane P, Asboe D, Cambiano V, Phillips A, et al. Persistence of HIV-1 transmitted drug resistance mutations. The Journal of infectious diseases. 2013; 208(9):1459–63. Epub 2013/08/02. https://doi.org/10.1093/infdis/jit345 PMID: 23904291; PubMed Central PMCID: PMCPMC3789571.

69. WHO. HIV drug resistance report 2017. http://apps.who.int/iris/bitstream/10665/255896/1/9789241512831-eng.pdf?ua=1 2017.

70. Bruhn CA, Audelin AM, Helleberg M, Bjorn-Mortensen K, Obel N, Gerstoft J, et al. The origin and emergence of an HIV-1 epidemic: from introduction to endemicity. Aids. 2014; 28(7):1031–40. https://doi.org/10.1097/QAD.0000000000000198 PMID: 24451158.

71. Kharsany AB, Karim QA. HIV Infection and AIDS in Sub-Saharan Africa: Current Status, Challenges and Opportunities. The open AIDS journal. 2016; 10:34–48. https://doi.org/10.2174/1874613601610010034 PMID: 27347270; PubMed Central PMCID: PMC4893541.

72. Little SJ, Holte S, Routy JP, Daar ES, Markowitz M, Collier AC, et al. Antiretroviral-drug resistance among patients recently infected with HIV. N Engl J Med. 2002; 347(6):385–94. https://doi.org/10.1056/NEJMoa013552 PMID: 12167680.

73. Yerly S, Junier T, Gayet-Ageron A, Amari EB, von Wyl V, Gunthard HF, et al. The impact of transmission clusters on primary drug resistance in newly diagnosed HIV-1 infection. Aids. 2009; 23(11):1415–23. https://doi.org/10.1097/QAD.0b013e32832d40ad PMID: 19487906.

74. Zeh C, Inzaule SC, Ondoa P, Nafisa LG, Kasembeli A, Otieno F, et al. Molecular Epidemiology and Transmission Dynamics of Recent and Long-Term HIV-1 Infections in Rural Western Kenya. PloS one. 2016; 11(2):e0147436. https://doi.org/10.1371/journal.pone.0147436 PMID: 26871567; PubMed Central PMCID: PMC4752262.

75. Brenner BG, Roger M, Routy JP, Moisi D, Ntemgwa M, Matte C, et al. High rates of forward transmission events after acute/early HIV-1 infection. The Journal of infectious diseases. 2007; 195(7):951–9. https://doi.org/10.1086/512088 PMID: 17330784.

76. Hughes GJ, Fearnhill E, Dunn D, Lycett SJ, Rambaut A, Leigh Brown AJ, et al. Molecular phylodynamics of the heterosexual HIV epidemic in the United Kingdom. PLoS pathogens. 2009; 5(9):e1000590. https://doi.org/10.1371/journal.ppat.1000590 PMID: 19779560; PubMed Central PMCID: PMC2742734.

77. Nazziwa J, Njai HF, Ndembi N, Birungi J, Lyagoba F, Gershim A, et al. Short communication: HIV type 1 transmitted drug resistance and evidence of transmission clusters among recently infected antiretroviral-naive individuals from Ugandan fishing communities of Lake Victoria. AIDS research and human retroviruses. 2013; 29(5):788–95. Epub 2012/11/24. https://doi.org/10.1089/AID.2012.0123 PMID: 23173702; PubMed Central PMCID: PMCPMC3636596.

78. Vermund SH, Leigh-Brown AJ. The HIV Epidemic: High-Income Countries. Cold Spring Harbor perspectives in medicine. 2012; 2(5):a007195. https://doi.org/10.1101/cshperspect.a007195 PMID: 22553497; PubMed Central PMCID: PMC3331688.

79. Wallis CL, Godfrey C, Fitzgibbon JE, Mellors JW. Key Factors Influencing the Emergence of Human Immunodeficiency Virus Drug Resistance in Low- and Middle-Income Countries. The Journal of infectious diseases. 2017; 216(suppl_9):S851–S6. https://doi.org/10.1093/infdis/jix409 PMID: 29207000; PubMed Central PMCID: PMC5853971.

80. Kiguoya MW, Mann JK, Chopera D, Gounder K, Lee GQ, Hunt PW, et al. Subtype-Specific Differences in Gag-Protease-Driven Replication Capacity Are Consistent with Intersubtype Differences in HIV-1 Disease Progression. Journal of virology. 2017; 91(13). https://doi.org/10.1128/JVI.00253-17 PMID: 28424286; PubMed Central PMCID: PMC5469260.

81. Yang WL, Kouyos RD, Boni J, Yerly S, Klimkait T, Aubert V, et al. Persistence of transmitted HIV-1 drug resistance mutations associated with fitness costs and viral genetic backgrounds. PLoS pathogens. 2015; 11(3):e1004722. Epub 2015/03/24. https://doi.org/10.1371/journal.ppat.1004722 PMID: 25798934; PubMed Central PMCID: PMCPMC4370492.

82. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ. Episodic sexual transmission of HIV revealed by molecular phylodynamics. PLoS medicine. 2008; 5(3):e50. https://doi.org/10.1371/journal.pmed.0050050 PMID: 18351795; PubMed Central PMCID: PMC2267814.

Paper III

Open Forum Infectious Diseases

MAJOR ARTICLE

Drug Resistance in HIV-Positive Adults During the Initial Year of Antiretroviral Treatment at Ethiopian Health Centers

Anton Reepalu,[1] Dawit A. Arimide,[2,3] Taye T. Balcha,[1] Habtamu Yeba,[4] Adinew Zewdu,[4] Patrik Medstrand,[2] and Per Björkman[1]

[1]Clinical Infection Medicine, Department of Translational Medicine, Lund University, Malmö, Sweden, [2]Clinical Virology, Department of Translational Medicine, Lund University, Malmö, Sweden, [3]Ethiopian Public Health Institute, Addis Ababa, Ethiopia, and [4]Adama Public Health Research and Referral Laboratory Center, Adama, Ethiopia

Background. The increasing prevalence of antiretroviral drug resistance in Sub-Saharan Africa threatens the success of HIV programs. We have characterized patterns of drug resistance mutations (DRMs) during the initial year of antiretroviral treatment (ART) in HIV-positive adults receiving care at Ethiopian health centers and investigated the impact of tuberculosis on DRM acquisition.

Methods. Participants were identified from a cohort of ART-naïve individuals aged ≥18 years, all of whom had been investigated for active tuberculosis at inclusion. Individuals with viral load (VL) data at 6 and/or 12 months after ART initiation were selected for this study. Genotypic testing was performed on samples with VLs ≥500 copies/mL obtained on these occasions and on pre-ART samples from those with detectable DRMs during ART. Logistic regression analysis was used to investigate the association between DRM acquisition and tuberculosis.

Results. Among 621 included individuals (110 [17.5%] with concomitant tuberculosis), 101/621 (16.3%) had a VL ≥500 copies/mL at 6 and/or 12 months. DRMs were detected in 64/98 cases with successful genotyping (65.3%). DRMs were detected in 7/56 (12.5%) pre-ART samples from these individuals. High pre-ART VL and low mid-upper arm circumference were associated with increased risk of DRM acquisition, whereas no such association was found for concomitant tuberculosis.

Conclusions. Among adults receiving health center–based ART in Ethiopia, most patients without virological suppression during the first year of ART had detectable DRM. Acquisition of DRM during this period was the dominant cause of antiretroviral drug resistance in this setting. Tuberculosis did not increase the risk of DRM acquisition.

Keywords. drug resistance; Ethiopia; HIV; primary health care; tuberculosis.

INTRODUCTION

Antiretroviral treatment (ART) blocks viral replication, with improved survival and minimized risk of HIV transmission among people with HIV (PWH) [1]. In contrast, inadequate virological suppression is associated with worse prognosis and promotes selection of viruses carrying mutations conferring antiretroviral drug resistance [2, 3], which may also be transmitted onward [4]. Although the global rollout of ART has resulted in reduced AIDS incidence and HIV-related mortality, a successive increase in the prevalence of drug-resistance mutations (DRMs) in treatment-naïve PWH (termed pretreatment drug resistance [PDR]) has been observed in many world regions [5, 6], implying community transmission of drug-resistant viruses [7].

Several factors are involved in the emergence of HIV drug resistance, including irregular drug supply, suboptimal adherence, and drug–drug interactions [8]. Importantly, insufficient capacity for virological treatment monitoring leads to delayed recognition of patients with treatment failure [9], which in turn can result in further accumulation of DRMs [2, 10].

In low-income countries, most PWH receive nurse-based care, often decentralized to primary health centers [11]. In these settings, many individuals have advanced disease at ART initiation, with high viral loads (VLs) and low CD4 cell counts [12], factors that may compromise the chances of virologic suppression, with ensuing risk of acquisition of drug resistance [13, 14]. Furthermore, concurrent opportunistic infections are common in PWH starting ART in resource-limited settings. In this context, tuberculosis (TB) is of special importance. Individuals with TB co-infection at ART initiation have higher VLs than HIV mono-infected individuals [15], and could therefore be at increased risk of acquiring DRMs during ART.

Received 26 November 2020; editorial decision 1 March 2021; accepted 3 March 2021.

Correspondence: Anton Reepalu, MD, PhD, Ruth Lundskogs gata 3, fl. 6, 214 28 Malmö, SWEDEN (anton.reepalu@med.lu.se).

Open Forum Infectious Diseases®2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

DOI: 10.1093/ofid/ofab106

Downloaded from https://academic.oup.com/ofid/article/8/4/ofab106/6160155 by Tulund Library user on 31 May 2022

We have previously reported data on patterns of long-term virological outcomes in a cohort of 630 adults investigated for active TB before starting ART at Ethiopian health centers. Whereas 68% achieved and maintained virological suppression <150 copies/mL for up to 4 years after treatment initiation, 21% had a VL result ≥1000 copies/mL on at least 1 occasion during follow-up. Lack of persistent virological suppression was associated with male sex, pretreatment CD4 count <100 cells/mm^3, and malnutrition, but not with active TB [16].

In this study, we have characterized antiretroviral drug resistance mutations among participants with inadequate viral suppression during the first year after ART initiation in this cohort and assessed the relative contribution of acquired and pretreatment drug resistance. In addition, we have determined factors associated with acquisition of DRMs, with particular regard to concomitant TB.

METHODS

Participants in the study cohort were recruited and followed at public health centers in an uptake area in and around the city Adama, Ethiopia, from 2011 to 2015. Consenting HIV-positive ART-naïve adults (≥18 years) who were eligible to start ART according to Ethiopian National ART guidelines at the time of the study (CD4 count <350 cells/mm^3 and/or World Health Organization [WHO] stage 4 disease) were included [17]. The study cohort has been described in detail previously [16, 18].

At inclusion, all participants were investigated for active TB, irrespective of symptoms. Sputum samples (and fine-needle aspirates of enlarged lymph nodes, if present) were analyzed with smear microscopy, GeneXpert MTB/Rif, and liquid culture. Sociodemographic and medical information was collected with structured questionnaires. Blood samples were obtained for CD4 count testing, with storage of plasma at −80°C. Medical information was updated, along with repeated blood sampling on subsequent follow-up visits scheduled at months 1, 3, 6, and 12, and biannually thereafter for up to 4 years after ART initiation. If incident TB was suspected at any time during follow-up, bacteriological TB investigations were repeated. Non-nucleoside reverse transcriptase inhibitor (NNRTI)–based ART was initiated according to Ethiopian national guidelines by nonphysician clinicians at the study sites. Participants diagnosed with TB received TB treatment according to Ethiopian national guidelines, provided at the same facilities [17].

VL was performed on stored plasma in batches during the study period using the Abbott Real-Time HIV-1 assay (Abbott Molecular Inc., Des Plaines, IL, USA; detection limit 40 copies/mL) or the Abbott m2000 RealTime System Automated molecular platforms (Abbott Molecular Inc., Des Plaines, IL, USA; detection limit 150 copies/mL). VL results were communicated to the responsible clinicians at the respective health center.

For this study, all individuals with VL data available at 6 and/or 12 months after ART initiation were included. This 6-month time span was used in order to include all patients with lack of virological suppression (defined as ≥1 VL result ≥500 copies/mL) during the first 12 months of ART.

HIV Genotype and Drug Resistance Mutation Analysis

Genotypic testing was performed on stored samples with a VL ≥500 copies/mL obtained at 6 and/or 12 months after starting ART. A 1084-bp fragment of HIV-1 pol (corresponding to the position 2243–3326 of HXB2, Genbank Accession Number K03455) comprising amino acids 6–99 of the protease (PR) and 1–251 of the reverse transcriptase (RT) was amplified using an in-house genotyping assay [19, 20]. Polymerase chain reaction products were directly sequenced using the Sanger method with 6 primers (3 on each strand) on an ABI 3100 or an ABI 3500xl DNA Genetic Analyzer (Applied Biosystems). Sequence assembly and editing were performed using the RECall, version 2.0, HIV-1 sequencing analysis tool [21]. Sequence quality control was performed to rule out contamination and mislabeling of samples using the online Quality Control program of the Los Alamos HIV sequence database (hiv.lanl.gov). Individuals with contaminated samples were excluded from this study. The presence of DRM was determined using the Stanford HIVdb database algorithm 8.6 (hivdb.stanford.edu) [22].

To determine whether detected DRMs had evolved during ART (acquired drug resistance [ADR]) or were present before ART initiation (pretreatment drug resistance [PDR]), genotypic analysis was also performed on samples obtained before starting ART for such participants. In order to estimate the prevalence of pre-ART DRMs among participants who had died or were lost to follow-up before scheduled sampling at 6 and 12 months (and could hence not be classified with regard to DRM after starting ART), we also genotyped pretreatment samples from these individuals. PDR mutations were examined according to the Stanford Genotypic Resistance calibrated population resistance (CPR) tool, version 6.0, based on the WHO surveillance transmitted drug resistance mutation list of 2009 [23, 24].

Statistical Analysis

Comparison of characteristics of cohort participants who were included and excluded from this study was performed using the Mann-Whitney *U* test for continuous variables and the chi-square test for categorical variables.

We used logistic regression analysis to investigate the association between TB and DRM acquisition. For this analysis, individuals with a VL <500 copies/mL in all available samples at 6 and/or 12 months were compared with those with ADR. Individuals with a VL ≥500 copies/mL without detectable ADR, as well as those without genotypic data, were excluded from this analysis. As we specifically aimed to investigate the risk of DRM acquisition during ART, those with

Downloaded from https://academic.oup.com/ofid/article/8/4/ofab106/6160155 by Tulund Library user on 31 May 2022

DRMs detected before ART initiation were also excluded from this analysis. In addition to TB (defined as bacteriologically or clinically diagnosed TB) at ART initiation, age and gender were included in the regression analysis, as well as pretreatment CD4 count and VL, and mid-upper arm circumference (MUAC) was included as a marker of malnutrition. Age was divided into 5-year intervals and CD4 counts into intervals of 25 cells/mm^3 for interpretation of odds ratios. All variables included in the univariate analysis were also included in the multivariate analysis.

Statistical analyses were performed using SPSS, version 26 (IBM Corp, Armonk, NY, USA). *P* values <.05 were considered statistically significant.

Patient Consent Statement

Ethical approval was obtained from the national Research Ethics Review Committee at the Ministry of Technology and Innovation of Ethiopia and the Regional Ethical Review Board of Lund University, Sweden. All study participants provided written informed consent.

RESULTS

Participant Characteristics

A total of 729/812 (89.8%) individuals enrolled in the original cohort started ART. Among these, 621 (85.2%) had VL data at 6 and/or 12 months after treatment initiation and were included in this study (Figure 1).

Among the 108 excluded individuals, 84 (77.8%) were lost to follow-up, died, or transferred out before the 6- and/or 12-month visit, and 24 (21.3%) did not have follow-up VL results (Figure 1). Among the 621 included individuals, 377 (60.7%) were women, the median CD4 count at ART initiation (interquartile range [IQR]) was 191 (121–274) cell/mm^3, and 110 (17.7%) had concomitant TB. Efavirenz (EFV) was the most common NNRTI used (83.9%). All participants received lamivudine (3TC), with the third nucleoside reverse transcriptase inhibitor (NRTI) being tenofovir disoproxil fumarate (TDF) in 89.0%, zidovudine (AZT) in 10.0%, and stavudine (d4T) in 1.0% (Table 1). Patients who were excluded were more likely to be male and had lower CD4 counts and MUAC; furthermore, the proportion of concomitant TB was higher among excluded patients (25% vs 18%; *P* = .07) (Table 1).

Among those with available VL data, 60/534 (10.1%) and 72/520 (13.8%) had VL ≥500 copies/mL at 6 and 12 months, respectively. The median logVL in patients with VL ≥500 copies/mL (IQR) was 4.54 (3.73–5.33) and 4.58 (4.02–5.15) at 6 and 12 months, respectively. Among participants with VL ≥500 copies/mL at these time points, 55/60 (91.7%) and 66/72 (91.7%) had ≥1000 copies/mL at 6 and 12 months, respectively.

Drug Resistance During the First Year After Starting ART

In total, 98 individuals with ≥1 VL ≥500 copies/mL at 6 and/or 12 months had samples available for genotyping (both 6 and 12 months: 29; only 6 months: 29; only 12 months: 40) (Figure 1). All of the specimens were successfully amplified

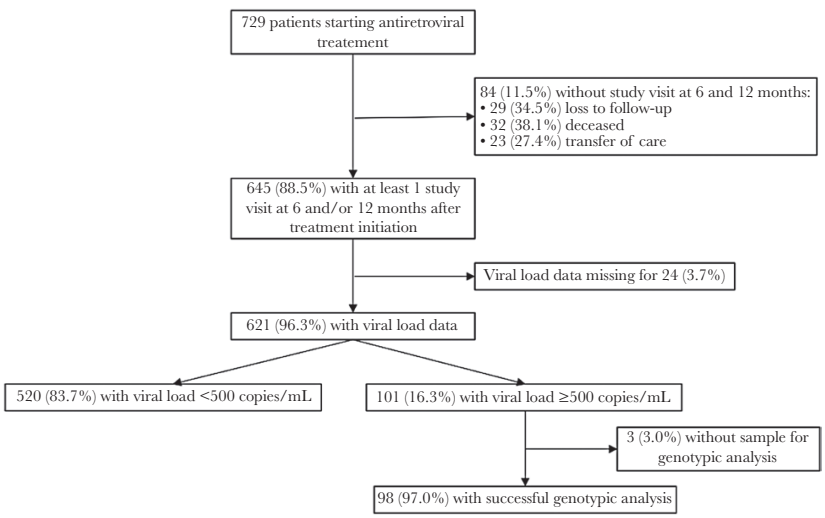


Figure 1. Flowchart of study participants eligible for genotypic analysis. Whereas 493 (79.4%) of the 621 included individuals had viral load (VL) data at both 6 and 12 months, VL data were missing from 27 (4.3%) and 101 (16.3%) participants at 6 and 12 months, respectively. For 7/27 participants with missing VL data at 6 months, this represented a missed study visit; the respective proportion at 12 months was 33/74; 27 individuals had not reached the 12-month visit at study closure. For the remaining cases, study visits were registered but blood samples for VL testing were not available.

Downloaded from https://academic.oup.com/ofid/article/8/4/ofab106/6160155 by Tulund Library user on 31 May 2022

Table 1. Characteristics of Cohort Participants at Antiretroviral Treatment Initiation With Comparison of Individuals Included and Excluded in the Current Study

		Included (n = 621)	Excluded (n = 108)	*P* Value
Age	years	32 (28–40)	31 (28–39)	.287
Female sex		377 (61)	54 (50)	.037
Viral load	log copies/mL	5.11 (4.50–5.55)	5.15 (4.47–5.67)	.595
CD4 count	cells/mm^3	191 (121–274)	154 (101–274)	.042
CD4 strata	<100 cells/mm^3	112 (18)	26 (24)	
	100–200 cells/mm^3	220 (36)	41 (38)	
	>200 cells/mm^3	288 (47)	40 (37)	
MUAC	cm	23.0 (21.0–25.0)	22.0 (20.0–24.0)	<.01
TB co-infection		110 (18)	27 (25)	.074
NNRTI	NVP	100 (16)	17 (14)	.611
	EFV	521 (84)	91 (86)	.611
NRTI	3TC	621 (100)	108 (100)	
	TDF	553 (89)	90 (85)	.217
	AZT	62 (10)	10 (9)	.861
	d4T	6 (1)	6 (6)	<.01

P values were derived using the Mann-Whitney *U* test for continuous variables and the chi-square test for categorical variables. Data presented as No. (%) or median (interquartile range). Viral load data were available for 703/729 (96.4%), and CD4 counts were available for 727/729 (99.7%).

Abbreviations: 3TC, lamivudine; AZT, zidovudine; d4T, stavudine; EFV, efavirenz; MUAC, mid-upper arm circumference; NNRTI, non-nucleoside reverse transcriptase inhibitor; NRTI, nucleoside reverse transcriptase inhibitor; NVP, nevirapine; TB, tuberculosis; TDF, tenofovir disoproxil fumarate.

and genotyped. All genotyped viruses belonged to subtype C, and DRMs were detected in 64 (65.3%) of these 98 individuals (Table 2; Supplementary Data). NNRTI-associated mutations were present in all individuals with DRMs. Additionally, 35/98 (35.7%) had NRTI-associated DRMs (Table 2). No protease inhibitor (PI)–associated mutations were detected.

The median logVL at the time of VL ≥500 copies/mL (IQR) was 4.60 (3.98–5.11) for patients with DRMs, compared with 4.45 (3.55–5.19) for those without detectable DRMs.

Table 2. Frequency of the 4 Most Common NNRTI and NRTI Drug Resistance Mutations Detected in Individuals With Viral Loads ≥500 Copies/mL at 6 and/or 12 Months After Treatment Initiation

	Total (n = 98)	6 mo (n = 58)	12 mo (n = 69)
Any NNRTI and/or NRTI	64 (65.3)	41 (70.7)	46 (66.7)
NNRTI	64 (65.3)	41 (70.7)	46 (66.7)
K103N	39 (39.8)	23 (39.7)	29 (42.0)
V106A/M	16 (16.3)	9 (15.5)	11 (15.9)
Y181C/I	15 (15.3)	12 (20.7)	12 (17.4)
G190A/C/E/Q/S	12 (12.2)	12 (20.7)	4 (5.8)
NRTI	35 (35.7)	26 (44.8)	25 (36.2)
M184V/I	30 (30.6)	20 (34.5)	23 (33.3)
K65R	24 (24.5)	20 (34.5)	17 (24.6)
A62V	9 (9.2)	7 (12.1)	7 (10.1)
Y115F	7 (7.1)	5 (8.6)	7 (10.1)

Data are presented as No. (%).

Abbreviations: NNRTI, non-nucleoside reverse transcriptase inhibitor; NRTI, nucleoside reverse transcriptase inhibitor.

Pretreatment Drug Resistance

Samples obtained before ART initiation were available for 56/64 (87.5%) of those with DRMs during ART. DRMs were detected in 7/56 (12.5%) of these samples. All 7 had NNRTI resistance (K103N = 5, K181V = 1, K103N and G190S combined = 1). The sample with dual NNRTI resistance also had multiple thymidine analogue mutations (TAMs): D67N, T215C, and K219E (Table 3). Only minor changes were observed in patterns of DRMs comparing pre-ART and ART samples from the same individuals (data not shown).

Among the 61 subjects who were LTFU or died before reaching a 6- or 12-month visit, pretreatment samples were available for 49 (80.3%), with successful genotyping in 43/49 (87.8%). Pretreatment DRM was detected in 2 of these (4.7%; both NNRTI-associated) (Table 3).

Factors Associated With Drug Resistance Acquisition During Antiretroviral Treatment

In this analysis, 57 cases with ADR were compared with 520 individuals with VL <500 copies/mL at 6 and/or 12 months. Concomitant TB was not significantly associated with ADR. In univariate analysis, CD4 count, VL, and MUAC were associated with ADR (Table 4). In multivariate analysis, the statistically significant association remained for pretreatment VL and MUAC (Table 4).

DISCUSSION

In this cohort of PWH receiving care at Ethiopian health centers, we detected DRMs in a majority of patients with VLs ≥500 copies/mL at 6–12 months after ART initiation. In most of these treatment-naïve individuals (87.5%), DRMs were not detected in samples obtained before starting ART, implying acquired drug resistance as the major mechanism for drug resistance in this setting.

Table 3. Frequency of Drug Resistance Mutations Detected in Pretreatment Samples

	Total (n = 125)	DRM at 6 and/or 12 mo (n = 64)	Deceased or LTFU Before Providing 6- or 12-mo Samples (n = 61)
Genotype missing	26 (20.8)	8 (12.5)	18 (29.5)
Genotype available	99 (79.2)	56 (87.5)	43 (70.5)
Any sDRM detected	9 (9.1)	7 (12.5)	2 (4.7)
NNRTI sDRM	9 (9.1)	7 (10.9)	2 (4.7)
NRTI sDRM	1 (1.0)	1 (1.8)	0 (0)

Data are presented as No. (%).

Abbreviations: DRM, drug resistance mutation; LTFU, lost to follow-up; NNRTI, non-nucleoside reverse transcriptase inhibitor; NRTI, nucleoside reverse transcriptase inhibitor; sDRM, surveillance drug resistance mutation included in the World Health Organization 2009 sDRM list.

Downloaded from https://academic.oup.com/ofid/article/8/4/ofab106/6160155 by Tulund Library user on 31 May 2022

Table 4. Factors Associated With Drug Resistance Acquisition During Antiretroviral Treatment

	OR (95% CI)	*P*	aOR (95% CI)	*P*
Tuberculosis	1.09 (0.53–2.24)	.817	0.76 (0.35–1.68)	.503
Age, per 5 y	1.09 (0.96–1.25)	.198	1.00 (0.85–1.17)	.978
Male sex	1.73 (1.00–3.00)	.050	1.50 (0.81–2.78)	.197
CD4 count, per 25 cells/mm^3	0.88 (0.82–0.95)	.001	0.93 (0.87–1.01)	.069
Viral load, log copies/mL	2.57 (1.64–4.03)	<.001	1.96 (1.21–3.16)	.006
MUAC, per cm	0.85 (0.77–0.94)	.002	0.89 (0.80–0.99)	.031

Abbreviations: aOR, adjusted odds ratio; MUAC, mid-upper arm circumference; OR, odds ratio.

Incomplete virological suppression during ART promotes selection of HIV variants carrying DRMs [2]. In particular, NNRTI-associated DRMs have been reported to emerge early in individuals who fail to achieve suppression after starting ART, while NRTI-associated DRMs tend to accumulate after longer periods of persistent replication [10]. In agreement with our findings, 87% of participants in a study conducted in 6 Sub-Saharan African countries had DRM at first occasion of VL ≥1000 copies/mL after at least 6 months of ART (a median of 1 year after first-line ART initiation) [2]. Previous data from Ethiopia also imply high rates of DRMs in patients with virological failure. Two repeated surveys performed in a hospital clinic in Northwestern Ethiopia in 2011 and 2015, respectively, showed an increase in the proportion of patients with VLs >400 copies/mL with detectable DRMs (40% vs 66%) [25, 26]. In another study, based on data from 7 Ethiopian teaching hospitals, DRMs were detected in 76.6% and 66.7% of patients with VLs >1000 copies/mL after 6 and 12 months of ART, respectively [27].

In contrast to most other studies from Sub-Saharan Africa, our cohort was recruited and followed at health centers, where the majority of PWH receive ART. To our knowledge, only 1 study in Ethiopia has previously investigated antiretroviral drug resistance at the health center level; among 11 patients with VLs >1000 copies/mL, 9 (81.8%) had DRMs [28]. However, in contrast to our findings, 6/9 (66.7%) of these individuals had detectable DRMs in samples obtained before starting ART. Instead, our results suggest acquisition of DRMs during the first 6–12 months of ART to be the dominant mechanism of drug resistance in this health care setting. In turn, this emphasizes the importance of adherence and implies that adherence support needs to be strengthened in Ethiopian ART programs in order to secure effective treatment options.

We could not determine the exact time point after ART initiation that DRM mutations emerged, but as the prevalence of NNRTI mutations was similar at 6 and 12 months, it is likely that these mutations occur during the first months of ART. VL testing and early identification of those with failing treatment during the initial 6 months of ART could therefore be effective for saving first-line options. In line with this, Kerschberger et al. showed superior ART outcomes when VL was first measured at 3 months compared with 6 months, potentially shortening the time on failing treatment [29]. Although virologic suppression can occur despite the presence of NNRTI-associated DRMs [30, 31], the recommendation of enhanced adherence counseling followed by repeat VL testing is unlikely to be successful in most patients with incomplete viral suppression due to drug resistance, constituting two-thirds in our population. In these cases, change to second-line ART regimens is indicated, whereas persons without DRM will not benefit from treatment modification. This dichotomy illustrates the need for access to methods to determine the presence of major drug resistance in patients with virologic failure in order to provide effective interventions to optimize treatment outcomes.

As expected, and in agreement with other studies, mutations conferring NNRTI resistance were the most commonly observed type of DRM. Importantly, the VL for those with ADR was high (4.60 log10 copies/mL), indicating the potential of onward transmission of viruses harboring DRM. Several studies, performed in different parts of Sub-Saharan Africa (as well as other low- and middle-income settings), show increasing rates of pre-ART resistance, paralleling scale-up of ART programs [6]. In particular, rates of NNRTI mutations are high, with levels >10% in some areas [6]. This situation has prompted recommendations to replace NNRTI with the integrase strand transfer inhibitor dolutegravir in first-line regimens [32]. Although the genetic barrier to resistance of dolutegravir is higher than for NNRTIs, dolutegravir monotherapy promotes selection of resistant variants [33]. Functionally, this situation could arise if NNRTIs are replaced with dolutegravir in patients with combined NRTI mutations. In our cohort (in which nearly 90% had TDF as the NRTI backbone), combined NRTI resistance with K65R and M184V/I was present in 25.5% and 21.7% with VLs ≥500 copies/mL at 6 and 12 months, respectively. In such patients, a regimen switch from NNRTI to dolutegravir could lead to functional dolutegravir monotherapy, with a risk of emergence of dolutegravir resistance [34].

Dolutegravir is also recommended as a second-line alternative for patients failing NNRTI-based ART [32]. The pattern of NRTI DRMs found in this study supports this recommendation if TDF is replaced by AZT, as mutations conferring AZT resistance were rare in our population.

The proportion of PDR among individuals starting ART in Ethiopia is not well known. In a study conducted at 7 Ethiopian hospitals from 2009 to 2011, PDR was detected in 18/461 (3.9%) randomly selected ART-naïve individuals [35]. We did not aim to assess PDR in this study. Nonetheless, in order to differentiate between PDR and ADR in our participants, we genotyped samples obtained before ART initiation for those with DRMs detected at 6 or 12 months of ART. Among these, PDR was detected in 7/56 (12.5%).

Downloaded from https://academic.oup.com/ofid/article/8/4/ofab106/6160155 by Tulund Library user on 31 May 2022

In this cohort, concomitant TB was not associated with increased risk of acquired drug resistance in patients receiving NNRTI-based ART. This is in line with previously reported findings from this cohort of similar short- and long-term ART outcomes with regard to TB co-infection [16, 36]. Factors that have been associated with acquisition of DRMs in other studies include male sex, higher pretreatment VL, and lower CD4 counts [37–39]. Interestingly, although participants with TB were more likely to have these characteristics [18], they were not at increased risk of DRM acquisition. This could suggest an indirect protective effect of concomitant TB related to closer contact with health care.

The only variables independently associated with ADR in this cohort were high pretreatment VL and low MUAC. Both of these factors indicate more advanced HIV disease. We have previously shown that low MUAC is associated with concomitant TB in ART-naïve PWH [40], as well as unfavorable ART outcomes [16, 36]. Low MUAC could also reflect unrecognized opportunistic infections, as well as poverty and food insecurity [41, 42].

This study was based on a well-characterized cohort in which all participants had been subjected to intensified TB case-finding. These patients received nurse-based care at health centers, which we consider to be a representative setting for Ethiopia, as well as for other countries in Sub-Saharan Africa. Nonetheless, this study has several limitations. Genotyping was performed with Sanger sequencing, which has a lower sensitivity compared with next-generation sequencing [43]. It is therefore possible that DRMs occurring at low frequencies were missed and that some of the DRMs detected during ART (and hence categorized as ADR) could also have been detected at inclusion if a sequencing technology with higher resolution were used. Furthermore, pretreatment genotypic data were missing for some of these individuals. Although emergence of DRMs is most common in the setting of high viral replication, this can occur also during low-level viremia [44]. For this reason, we chose 500 copies/mL to select cases for genotypic testing, in contrast to most prior studies on antiretroviral drug resistance in Sub-Saharan Africa (which have used a threshold of 1000 copies/mL [2, 27]). Therefore, direct comparisons with our findings require consideration of this circumstance. However, 83.3% of nonsuppressed individuals had VL >1000 copies/mL. Finally, this study was not specifically powered to test the hypothesis that concomitant TB increases the risk of ADR. However, the 95% confidence intervals of both the unadjusted and adjusted odds ratio do indicate that a clinically relevant association was not missed. The prevalence of concomitant TB tended to be higher among the 108 individuals excluded due to lack of follow-up viral load and genotypic data, which could imply selection bias, which may have had an impact on these results.

In conclusion, antiretroviral drug resistance was observed in a majority of individuals not achieving virological suppression after 6–12 months of ART. In most of these, DRMs were not detected in samples obtained before starting ART, implying DRM acquisition during the initial year of ART as the dominant cause of drug resistance in this population. This demonstrates the importance of earlier identification of patients without virological suppression, so that interventions can be implemented before drug resistance acquisition has occurred.

Supplementary Data

Supplementary materials are available at Open Forum Infectious Diseases online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Acknowledgments

We acknowledge all patients who participated in this study as well as the staff at participating health centers and the Adama Public Health Research and Referral Laboratory Center for their excellent work.

Financial support. This work was supported by the Swedish Research Council (P.M.; 2015–05842), the Swedish Civil Contingencies Agency (P.B.; 2009–7551), the Crafoord Foundation (P.B.; 20130644), the Swedish Physicians Against AIDS Research Foundation (A.R.; 2015-0002), and a private donation to the Faculty of Medicine, Lund University (P.B. and P.M.).

Potential conflicts of interest. A.R., D.A.A., T.T.B., H.Y., A.Z., P.M., and P.B. report no conflicts of interest.

Author contributions. A.R., P.B., D.A.A., T.T.B., and P.M. designed the study; A.R., T.T.B., and P.B. collected clinical data; D.A.A., P.M., H.Y., and A.Z. collected biological data and performed experiments; A.R., D.A.A., P.B., and P.M. analyzed results; A.R., P.B., D.A.A., and P.M. wrote the article. All authors reviewed and accepted the final version of the article.

Prior presentation. Preliminary results from this study were presented at the 2019 Conference on Retroviruses and Opportunistic Infections, Seattle, WA, USA (poster 0536).

References

1. World Health Organization (WHO). Guidelines for Managing Advanced HIV Disease and Rapid Initiation of Antiretroviral Therapy. Geneva: World Health Organisation; **2017**.
2. Boender TS, Kityo CM, Boerma RS, et al. Accumulation of HIV-1 drug resistance after continued virological failure on first-line ART in adults and children in Sub-Saharan Africa. J Antimicrob Chemother **2016**; 71:2918–27.
3. Clutter DS, Jordan MR, Bertagnolio S, Shafer RW. HIV-1 drug resistance and resistance testing. Infect Genet Evol **2016**; 46:292–307.
4. Stekler JD, Milne R, Payant R, et al. Transmission of HIV-1 drug resistance mutations within partner-pairs: a cross-sectional study of a primary HIV infection cohort. PLoS Med **2018**; 15:e1002537.
5. Gupta RK, Gregson J, Parkin N, et al. HIV-1 drug resistance before initiation or re-initiation of first-line antiretroviral therapy in low-income and middle-income countries: a systematic review and meta-regression analysis. Lancet Infect Dis **2018**; 18:346–55.
6. World Health Organization (WHO). HIV Drug Resistance Report. Geneva: World Health Organization; **2019**.
7. Hofstra LM, Sauvageot N, Albert J, et al; SPREAD Program. Transmission of HIV drug resistance and the predicted effect on current first-line regimens in Europe. Clin Infect Dis **2016**; 62:655–63.
8. Bertagnolio S, De Luca A, Vitoria M, et al. Determinants of HIV drug resistance and public health implications in low- and middle-income countries. Antivir Ther **2012**; 17:941–53.
9. De Luca A, Marazzi MC, Mancinelli S, et al. Prognostic value of virological and immunological responses after 6 months of antiretroviral treatment in adults with HIV-1 infection in Sub-Saharan Africa. J Acquir Immune Defic Syndr **2012**; 59:236–44.

Downloaded from https://academic.oup.com/ofid/article/8/4/ofab106/6160155 by Tulund Library user on 31 May 2022

10. De Luca A, Sidumo ZJ, Zanelli G, et al. Accumulation of HIV-1 drug resistance in patients on a standard thymidine analogue-based first line antiretroviral therapy after virological failure: implications for the activity of next-line regimens from a longitudinal study in Mozambique. BMC Infect Dis **2017**; 17:605..
11. Joint United Nations Programme on HIV/AIDS (UNAIDS). Ending AIDS: Progress Towards the 90–90–90 Targets. Geneva: UN Joint Programme on HIV/AIDS (UNAIDS); **2017**..
12. Ford N, Mills EJ, Egger M. Immunodeficiency at start of antiretroviral therapy: the persistent problem of late presentation to care. Clin Infect Dis **2014**; 60:1128–30..
13. The PLATO II Project Team and COHERE group. Triple-class virologic failure in hiv-infected patients undergoing antiretroviral therapy for up to 10 years. Arch Intern Med **2010**; 170:410–9..
14. Raffi F, Hanf M, Ferry T, et al; Dat'AIDS Study Group. Impact of baseline plasma HIV-1 RNA and time to virological suppression on virological rebound according to first-line antiretroviral regimen. J Antimicrob Chemother **2017**; 72:3425–34..
15. Olsson O, Björkman P, Jansson M, et al. Plasma profiles of inflammatory markers associated with active tuberculosis in antiretroviral therapy-naive human immunodeficiency virus-positive individuals. Open Forum Infect Dis **2019**; 6:XXX–XX..
16. Reepalu A, Balcha TT, Sturegård E, Medstrand P, Björkman P. Long-term outcome of antiretroviral treatment in patients with and without concomitant tuberculosis receiving health center–based care—results from a prospective cohort study. Open Forum Infect Dis **2017**; 4:1–8..
17. Ethiopian Federal Ministry of Health. Guidelines for Clinical and Programmatic Management of TB, Leprosy and TB/HIV in Ethiopia. Addis Ababa: Ethiopian Federal Ministry of Health; **2012**. Available at https://etharc.org/index.php/resources/download/finish/33/709. Accessed 16 February 2021..
18. Balcha TT, Sturegård E, Winqvist N, et al. Intensified tuberculosis case-finding in HIV-positive adults managed at Ethiopian health centers: diagnostic yield of Xpert MTB/RIF compared with smear microscopy and liquid culture. PLoS One **2014**; 9:e85478..
19. Zhou Z, Wagar N, DeVos JR, et al. Optimization of a low cost and broadly sensitive genotyping assay for HIV-1 drug resistance surveillance and monitoring in resource-limited settings. PLoS One **2011**; 6:e28184..
20. Arimide DA, Abebe A, Kebede Y, et al. HIV-genetic diversity and drug resistance transmission clusters in Gondar, Northern Ethiopia, 2003–2013. PLoS One **2018**; 13:2003–13..
21. Woods CK, Brumme CJ, Liu TF, et al. Automating HIV drug resistance genotyping with RECall, a freely accessible sequence analysis tool. J Clin Microbiol **2012**; 50:1936–42..
22. Rhee SY, Kantor R, Katzenstein DA, et al; International Non Subtype B HIV-1 Working Group. HIV-1 pol mutation frequency by subtype and treatment experience: extension of the HIVseq program to seven non-B subtypes. AIDS **2006**; 20:643–51..
23. Bennett DE, Camacho RJ, Otelea D, et al. Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. PLoS One **2009**; 4:e4724..
24. Gifford RJ, Liu TF, Rhee SY, et al. The calibrated population resistance tool: standardized genotypic estimation of transmitted HIV-1 drug resistance. Bioinformatics **2009**; 25:1197–8..
25. Mulu A, Maier M, Liebert UG. Low incidence of HIV-1C acquired drug resistance 10 years after roll-out of antiretroviral therapy in Ethiopia: a prospective cohort study. PLoS One **2015**; 10:e0141318..
26. Mulu A, Maier M, Liebert UG. Upward trends of acquired drug resistances in Ethiopian HIV-1C isolates: a decade longitudinal study. PLoS One **2017**; 12:e0186619..
27. Telele NF, Kalu AW, Gebre-Selassie S, et al. A viral genome wide association study and genotypic resistance testing in patients failing first line antiretroviral therapy in the first large countrywide Ethiopian HIV cohort. BMC Infect Dis **2019**; 19:569..
28. Abdissa A, Yilma D, Fonager J, et al. Drug resistance in HIV patients with virological failure or slow virological response to antiretroviral therapy in Ethiopia. BMC Infect Dis **2014**; 14: 181..
29. Kerschberger B, Boulle AM, Kranzer K, et al. Superior virologic and treatment outcomes when viral load is measured at 3 months compared to 6 months on antiretroviral therapy. J Int AIDS Soc **2015**; 18:20092..
30. Hoffmann CJ, Charalambous S, Sim J, et al. Viremia, resuppression, and time to resistance in human immunodeficiency virus (HIV) subtype C during first-line antiretroviral therapy in South Africa. Clin Infect Dis **2009**; 49:1928–35..
31. Derache A, Iwuji CC, Baisley K, et al. Impact of next-generation sequencing defined human immunodeficiency virus pretreatment drug resistance on virological outcomes in the ANRS 12249 treatment-as-prevention trial. Clin Infect Dis **2019**; 69:207–14..
32. World Health Organization (WHO). Update of First and Second Line Antiretroviral Regimens. Geneva: World Health Organization; **2019**..
33. Wijting I, Rokx C, Boucher C, et al. Dolutegravir as maintenance monotherapy for HIV (DOMONO): a phase 2, randomised non-inferiority trial. Lancet HIV **2017**; 4:e547–54..
34. Rhee SY, Grant PM, Tzou PL, et al. A systematic review of the genetic mechanisms of dolutegravir resistance. J Antimicrob Chemother **2019**; 74:3135–49..
35. Telele NF, Kalu AW, Gebre-Selassie S, et al. Pretreatment drug resistance in a large countrywide Ethiopian HIV-1C cohort: a comparison of Sanger and high-throughput sequencing. Sci Rep **2018**; 8:7556..
36. Reepalu A, Balcha TT, Skogmar S, et al. High rates of virological suppression in a cohort of human immunodeficiency virus-positive adults receiving antiretroviral therapy in Ethiopian health centers irrespective of concomitant tuberculosis. Open Forum Infect Dis **2014**; 1:XXX–XX..
37. Palumbo PJ, Fogel JM, Hudelson SE, et al. HIV drug resistance in adults receiving early vs delayed antiretroviral therapy. J Acquir Immune Defic Syndr **2018**; 77:484–91..
38. Fogel JM, Hudelson SE, Ou S-S, et al. HIV drug resistance in adults failing early antiretroviral treatment: results from the HIV Prevention Trials Network 052 trial. J Acquir Immune Defic Syndr **2016**; 72:304–9..
39. Xing H, Wang X, Liao L, et al. Incidence and associated factors of HIV drug resistance in Chinese HIV-infected patients receiving antiretroviral treatment. PLoS One **2013**; 8:4–9..
40. Skogmar S, Balcha TT, Jemal ZH, et al. Development of a clinical scoring system for assessment of immunosuppression in patients with tuberculosis and HIV infection without access to CD4 cell testing—results from a cross-sectional study in Ethiopia. Glob Health Action **2014**; 7:23105..
41. Koethe JR, Heimburger DC. Nutritional aspects of HIV-associated wasting in Sub-Saharan Africa. Am J Clin Nutr **2010**; 91:1138–42S..
42. Weiser SD, Tuller DM, Frongillo EA, et al. Food insecurity as a barrier to sustained antiretroviral therapy adherence in Uganda. PLoS One **2010**; 5:e10340..
43. Inzaule SC, Hamers RL, Noguera-Julian M, et al; PanAfrican Studies to Evaluate Resistance. Clinically relevant thresholds for ultrasensitive HIV drug resistance testing: a multi-country nested case-control study. Lancet HIV **2018**; 5:e638–46..
44. Delaugerre C, Gallien S, Flandre P, et al. Impact of low-level-viremia on HIV-1 drug-resistance evolution among antiretroviral treated-patients. PLoS One **2012**; 7:e36673..

Downloaded from https://academic.oup.com/ofid/article/8/4/ofab106/6160155 by Tulund Library user on 31 May 2022

Paper IV

High Level of HIV Drug Resistance and Virologic Nonsuppression Among Female Sex Workers in Ethiopia: A Nationwide Cross-Sectional Study

Dawit Assefa Arimide,[a,b] Minilik Demissie Amogne, PhD,[b,c] Yenew Kebede, MD,[d] Taye T. Balcha, MD, PhD,[a] Fekadu Adugna, MD,[e] Artur Ramos, PhD,[f] Joshua DeVos, MSc,[f] Clement Zeh, PhD,[f] Anette Agardh, PhD,[c] Joy Chih-Wei Chang, PhD,[f] Per Björkman, PhD,[a] and Patrik Medstrand, PhD[a]

Objective: To determine viral load (VL) nonsuppression (VLN) rates, HIV drug resistance (HIVDR) prevalence, and associated factors among female sex workers (FSWs) in Ethiopia.

Methods: A cross-sectional biobehavioral survey was conducted among FSWs in 11 cities in Ethiopia in 2014. Whole blood was collected, and HIVDR genotyping was performed. Logistic regression analysis was performed to identify factors associated with VLN and HIVDR.

Results: Among 4900 participants, 1172 (23.9%) were HIV-positive and 1154 (98.5%) had a VL result. Participants were categorized into antiretroviral therapy (ART) (n = 239) and ART-naive (n = 915) groups based on self-report. From the 521 specimens (ART, 59; ART-naive, 462) with VL ≥1000 copies/mL, genotyping was successful for 420 (80.6%) and 92 (21.9%) had drug resistance mutations (DRMs). Pretreatment drug resistance (PDR) was detected in 16.5% (63/381) of the ART-naive participants. Nucleoside reverse transcriptase inhibitor (NRTI), non-NRTIs (NNRTIs), and dual-class DRMs were detected in 40 (10.5%), 55 (14.4%), and 35 (9.2%) of the participants, respectively. Among 239 participants on ART, 59 (24.7%) had VLN. Genotyping was successfully performed for 39 (66.1%). DRMs were detected in 29 (74.4%). All 29 had NNRTI, 23 (79.3%) had NRTI or dual-class DRMs. VLN was associated with age 35 years or older, $CD4^+$ T-cell count <350 cells/mm^3, and being forced into selling sex. PDR and acquired drug resistance were associated with $CD4^+$ T-cell count <350 cells/mm^3 ($P < 0.001$).

Conclusions: The high VLN and HIVDR rates among FSWs underscore the need for targeted interventions to improve ART access and virologic monitoring to maximize the benefit of ART and limit the spread of HIV and HIVDR.

Key Words: female sex worker, HIV drug resistance, pretreatment drug resistance, acquired drug resistance, virologic failure, Ethiopia

(*J Acquir Immune Defic Syndr* 2022;89:566–574)

Received for publication August 19, 2021; accepted December 20, 2021.

From the [a]Department of Translational Medicine, Lund University, Malmo, Sweden; [b]TB/HIV Department, Ethiopian Public Health Institute, Addis Ababa, Ethiopia; [c]Department of Clinical Sciences, Lund University, Malmo, Sweden. [d]Africa Centre for Disease Prevention and Control, Africa Union Commission, Addis Ababa, Ethiopia; [e]NPO - HIV/AIDS, World Health Organization, Addis Ababa, Ethiopia; [f]Division of Global HIV & TB, Center for Global Health, Center for Disease Control and Prevention, Atlanta, GA

Supported by the President's Emergency Plan for AIDS Relief (PEPFAR) through HHS/CDC under the terms of 5U2GGH001178. P.M. was supported by the Swedish Research Council (grant numbers 2019-05235 and 2020-02344), the Österlund Foundation, and a donation to P.M. and P.B. through the Medical Faculty, Lund University, supported the research.

The authors have no conflicts of interest to disclose.

D.A.A., P.B., and P.M. were responsible for the overall study design. D.A.A., Y.K., A.R., J.H., C.Z., P.M., and T.T.B. were responsible for overall project coordination. A.R., J.C., J.H., and C.Z. coordinated the laboratory tests. D.A.A. and M.D. performed database entry and data cleaning and analysis. D.A.A. and P.M. performed the resistance testing and sequencing analyses. D.A.A., P.M., M.D., and P.B. interpreted the results. D.A.A., M.D., and P.M. wrote the manuscript. All authors revised the manuscript, provided important intellectual content, and approved the manuscript.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.jaids.com).

Correspondence to: Dawit Assefa Arimide, Department of Translational Medicine, Lund University, Clinical Virology, BMC B13, 22184 Lund, Sweden (e-mail: dawit.assefa@med.lu.se).

Copyright © 2021 Wolters Kluwer Health, Inc. All rights reserved.

INTRODUCTION

Female sex workers (FSWs) are at high risk of HIV infection and transmission and bear a disproportionately large burden of the disease.[1–4] As in many low- and middle-income countries (LMICs), Ethiopia has a generalized HIV epidemic primarily through heterosexual transmission. Since the beginning of the epidemic, FSWs have had high risk of HIV infection and were considered key drivers of HIV transmission.[5–8] According to the 2014 most at-risk population survey (MARPS), HIV prevalence among FSWs in Ethiopia was 24%, more than 5 times the prevalence of HIV in the general female population of reproductive age.[9]

In Ethiopia, antiretroviral therapy (ART) was rolled out free of charge in 2005, and since then, ART has been scaled up to provide access to all HIV-infected individuals.[10] In 2019, 473,261 people living with HIV were receiving ART in Ethiopia (75% coverage). Ethiopia has also implemented the test and treat recommendation since 2017. Accordingly, every person tested positive for HIV will start treatment, irrespective of his/her immunological and virologic status.[11]

Copyright © 2021 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

However, with rapid scale-up of ART, increased trends in emergency and transmission of HIV drug resistance (HIVDR) particularly to nonnucleoside reverse transcriptase inhibitors (NNRTI) have been reported from several LMICs.[12]

Despite advances in expanding access to HIV treatment and prevention, Ethiopia has limited access to regular virologic monitoring and HIVDR testing, delaying identification of patients with treatment failure, and increasing the risk of drug resistance mutations (DRMs) and onward transmission of HIVDR.[12,13] This may be more pronounced among FSWs who are highly mobile; are hard to reach; have low access to ART, adherence support, and VL monitoring; and have low care retention rates.[3,14] Moreover, FSWs are frequently exposed to violence, and women who report violence have poor ART adherence and viral suppression.[15,16] Addressing these barriers has the potential to reduce HIV infection and improve HIV treatment outcome.[17]

Data about ART uptake and treatment outcomes among FSWs in Ethiopia and other LMICs are limited. Given the potential risk of transmission to the general population, monitoring risk behavior and testing for VL nonsuppression (VLN; VL ≥ 1000 copies/mL) and HIVDR among FSWs can help inform prevention strategies to decrease HIVDR rates and onward transmission. Although FSWs are known to be at high risk of HIV infection and play an important role in HIV transmission dynamics, there is a lack of data on VLN and HIVDR among FSWs in Ethiopia. This study describes the prevalence of VLN, HIVDR mutations, and associated factors among FSWs in Ethiopia.

METHODS

This study was part of a larger cross-sectional study that assessed HIV prevalence and related risk factors among FSWs in Ethiopia in 2014. Data were collected through respondent-driven sampling in 11 cities (Addis Ababa, Mekele, Bahir Dar, Adama, Dire Dawa, Gambela, Hawassa, Metema, Kombolcha, Semera, and Shashamene) (Fig. 1). We defined FSWs as women who engage in sexual activity with the precondition of financial or in-kind benefits. The inclusion criteria for the study were women receiving money or other benefits for sex with 4 or more people within the past 30 days, aged 15 years or older, recruited by a peer, and providing consent for the interview and blood tests. The study methods have previously been described.[16] For this study, only women aged 18 years or older were included. In brief, 6 seed FSWs were selected to use coupons to recruit peers in each town. Eligible FSWs who provided informed consent participated in a face-to-face interview with nurses using a structured questionnaire in a private room. After completing the interview, participants provided blood specimens for HIV, CD4$^+$ T-cell counts, VL, and HIVDR testing and were given 3 coupons to recruit their peers into the study.

During the survey, sociodemographic characteristics and biobehavioral data were collected. Awareness of HIV status and prior ART exposure were used to classify study participants. Participants who reported that they were currently receiving ART were in the ART group, whereas those who reported not receiving ART (either ongoing or previous

FIGURE 1. Map of the cities in Ethiopia included in the 2014 study of HIV drug resistance among female sex workers. Details of the study are shown in the box. This figure was modified from Google Maps (https://www.google.com/maps/place/Ethiopia). full color online

treatment including antiretroviral for prevention of mother-to-child HIV transmission) were categorized as the ART-naive group. This categorization was also used to classify pretreatment drug resistance (PDR) in ART-naive participants and acquired HIVDR (ADR) in the ART group.

Participants were screened for HIV at the collection site through point-of-care rapid testing, which is used for HIV diagnosis in Ethiopia.[10] CD4$^+$ T-cell counts were obtained in nearby health facilities using the FACSCalibur and FACSCount systems (Becton Dickinson, San Jose, CA) according to the manufacturer's recommendations. Plasma was separated from whole blood and transported to the Ethiopian Public Health Institute (EPHI) where HIV-1 VL was determined using the Abbott RealTime HIV-1 assay (Abbott Molecular, Inc., Des Plaines, IL). Using 1000 copies/mL as a VL suppression threshold based on WHO recommendation,[18] all samples with VL ≥ 1000 copies/mL were shipped to the International Laboratory Branch of the Division of Global HIV & Tuberculosis, Center for Global Health, CDC (Atlanta, GA) for HIVDR genotyping (for details, see the Supplemental Digital Content, http://links.lww.com/QAI/B785).

HIV-1 Genotyping

Genotyping was performed using the ABI HIV-1 Genotyping Kit (Thermo Fisher Scientific, Waltham, MA).[19] In brief, a 1084-base pair fragment of HIV-1 pol (corresponding to the position 2243–3326 of HXB2; GenBank Accession Number: K03455) comprising amino acids 6–99 of the protease and 1–251 of the reverse transcriptase was generated by reverse transcription polymerase chain reaction (PCR) and nested PCR. The purified PCR fragments were then sequenced and analyzed on the ABI Prism 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA). Sequence assembly and editing were performed using the

Copyright © 2021 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2021 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

RECall V 2.0 HIV-1 sequencing analysis tool (University of British Columbia, Vancouver, Canada).[20] Sequence quality control was performed using the online Quality Control program of the Los Alamos HIV sequence database (https://www.hiv.lanl.gov/).

Drug Resistance Mutation Analysis

Surveillance drug resistance mutations were examined according to the Stanford Genotypic Resistance calibrated population resistance tool, version 6.0 (https://hivdb.stanford.edu/cpr). PDR levels were classified (low, <5%; moderate, 5%–15%; or high, >15%) using the World Health Organization (WHO) threshold survey protocol.[21] ADR was analyzed using the Stanford HIVdb program. Genotypic susceptibility scores ≥60 for each NNRTI and/or NRTI were considered a high level of resistance.[22]

Statistical Analysis

Statistical analysis was performed using SPSS, version 20 (Chicago, IL). We performed logistic regression analysis to identify potential risk factors for VLN and for PDR and ADR mutations. We used a multivariable model to assess biologically plausible interactions. Variables considered were age, education status, income from selling sex, khat chewing, heavy episodic drinking, sex-selling venues, frequency of sexual encounters per month, violence, being forced to sell sex, $CD4^+$ T-cell counts, vaginal discharge, and genital ulcers. In the model, we included a binary response, indicating detection of any VLN, PDR, and ADR mutations from each participant as an outcome. We analyzed all variables separately and entered those associated ($P < 0.2$) with the outcomes into the multivariable model. Odds ratios (crude and adjusted OR) with 95% confidence intervals (CI) were obtained by performing logistic regression analysis. *P*-values ≤0.05 were considered statistically significant. Although the data were collected using RDS sampling, our study focuses on a segment of samples (ie, participants with VL ≥ 1000 copies/mL) to extrapolate the HIVDR (ADR and PDR) prevalence among FSWs, and sample RDS weighting was not included in our analysis.

Ethical Considerations

The protocol was cleared by the Scientific and Ethical Research Office of EPHI and the Ethiopian Science and Technology Ministry Ethical Committee Institutional Review Boards (NHSBS-Round 1). This project was reviewed in accordance with CDC human research protection procedures (CDC-IRB #6343.0) and was determined to be research, but CDC investigators did not interact with human subjects or have access to identifiable data or specimens for research purposes. Individual written informed consent was obtained from each participant.

RESULTS

Figure 2 summarizes how participants were selected for HIVDR genotyping using the HIV test, VL, and the genotyping results. Of 4900 participants, 1172 (23.9%) were HIV-positive; of these, 1154 (98.5%) had VL results and were grouped based on self-report in the ART-naive or ART groups. The threshold for VL suppression was ≥1000 copies/mL per WHO recommendations.[18] Among 915 participants in the ART-naive group, 453 had VL <1000 copies/mL, indicating that they may have been exposed to ART but did not report it. The 521 samples (ART group, 59; ART-naive group, 462) with VL ≥1000 copies/mL were subjected for HIVDR genotyping. The genotyping success rates were 82.5% (381/462) for the ART-naive group and 66.1% (39/59) for the ART group. Overall HIVDR prevalence rates were 16.5% (63/381) for the ART-naive group and 74.4% (29/39) for the ART group.

We also calculated the ART uptake of participants (proportion of FSWs who tested HIV-positive and were receiving ART). Self-report of ART uptake was 20.7% (239/1154). However, including participants with VL <1000 copies/mL but who self-reported being ART naive, ART uptake was 60.0% (692/1154).

Prevalence of Pretreatment Drug Resistance

In the ART-naive group, 462 participants had VL ≥1000 copies/mL and 381 had genotyping results that were included in the PDR analysis. The median age was 25 years [interquartile range (IQR), 22–29 years]. The median HIV VL and $CD4^+$ T-cell count were 28,823 copies/mL (IQR, 7809–122,812 copies/mL) and 421 cells/mm^3 (IQR, 251–606 cells/mm^3), respectively.

Sixty-three (16.5% [95% CI: 12.8% to 20.3%]) of the genotyped specimens were associated with at least 1 major DRM. The highest prevalence of PDR was found against NNRTIs (55/381 [14.4%]), and five DRMs (K103N, Y181C, G190A/E/S, K101E/P, and V106M) accounted for most of the NNRTI PDR mutations (90.0%) (Table 1).

NRTI PDR mutations were detected in 10.5% (40/381) of the specimens, and 9.2% (35/381) had dual-class (NRTI and NNRTI) DRMs. The most prevalent NRTI DRMs were M184V and thymidine-analog mutations (TAMs; M41L, D67G/N, K70R, L210W, T215F/Y, and K219E/Q), accounting for 58.7% (37/63) and 27.0% (17/63) of the NRTI PDR, respectively. PI PDR mutations were detected in 0.8% (3/381) of the specimens (Table 1). According to the WHO classification of HIVDR prevalence, the overall PDR level among our participants was high (≥15%) but was moderate for NNRTIs and NRTIs and was low for PIs.

Prevalence and Patterns of Acquired Drug Resistance

Among 239 participants receiving ART, 59 (24.7%) had VL ≥ 1000 copies/mL. The median $CD4^+$ T-cell count and VL were 384 cells/mm^3 (IQR, 163–568 cells/mm^3) and 10,225 copies/mL (IQR, 2,802–95,220 copies/mL), respectively. Genotyping was successful for 39 (66.1%) of the specimens. Twenty-nine (74.4% [95% CI: 60.7% to 88.1%]) of the genotyped specimens had at least 1 major DRM. All 29 specimens had NNRTI DRMs, 23 (79.3%) had NRTI DRMs,

Copyright © 2021 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2021 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

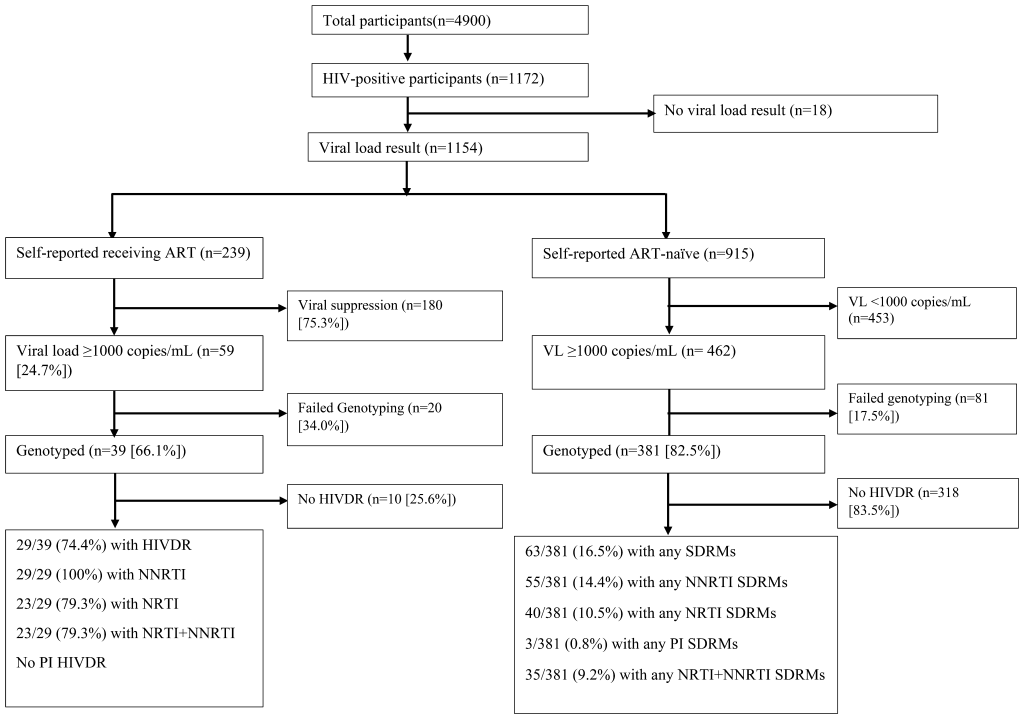


FIGURE 2. Flow chart of selection of female sex workers who participated in a biobehavioral survey and genotypic analysis of drug-resistant HIV in Ethiopia (2014).

and none had PI DRMs (Table 2). The most prevalent NNRTI DRMs were K103N, Y181C, and G190A. The most frequent NRTI DRMs were M184V (20 [69.0%]) and TAMs (18 [62.1%]; Table 2).

Dual-class resistance was present in 79.3% (23/29) of the specimens. Overall, the mean numbers of NRTI and NNRTI DRMs detected per specimen were 3.4 and 4.7, respectively. Four of the sequences had only one mutation (all NNRTI DRMs), three sequences had two mutations, and 22 (76.0%) of the sequences had ≥3 mutations.

Genotypic susceptibility scores of individual antiretroviral drugs indicated that many of the specimens had high levels of resistance to several of the most used first-line ART drugs in Ethiopia. Most specimens (69.0%) showed high-level resistance to lamivudine and tenofovir, nevirapine (100%), efavirenz (86.2%), and rilpivirine (51.7%; see Table 1, Supplemental Digital Content, http://links.lww.com/QAI/B784).

Factors Associated With VLN and HIVDR

In both bivariate and multivariate analyses, VLN was significantly associated with being forced into selling sex

TABLE 1. Frequency of Pretreatment Drug-Resistance Mutations Detected Among Female Sex Workers (n = 63) in Ethiopia (2014)

NNRTI SDRMs	N (%)*	NRTI SDRMs	N (%)*	PI SDRMs	N (%)*
K103N/S	30 (47.6)	M184V/I	37 (58.7)	L23I	1 (1.6)
Y181C	17 (27.0)	K65R	10 (15.9)	M46I	1 (1.6)
G190A/E/S	14 (22.2)	T215F/Y	8 (12.7)	I85V	1 (1.6)
K101E/P	8 (12.7)	Y115F	3 (4.8)		
V106M	8 (12.7)	L210W	3 (4.8)		
Y188H	3 (4.8)	M41L	2 (3.2)		
M230L	3 (4.8)	K70R	2 (3.2)		
L100I	1 (1.6)	L74V/I	2 (3.2)		
V179F	1 (1.6)	D67N	1 (1.6)		
P225H	1 (1.6)	T69D	1 (1.6)		
		K219R/Q	1 (1.6)		

SDRM, surveillance drug resistance mutation included in the WHO 2009 SDRM list; NNRTI, nonnucleoside reverse transcriptase inhibitor; NRTI, nucleoside reverse transcriptase inhibitor; PI, protease inhibitor.

*To calculate the percentages of each SDRM, we used 63 as the denominator, corresponding to the number of specimens with a PDR in the study.

Copyright © 2021 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2021 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

TABLE 2. Type and Frequency of Acquired Drug Resistance Mutations Detected Among Female Sex Workers With Viral Load Nonsuppression (n = 29) in Ethiopia (2014)

NNRTI DRMs	N (%)	NRTI DRMs	N (%)
K103N/S	18 (62.1)	M184IV	20 (69.0)
Y181C	10 (34.5)	K65R	6 (20.7)
G190A/E/S	7 (24.1)	K70R/E	6 (20.7)
H221HY	6 (20.7)	T215F/Y	5 (17.2)
A98G	5 (17.2)	K219Q	4 (13.8)
K101E/P	5 (17.2)	A62V	3 (10.3)
V106M	4 (13.8)	Y115F	3 (10.3)
V108I	4 (13.8)	D67N	2 (6.9)
L100I	3 (10.3)	M41L	1 (3.4)
E138A	2 (6.9)	L74V/I	1 (3.4)
V179D	2 (6.9)		
P225H	2 (6.9)		
F227FL	1 (3.4)		
M230L	1 (3.4)		
K238T	1 (3.4)		

($P < 0.036$), age 35 years or older ($P < 0.037$), and low $CD4^+$ T-cell counts [<350 cells/mm^3 ($P < 0.001$; Table 3]. In bivariate analysis, PDR was significantly associated with low CD4 counts ($P < 0.001$) and ever giving birth ($P < 0.03$). However, in multivariate analysis, only low CD4 counts remained significantly associated with PDR ($P < 0.001$). Moreover, low CD4 counts were significantly associated with ADR in both bivariate and multivariate analyses ($P < 0.001$).

DISCUSSION

To the best of our knowledge, this is the first national study that comprehensively describes the level of VLN and HIVDR among FSWs in Ethiopia. Overall, our results showed a high prevalence of HIVDR (PDR, 16.5%; ADR, 74.4%), poor ART uptake (20.7%), and high VLN (24.7%) with multiple DRMs among participants, which indicates high risk of HIVDR transmission to the general population.

We found high prevalence of PDR, particularly toward NNRTIs. This level is higher than the PDR level reported among the general population in Ethiopia (4%–6%).[10,23–29] Consistent with our findings, other studies have reported high PDR levels (10%–48%) among FSWs in different countries, including those in sub-Saharan Africa.[28,30–34] Moreover, previous studies also have shown a higher PDR rate among communities and groups with high-risk behaviors.[29,35] This highlights the vulnerability of FSWs to HIVDR and the risk of onward transmission to the general population.

After 10 years of ART roll out in Ethiopia, the prevalence of NNRTI PDR in our study is above the WHO-recommended levels to replace NNRTIs with dolutegravir in first-line regimens.[36] Similar findings have been reported in other LMICs, which depend on standardized first-line ART.[37,38] The high NNRTI DRM prevalence might in part be due to the low genetic barrier of these drugs and their wide use for prevention of mother-to-child HIV transmission and as part of the standard first-line ART regimen.[12,39]

In our study, K103N, Y181C, G190A/E/S, K101E/P, and V106M accounted for most of the NNRTI PDR mutations. Strains with K103N and other NNRTI mutations have a fitness similar to wild-type virus, and the mutation can persist for years in HIV-positive individuals.[39–41] It is, therefore, likely that the high prevalence of these mutations is a consequence of frequent transmission from sexual partners with unsuppressed viremia to FSWs. Consistent with our study, 2 meta-analyses have shown that these DRMs are the dominant SRDRMs in sub-Saharan Africa.[37,38]

The most common NRTI PDR mutations detected in our study were M184V, K65R, and TAMs. However, both M184V and K65R revert to wild type relatively quickly in the absence of ART[42,43] and would be expected to be found at low frequencies among individuals with PDR. Nevertheless, M184V is one of the most detected PDR mutations in most countries, including sub-Saharan African countries.[38]

We found that FSWs had poor ART uptake. Only 1 in 5 HIV-positive participants were receiving ART, which is consistent with results of other studies in sub-Saharan Africa, showing generally poor ART uptake among FSWs (range, 26%–38%).[44–47] However, in our study, more than half of the participants with self-reported ART-naive status had VL < 1000 copies/mL, indicating that they may have been exposed to ART but did not disclose this history.[48–51] A recent report from Ethiopia also showed that only 26% of HIV-positive FSWs were receiving ART.[52] Consistent with our results, several studies in sub-Saharan Africa have shown high levels of VLN among FSWs.[28,53–56] This might be due to multiple barriers, such as stigma related to HIV and sex work or high mobility, that prevent FSWs from accessing the HIV care continuum.[20] Moreover, FSWs are frequently exposed to violence, and women who report violence have poor ART adherence and viral suppression.[15,16]

Improving access to ART for FSWs not only will improve the survival and health of this population but also will reduce the risk of HIV transmission to their clients and could lower HIV transmission at the general population level.[4,57–59] Our findings highlight the importance of identifying potential factors that prevent FSWs from accessing HIV treatment services. Improving ART uptake could help improve outcomes for clients in national HIV control programs.[57,58] Furthermore, targeting scale-up of VL monitoring among FSWs could help ensure timely therapy changes for those with virologic failure, according to the national treatment guidelines.[60]

In our study, a high proportion of FSWs with VLN carried dual-class DRMs with high genotypic susceptibility scores to several commonly used first-line ART drugs. Consistent with our study, other studies have reported high DRM frequency with a complex pattern in patients with the prolonged use of failing regimens in the absence of VL monitoring.[61,62] Besides the resultant limitations in the choice of effective treatment regimens for patients with VLN, the high prevalence of HIVDR detected among participants in our study highlights the potential risk of HIVDR transmission to the general population. Furthermore, when individuals

Copyright © 2021 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2021 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

TABLE 3. Bivariate and Multivariate Analyses for Factors Associated With Virologic Failure and HIV Drug Resistance Among Female Sex Workers in Ethiopia (2014)

	VLN			PDR			ADR		
	N	OR (95% CI)	aOR (95% CI)	N	OR (95% CI)	aOR (95% CI)	N	OR (95% CI)	aOR (95% CI)
Age, yr									
18–24	30	Ref		163	Ref		128	Ref	
25–34	139	2.36 (0.77 to 7.21)	3.02 (0.83 to 11.06)	175	1.57 (0.87 to 2.85)†	1.47 (0.74 to 2.92)	392	1.78 (0.51 to 6.18)	
≥35	70	2.25 (0.69 to 7.33)†	4.09 (1.04 to 16.1)*	42	1.84 (0.77 to 4.39)†	1.69 (0.65 to 4.41)	141	3.2 (0.86 to 11.83)	
Income (monthly; currency in USD)									
<$100)	169	Ref		230	Ref		432	Ref	
≥$100	70	1.20 (0.64 to 2.27)		149	1.19 (0.69 to 2.06)		229	0.53 (0.19 to 1.46)	
Level of education									
No education	89	Ref		128	Ref		241	Ref	
Primary 1st cycle (grade 1–4)	32	0.68 (0.23 to 2.01)		63	0.49 (0.20 to 1.20)		99	0.24 (0.03 to 1.92)	
Primary 2nd cycle (grade 5–8)	96	1.37 (0.70 to 2.70)		139	0.70 (0.37 to 1.32)		250	1.36 (0.59 to 3.18)	
Secondary and above	22	2.55 (0.95 to 6.86)		50	0.86 (0.37 to 1.20)		71	0.98 (0.19 to 4.94)	
Ever given birth									
No	52	Ref		130	Ref		194	Ref	
Yes	187	1.12 (0.54 to 2.31)		250	2.02 (1.07 to 3.82)*	1.56 (0.76 to 3.20)	467	1.37 (0.49 to 3.83)	
Number of sexual partners/month									
4–10	114	Ref		164	Ref		267	Ref	
≥11	125	1.76 (0.96 to 3.21)¥	1.82 (0.89, 3.73)	216	1.02 (0.59 to 1.75)		394	1.99 (0.88 to 4.51)¥	1.85 (0.71–4.83)
Sex selling venue									
Street	23	Ref		89	Ref		152	Ref	
Local drinking houses	86	1.26 (0.38 to 4.16)		83	1.74 (0.80 to 3.79)		196	0.93 (0.23 to 3.73)	
Spa/massage/beauty salon/own house	31	1.39 (0.35 to 5.44)		23	1.23 (0.36 to 4.20)		56	0.79 (0.14 to 4.38)	
Red light houses	33	1.52 (0.40 to 5.81)		33	1.30 (0.45 to 3.76)		78	1.01 (0.20 to 5.08)	
Bar/hotel	49	1.90 (0.55 to 6.59)		136	1.01 (0.47 to 2.15)		140	1.09 (0.24 to 4.84)	
Others	17	4.22 (1.00 to 17.80)		16	0.39 (0.05 to 3.21)		39	2.11 (0.35 to 12.59)	
Heavy episodic drinking in the past month									
No	72	Ref		172	Ref		260	Ref	
Yes	32	1.62 (0.62 to 4.26)		103	0.60 (0.29 to 1.23)		158	0.32 (0.04 to 2.66)	
Frequency of khat chewing per week									
Never	168	Ref		168	Ref		347	Ref	
Less than once	23	1.17 (0.43 to 3.17)		51	0.40 (0.15 to 1.08)		77	1.08 (0.29 to 4.02)	
1–2 d	9	0.95 (0.19 to 4.74)		40	0.30 (0.09 to 1.02)		49	1.76 (0.34 to 9.03)	
3–4 d	3	1.65 (0.15 to 8.73)		24	0.52 (0.15 to 1.86)		30	0	
5–7 d	36	1.46 (0.66 to 3.22)		97	0.72 (0.38 to 1.39)		158	0.74 (0.20 to 2.66)	

(*continued on next page*)

Copyright © 2021 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2021 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

TABLE 3. (*Continued*) Bivariate and Multivariate Analyses for Factors Associated With Virologic Failure and HIV Drug Resistance Among Female Sex Workers in Ethiopia (2014)

	VLN			PDR			ADR		
	N	OR (95% CI)	aOR (95% CI)	N	OR (95% CI)	aOR (95% CI)	N	OR (95% CI)	aOR (95% CI)
Physically beaten in the past 12 mo									
No	209	Ref		313	Ref		599	Ref	
Yes	30	1.13 (0.47 to 2.69)		67	0.74 (0.35 to 1.59)		61	0.52 (0.19 to 2.39)	
Forced into selling sex									
No	213	Ref		326	Ref		581	Ref	
Yes	26	3.03 (1.31 to 6.99)*	2.79 (1.07 to 7.27)*	54	0.59 (0.24 to 1.44)		80	3.77 (1.37 to 10.36)*	3.21 (0.99–10.38)
Unusual vaginal discharge in the past 12 months									
No	194	Ref		311	Ref		583	Ref	
Yes	45	1.14 (0.54 to 2.38)		69	1.36 (0.70 to 2.64)		78	0.71 (0.23 to 2.19)	
Genital ulcer in the past 12 mo									
No	215	Ref		335	Ref		260	Ref	
Yes	24	1.02 (0.38 to 2.70)		45	1.30 (0.59 to 2.86)		158	0.32 (0.04 to 2.51)	
CD4 count (cell/mm^3)									
Lower (<350)	48	4.19 (2.11 to 8.32)*	4.67 (2.23 to 9.77)*	129	3.44 (1.90 to 6.23)*	3.24 (1.78 to 5.89)*	124	6.51 (2.77 to 15.32)*	7.25 (2.95–17.83)*
Higher (≥350)	172	Ref		215	Ref		491	Ref	

ADR, acquired drug resistance; CI, confidence intervals; OR, odds ratio; aOR, adjusted odds ratio.
*$P \leq 0.05$.
†$P < 0.2$.
‡Facilities other than those mentioned in the list.

carrying multiple DRMs are switched to second-line therapy, there is a risk of introducing a functional monotherapy, which may be associated with substantial risk of subsequent virologic failure and emergence of HIVDR.

Among FSWs receiving ART with VL ≥1000 copies/mL, 26% had no HIVDR mutations, which suggests that nonadherence could be the possible cause for the detected virologic failure. This shows the importance of strengthening adherence among FSWs and using HIVDR testing before treatment switches to reduce the cost associated with prematurely switching to costly second-line regimens.

We found that participants aged 35 years or older experienced higher prevalence of VLN compared with younger participants aged 18–24 years. This finding contrasts with those of a study in Uganda, where young FSWs (18–24 years) experienced higher prevalence of virologic failure than elder FSWs (>35 years).[28] This difference might in part be due to the difference in the research design of the studies. The study in Uganda was conducted among FSWs with virologic failure identified during the follow-up, whereas our study collected lifetime ART status, and the elder participants in our study might be more likely to have treatment failure due to prolonged ART exposure compared with younger participants. The longer the duration of ART treatment, the higher the odds of developing drug resistance leading to treatment failure.[65]

Moreover, participants who reported being forced into sex work had higher prevalence of VLN. Women forced into sex work are especially vulnerable because they cannot control their environment.[16] This may increase the risk of substance use as a coping mechanism, which can decrease the efficacy of ART (including poor adherence), potentially leading to treatment failure.[66] Our results showed that low CD4$^+$ T-cell count (<350 cells/mm^3) was associated with VLN and DRMs among ART-experienced participants, suggesting disease progression among those with VLN and ADR and underlining the importance of DRM monitoring to improve individual outcomes.

Our study has several limitations. One limitation of our study and similar studies is that the duration of HIV infection before sampling is unknown. Because our classification of ART status among participants was based on self-report, there is a risk of misclassification if participants did not disclose previous ART exposure for fear of discrimination, which has been documented in other studies.[50,51,67] We used 1000 copies/mL as the cutoff for VLN; however, other studies have shown the development of HIVDR among patients with low-level viremia.[68] The overall genotyping success rate was 80.6%, which might have affected the overall study results. Although the data used in our analysis were collected using RDS sampling, our study only focused on a segment of the samples (ie, participants with VL ≥1000

Copyright © 2021 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2021 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

copies/mL) to extrapolate the HIVDR prevalence among FSWs; therefore, weighting was not included in the data analysis. We also did not collect information about ART regimens, ART duration, or ART adherence, which could affect the level of ADR. Finally, some of the ART-experienced participants with VLN might have been infected with a DRM virus.

CONCLUSIONS

The suboptimal ART uptake and high VLN and HIVDR levels detected among FSWs underscore the importance of programmatic intervention to improve ART access and routine virologic monitoring among this population to maximize the benefit of ART and limit the spread of HIV, HIVDR, and disease progression. Our findings also demonstrate the need for implementation of HIVDR genotyping to optimize the selection of regimen and transition to dolutegravir-based first-line ART in Ethiopia.

ACKNOWLEDGMENTS

The authors express their deep appreciation to the study participants and all staff involved in the survey. The authors also acknowledge the support given by the EPHI research laboratory team and CDC in Ethiopia and Atlanta.

REFERENCES

1. Baral S, Beyrer C, Muessig K, et al. Burden of HIV among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis. *Lancet Infect Dis.* 2012;12:538–549.
2. Shannon K, Crago AL, Baral SD, et al. The global response and unmet actions for HIV and sex workers. *Lancet.* 2018;392:698–710.
3. Doshi RH, Sande E, Ogwal M, et al. Progress toward UNAIDS 90-90-90 targets: a respondent-driven survey among female sex workers in Kampala, Uganda. *PLoS One.* 2018;13:e0201352.
4. Prüss-Ustün A, Wolf J, Driscoll T, et al. HIV due to female sex work: regional and global estimates. *PLoS One.* 2013;8:e63476.
5. Aklilu M, Messele T, Tsegaye A, et al. Factors associated with HIV-1 infection among sex workers of Addis Ababa, Ethiopia. *AIDS.* 2001;15: 87–96.
6. Mehret MKL, Shanko B, Belete F. Sexual behaviours and some social features off female sex workers in the city of Addis Ababa. *Ethiopian J Health Develop.* 1990;4:133–13.
7. Mehret MKL, Zewdie D. Ayehunie S, et al. HIV-1 infection and some related risk factors among female sex workers in Addis Ababa. *Ethiopian J Health Develop.* 1990;4: 171 –176.
8. Mehret MKL, Zewdie D. et al. HIV-1 infection and related risk factors among female sex workers in urban areas of Ethiopia. *Ethiopian J Health Develop.* 1990;4(2, Suppl):163–170.
9. EPHI. *Ethiopian National Key Population HIV Bio-Behavioral Surveillance Round I, 2013 Report: EPHI.* Addis Ababa. Ethiopia,: EPHI; 2014.
10. Arimide DA, Abebe A, Kebede Y, et al. HIV-genetic diversity and drug resistance transmission clusters in Gondar, Northern Ethiopia, 2003-2013. *PLoS One.* 2018;13:e0205446.
11. Tesfaye B, Ermias D, Moges S, et al. Effect of the test and treat strategy on mortality among HIV-positive adult clients on antiretroviral treatment in public hospitals of addis Ababa, Ethiopia. *HIV AIDS (Auckl).* 2021;13: 349–360.
12. Gupta RK, Jordan MR, Sultan BJ, et al. Global trends in antiretroviral resistance in treatment-naive individuals with HIV after rollout of antiretroviral treatment in resource-limited settings: a global collaborative study and meta-regression analysis. *Lancet.* 2012;380:1250–1258.
13. Kityo C, Thompson J, Nankya I, et al. HIV drug resistance mutations in non-B subtypes after prolonged virological failure on NNRTI-based first-line regimens in sub-saharan Africa. *J Acquir Immune Defic Syndr.* 2017;75:e45–e54.
14. Van Blerk L. AIDS, mobility and commercial sex in Ethiopia: implications for policy. *AIDS Care.* 2007;19:79–86.
15. Hatcher AM, Smout EM, Turan JM, et al. Intimate partner violence and engagement in HIV care and treatment among women: a systematic review and meta-analysis. *AIDS.* 2015;29:2183–2194.
16. Amogne MD, Balcha TT, Agardh A. Prevalence and correlates of physical violence and rape among female sex workers in Ethiopia: a cross-sectional study with respondent-driven sampling from 11 major towns. *BMJ Open.* 2019;9:e028247.
17. Decker MR, Wirtz AL, Pretorius C, et al. Estimating the impact of reducing violence against female sex workers on HIV epidemics in Kenya and Ukraine: a policy modeling exercise. *Am J Reprod Immunol.* 2013;69suppl 1:122–132.
18. WHO. *World Health Organization Global Strategy for the Surveillance and Monitoring of HIV Drug Resistance*. Geneva, Switzerland: WHO press; 2012.
19. Rosemary A, Chika O, Jonathan O, et al. Genotyping performance evaluation of commercially available HIV-1 drug resistance test. *PLoS One.* 2018;13:e0198246.
20. Woods CK, Brumme CJ, Liu TF, et al. Automating HIV drug resistance genotyping with RECall, a freely accessible sequence analysis tool. *J Clin Microbiol.* 2012;50:1936–1942.
21. Bennett DE, Myatt M, Bertagnolio S, et al. Recommendations for surveillance of transmitted HIV drug resistance in countries scaling up antiretroviral treatment. *Antivir Ther.* 2008;13(suppl 2):25–36.
22. Liu TF, Shafer RW. Web resources for HIV type 1 genotypic-resistance test interpretation. *Clin Infect Dis.* 2006;42:1608–1618.
23. Kassu A, Fujino M, Matsuda M, et al. Molecular epidemiology of HIV type 1 in treatment-naive patients in north Ethiopia. *AIDS Res Hum Retroviruses.* 2007;23:564–568.
24. Mulu A, Lange T, Liebert UG, et al. Clade homogeneity and Pol gene polymorphisms in chronically HIV-1 infected antiretroviral treatment naive patients after the roll out of ART in Ethiopia. *BMC Infect Dis.* 2014;14:158.
25. Huruy K, Maier M, Mulu A, et al. Limited increase in primary HIV-1C drug resistance mutations in treatment naive individuals in Ethiopia. *J Med Virol.* 2015;87:978–984.
26. Abdissa A, Yilma D, Fonager J, et al. Drug resistance in HIV patients with virological failure or slow virological response to antiretroviral therapy in Ethiopia. *BMC Infect Dis.* 2014;14:181.
27. Telele NF, Kalu AW, Gebre-Selassie S, et al. Pretreatment drug resistance in a large countrywide Ethiopian HIV-1C cohort: a comparison of Sanger and high-throughput sequencing. *Sci Rep.* 2018;8:7556.
28. Namale G, Kamacooko O, Bagiire D, et al. Sustained virological response and drug resistance among female sex workers living with HIV on antiretroviral therapy in Kampala, Uganda: a cross-sectional study. *Sex Transm Infect.* 2019;95:405–411.
29. Chen I, Connor MB, Clarke W, et al. Antiretroviral drug use and HIV drug resistance among HIV-infected black men who have sex with men: HIV prevention trials network 061. *J Acquir Immune Defic Syndr.* 2015; 69:446–452.
30. Coetzee J, Hunt G, Jaffer M, et al. HIV-1 viraemia and drug resistance amongst female sex workers in Soweto, South Africa: a cross sectional study. *PLoS One.* 2017;12:e0188606.
31. da Costa LM, Frade PCR, Blandtt LDS, et al. HIV-1 genetic diversity and transmitted drug resistance mutations in female sex workers from a Brazilian municipality in the amazon region. *AIDS Res Hum Retroviruses.* 2020;36:99–100.
32. Diallo M, Behanzin L, Guedou FA, et al. HIV treatment response among female sex workers participating in a treatment as prevention demonstration project in Cotonou, Benin. *PLoS One.* 2020;15: e0227184.
33. Carobene M, Bolcic F, Farias MS, et al. HIV, HBV, and HCV molecular epidemiology among trans (transvestites, transsexuals, and transgender) sex workers in Argentina. *J Med Virol.* 2014;86:64–70.
34. Sampathkumar R, Shadabi E, La D, et al. Naturally occurring protease inhibitor resistance mutations and their frequencies in HIV proviral sequences of drug-naïve sex workers in Nairobi, Kenya. *Retrovirology.* 2014;11:P133.
35. Weinstock HS, Zaidi I, Heneine W, et al. The epidemiology of antiretroviral drug resistance among drug-naive HIV-1-infected persons in 10 US cities. *J Infect Dis.* 2004;189:2174–2180.

Copyright © 2021 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2021 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

36. WHO. *Guidelines on the Public Health Response to Pretreatment HIV Drug Resistance*. Geneva, Switzerland. 2017. Available at: http://apps.who.int/iris/bitstream/10665/255880/1/9789241550055-eng.pdf. Accessed October 16, 2020.
37. Chimukangara B, Lessells RJ, Rhee SY, et al. Trends in pretreatment HIV-1 drug resistance in antiretroviral therapy-naive adults in South Africa, 2000-2016: a pooled sequence analysis. *EClinicalMedicine.* 2019;9:26–34.
38. Rhee SY, Blanco JL, Jordan MR, et al. Geographic and temporal trends in the molecular epidemiology and genetic mechanisms of transmitted HIV-1 drug resistance: an individual-patient- and sequence-level meta-analysis. *PLoS Med.* 2015;12:e1001810.
39. Clutter DS, Jordan MR, Bertagnolio S, et al. HIV-1 drug resistance and resistance testing. *Infect Genet Evol.* 2016;46:292–307.
40. Castro H, Pillay D, Cane P, et al. Persistence of HIV-1 transmitted drug resistance mutations. *J Infect Dis.* 2013;208:1459–1463.
41. Kuhnert D, Kouyos R, Shirreff G, et al. Quantifying the fitness cost of HIV-1 drug resistance mutations through phylodynamics. *PLoS Pathog.* 2018;14:e1006895.
42. Castro H, Pillay D, Cane P, et al. Persistence of HIV-1 transmitted drug resistance mutations. *J Infect Dis.* 2013;208:1459–1463.
43. Wertheim JO, Oster AM, Johnson JA, et al. Transmission fitness of drug-resistant HIV revealed in a surveillance system transmission network. *Virus Evol.* 2017;3:vex008.
44. Lancaster KE, Cernigliaro D, Zulliger R, et al. HIV care and treatment experiences among female sex workers living with HIV in sub-Saharan Africa: a systematic review. *Afr J AIDS Res.* 2016;15:377–386.
45. Mountain E, Mishra S, Vickerman P, et al. Antiretroviral therapy uptake, attrition, adherence and outcomes among HIV-infected female sex workers: a systematic review and meta-analysis. *PLoS One.* 2014;9:e105645.
46. Holland CE, Papworth E, Billong SC, et al. Antiretroviral treatment coverage for men who have sex with men and female sex workers living with HIV in Cameroon. *J Acquir Immune Defic Syndr.* 2015;68(Suppl 2): S232–S240.
47. Cowan FM, Mtetwa S, Davey C, et al. Engagement with HIV prevention treatment and care among female sex workers in Zimbabwe: a respondent driven sampling survey. *PLoS One.* 2013;8:e77080.
48. Huerga H, Shiferie F, Grebe E, et al. A comparison of self-report and antiretroviral detection to inform estimates of antiretroviral therapy coverage, viral load suppression and HIV incidence in Kwazulu-Natal, South Africa. *BMC Infect Dis.* 2017;17:653.
49. Fogel JM, Wang L, Parsons TL, et al. Undisclosed antiretroviral drug use in a multinational clinical trial (HIV Prevention Trials Network 052). *J Infect Dis.* 2013;15:1624–1628.
50. Moyo S, Gaseitsiwe S, Powis KM, et al. Undisclosed antiretroviral drug use in Botswana: implication for national estimates. *AIDS.* 2018;32:1543–1546.
51. Kim AA, Mukui I, Young PW, et al. Undisclosed HIV infection and antiretroviral therapy use in the Kenya AIDS indicator survey 2012: relevance to national targets for HIV diagnosis and treatment. *AIDS.* 2016;30:2685–2695.
52. *PSI. Community HIV Care and Treatment for Female Sex Workers in Ethiopia: Successful Servic Provision through Drop- in Centers (THPEE774)* , 2016. Available at: https://www.psi.org/publication/community-hiv-care-and-treatment-for-female-sex-workers-in-ethiopia-successful-service-provision-through-drop-in-centers-thpee774. Accessed October 16, 2020.
53. Mountain E, Pickles M, Mishra S, et al. The HIV care cascade and antiretroviral therapy in female sex workers: implications for HIV prevention. *Expert Rev Anti Infect Ther.* 2014;12:1203–1219.
54. Cowan FM, Davey CB, Fearon E, et al. The HIV care cascade among female sex workers in Zimbabwe: results of a population-based survey from the sisters antiretroviral therapy programme for prevention of HIV, an integrated response (SAPPH-IRe) trial. *J Acquir Immune Defic Syndr.* 2017;74:375–382.
55. Lancaster KE, Powers KA, Lungu T, et al. The HIV care continuum among female sex workers: a key population in lilongwe, Malawi. *PLoS One.* 2016;11:e0147662.
56. Lindman J, Djalo MA, Biai A, et al. The HIV care continuum and HIV-1 drug resistance among female sex workers: a key population in Guinea-Bissau. *AIDS Res Ther.* 2020;17:33.
57. Delva W, Eaton JW, Meng F, et al. HIV treatment as prevention: optimising the impact of expanded HIV treatment programmes. *PLoS Med.* 2012;9:e1001258.
58. Alary M, Lowndes CM, Van de Perre P, et al. Scale-up of combination prevention and antiretroviral therapy for female sex workers in West Africa: time for action. *AIDS.* 2013;27:1369–1374.
59. Moses S, Ramesh BM, Nagelkerke NJ, et al. Impact of an intensive HIV prevention programme for female sex workers on HIV prevalence among antenatal clinic attenders in Karnataka state, south India: an ecological analysis. *AIDS.* 2008;22(suppl 5):S101–S108.
60. MOH E. *National Guideline for Comprehensive HIV Prevention, Care, and Treatment*; 2017. Available at: https://www.afro.who.int/publications/national-consolidated-guidelines-comprehensive-hiv-prevention-care-and-treatment. Accessed October 16, 2020.
61. Etta EM, Mavhandu L, Manhaeve C, et al. High level of HIV-1 drug resistance mutations in patients with unsuppressed viral loads in rural northern South Africa. *AIDS Res Ther.* 2017;14:36.
62. Gupta RK, Hill A, Sawyer AW, et al. Virological monitoring and resistance to first-line highly active antiretroviral therapy in adults infected with HIV-1 treated under WHO guidelines: a systematic review and meta-analysis. *Lancet Infect Dis.* 2009;9:409–417.
63. Kwon EH, Musema GMA, Boelter J, et al. HIV-1 subtypes and drug resistance mutations among female sex workers varied in different cities and regions of the Democratic Republic of Congo. *PLoS One.* 2020;15: e0228670.
64. Barth RE, van der Loeff MF, Schuurman R, et al. Virological follow-up of adult patients in antiretroviral treatment programmes in sub-Saharan Africa: a systematic review. *Lancet Infect Dis.* 2010;10: 155–166.
65. Feleke R, Geda B, Teji Roba K, et al. Magnitude of antiretroviral treatment failure and associated factors among adult HIV-positive patients in Harar public hospitals, Eastern Ethiopia. *SAGE Open Med.* 2020;8:2050312120906076.
66. Gupta J, Raj A, Decker MR, et al. HIV vulnerabilities of sex-trafficked Indian women and girls. *Int J Gynaecol Obstet.* 2009;107:30–34.
67. Kahle EM, Kashuba A, Baeten JM, et al. Unreported antiretroviral use by HIV-1-infected participants enrolling in a prospective research study. *J Acquir Immune Defic Syndr.* 2014;65:e90–4.
68. von Braun A, Sekaggya-Wiltshire C, Bachmann N, et al. HIV-1 drug resistance among Ugandan adults attending an urban out-patient clinic. *J Acquir Immune Defic Syndr.* 2018;78:566–573.

 Copyright © 2021 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2021 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

Paper V

viruses

MDPI

Article

Pre-Treatment Integrase Inhibitor Resistance and Natural Polymorphisms among HIV-1 Subtype C Infected Patients in Ethiopia

Dawit Assefa Arimide [1,2,*], Zsófia Ilona Szojka [3], Kidist Zealiyas [2,†], Atsbeha Gebreegziabxier [2,†], Fekadu Adugna [4], Sviataslau Sasinovich [1], Per Björkman [1,5] and Patrik Medstrand [1,6]

1 Department of Translational Medicine, Lund University, 205 02 Malmo, Sweden; sviataslau.sasinovich@med.lu.se (S.S.); per.bjorkman@med.lu.se (P.B.); patrik.medstrand@med.lu.se (P.M.)
2 TB/HIV Department, Ethiopian Public Health Institute, Addis Ababa 1242, Ethiopia; kzealiyas@gmail.com (K.Z.); atsbehag@gmail.com (A.G.)
3 Department of Laboratory Medicine, Lund University, 222 42 Lund, Sweden; zsofia_ilona.szojka@med.lu.se
4 NPO–HIV/AIDS, World Health Organization, Addis Ababa 3069, Ethiopia; adugnadadif@who.int
5 Department of Infectious Diseases, Skane University Hospital, 205 02 Malmo, Sweden
6 Clinical Microbiology and Infectious Disease Control, Skane University Hospital, 221 85 Lund, Sweden
* Correspondence: dawit.assefa@med.lu.se
† These authors contributed equally to this work.

Abstract: Dolutegravir-based antiretroviral therapy (ART) has been scaled up in many developing countries, including Ethiopia. However, subtype-dependent polymorphic differences might influence the occurrence of HIV-drug-resistance mutations (HIVDRMs). We analyzed the prevalence of pre-treatment integrase strand transfer inhibitor (INSTI) HIVDRMs and naturally occurring polymorphisms (NOPs) of the integrase gene, using plasma samples collected as part of the national HIVDR survey in Ethiopia in 2017. We included a total of 460 HIV-1 integrase gene sequences from INSTI-naïve (n = 373 ART-naïve and n = 87 ART-experienced) patients. No dolutegravir-associated HIVDRMs were detected, regardless of previous exposure to ART. However, we found E92G in one ART-naïve patient specimen and accessory mutations in 20/460 (4.3%) of the specimens. Moreover, among the 288 integrase amino acid positions of the subtype C, 187/288 (64.9%) were conserved (<1.0% variability). Analysis of the genetic barrier showed that the Q148H/K/R dolutegravir resistance pathway was less selected in subtype C. Docking analysis of the dolutegravir showed that protease- and reverse-transcriptase-associated HIVDRMs did not affect the native structure of the HIV-1 integrase. Our results support the implementation of a wide scale-up of dolutegravir-based regimes. However, the detection of polymorphisms contributing to INSTI warrants the continuous surveillance of INSTI resistance.

Keywords: dolutegravir; integrase strand transfer inhibitor (INSTI); naturally occurring polymorphisms (NOPs); pretreatment; HIV drug resistance (HIVDR); docking; genetic barrier; Ethiopia

check for updates

Citation: Arimide, D.A.; Szojka, Z.I.; Zealiyas, K.; Gebreegziabxier, A.; Adugna, F.; Sasinovich, S.; Björkman, P.; Medstrand, P. Pre-Treatment Integrase Inhibitor Resistance and Natural Polymorphisms among HIV-1 Subtype C Infected Patients in Ethiopia. *Viruses* **2022**, *14*, 729. https://doi.org/10.3390/v14040729

Academic Editor: Stefano Aquaro

Received: 4 March 2022
Accepted: 28 March 2022
Published: 30 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Following the global increase of pre-treatment drug resistance (PDR) to non-nucleoside reverse transcriptase inhibitors (NNRTIs), the World Health Organization (WHO) recommended the transition from NNRTI to integrase strand transfer inhibitor (INSTI)-based regimens in both treatment-naïve and treatment-experienced patients [1–3]. Several low- and middle-income countries have already transitioned to the dolutegravir (DTG)-based regimen, and many more are in the planning phase, so millions of people living with HIV will soon receive DTG combined with two nucleoside reverse transcriptase inhibitors (NRTIs) as first- and second-line therapies [2,4].

HIV-1 integrase (IN), which comprises 288 amino acids encoded by the 5′-end of the HIV *pol* (polymerase) gene, plays a vital role in HIV-1 replication by catalyzing two distinct

reactions: 3′-end processing and strand transfer [5–7]. IN consists of three functional domains: the N-terminal domain (NTD) (aa :1–50), which contains a highly conserved histidine–histidine–cysteine–cysteine ($H_{12}H_{16}C_{40}C_{43}$) motif that coordinates zinc binding and favors multimerization of the IN subunit [8]; the catalytic core domain (CCD) (aa: 51–212), which contains the catalytic triad $D_{64}D_{116}E_{152}$ (known as the DDE motif) that plays an essential role in IN enzymatic activity; and the C-terminal domain (CTD) (aa: 213–288), which is involved in binding to viral and cellular DNA, and in protein oligomerization and interactions with the reverse transcriptase [5–7,9].

INSTIs inhibit the HIV-1 integrase strand transfer steps to block the integration of HIV viral DNA into the host cell chromosomal DNA through competitive binding to the enzyme's active site [7,10]. There are currently five US Food and Drug Administration (FDA)-approved drugs belonging to this therapeutic class: raltegravir (RAL), elvitegravir (EVG), DTG, bictegravir (BIC), and cabotegravir (CAB) [11]. RAL and EVG were the first-generation INSTIs to be used clinically; however, their relatively low genetic barrier for resistance and the extensive cross-resistance between them limit their efficiency [12,13]. DTG and BIC are second-generation INSTIs shown to be highly effective in both treatment-naive and treatment-experienced individuals with good tolerability and a high genetic barrier to resistance [12,14]. Pooled analysis of resistance data conducted by Yang et al. (2019) indicated that the development of resistance to DTG and BIC was rare [12]. However, with the wide scale-up of DTG, gradual development and transmission of HIVDR against INSTIs will be inevitable and can render existing therapies ineffective, thereby increasing the risk of virological failure, disease progression, and mortality [12,15–28].

Although non-B subtypes dominate the global HIV epidemic, most clinical and virological studies on DTG were based on subtype B. However, subtype-dependent differences in naturally occurring polymorphisms (NOPs) have been implicated in the development of different mutational pathways, leading to varying levels of drug resistance against INSTIs among different HIV-1 subtypes [5,13,29–34]. Q148H and G140S, which confer resistance to RAL and EVG and cross-resistance to DTG, appear more frequently in subtype B than in non-B subtypes [31]. Similarly, R263K is mainly present in subtype B, while G118R has a pathway in selecting DTG resistance in non-B subtype viruses [13,22,35,36].

HIV-1 sequences and structure-based analyses also showed that subtype-specific NOPs, especially at the active site of IN, can affect the genetic barrier to drug resistance by influencing the selection of resistance mutations, native protein structure, and the function of the drug-mediated inhibition of the enzyme [29,30,32,33,37].

In 2019, an estimated 669,236 people were living with HIV in Ethiopia, and the epidemic was dominated by subtype C [38,39]. Similar to many other countries in sub-Saharan Africa, Ethiopia has implemented the test-and-treat strategy, with DTG-based regimens recommended as the first-line antiretroviral therapy (ART) [40]. However, there is limited knowledge of the frequency and characteristics of NOPs of IN or their effect on the development of INSTI resistance. This study aimed to investigate HIV-1 IN genotypic profile to evaluate the prevalence of pre-treatment DRMs and NOPs that might affect the genetic barrier to the emergence of resistance in INTSI-naïve patients in Ethiopia infected with HIV-1 subtype C.

2. Materials and Methods

2.1. Study Design

In this study, we used plasma samples collected from HIV-1-infected patients as part of a national HIVDR survey conducted in Ethiopia. A cross-sectional survey was conducted in 2017 among treatment-naïve patients and patients on first- and second-line regimens in selected health facilities from different parts of the country according to the WHO-recommended HIVDR survey [41]. After obtaining written informed consent from each participant, 10 mL of blood was collected by venipuncture for CD4+ T-cell count, viral load, and HIVDR genotyping. Basic demographic and clinical information were also collected during the survey using a standardized questionnaire. Specimens were

transported to the Ethiopian Public Health Institute (EPHI) on dry ice for viral load testing and long-term storage at −80 °C. HIV-1 VL was determined using the Abbott RealTime HIV-1 assay (Abbott Molecular Inc., Des Plaines, IL, USA). Using 1000 copies/mL as a viral load suppression threshold based on the WHO recommendation [42], all samples with a viral load ≥1000 copies/mL were then shipped to the National Institute of Respiratory Diseases-Mexico (INER) laboratory for HIVDR genotyping.

2.2. HIV-1 Genotyping

Genotyping of the integrase region was performed using an in-house-developed and -validated protocol for IN [43]. Amplicons obtained by the nested PCR method were used for Sanger sequencing using the BigDye technology on the ABI Prism 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Sequence assembly and editing were performed using the RECall V 2.0 HIV-1 sequencing analysis tool (University of British Columbia, Vancouver, BC, Canada) [44]. Sequence quality control was performed using the WHO tool (https://sequenceqc-dev.bccfe.ca/who_qc (accessed on 28 June 2021)) and the Quality Control program of the Los Alamos HIV sequence database (https://www.hiv.lanl.gov (accessed on 28 June 2021)).

2.3. Subtype Determination Using HIV-1 Integrase Sequences

The HIV-1 subtyping was performed using the online automated subtyping tools REGA v3.0 [45], COMET [46], and the jumping profile Hidden Markov Model (jpHMM) [47]. Subtyping was further confirmed by Maximum likelihood (ML) phylogenetic tree analysis with the IN references sequences from HIV-1 subtype (A-K) and recombinant virus downloaded from the Los Alamos database (http://www.hiv.lanl.gov (accessed on 3 July 2021)). Multiple sequence alignment was conducted using MAFFT version 7 [48] and was then manually edited using BioEdit V7.0.9.0 [49,50] until a perfect codon alignment was obtained. ML tree topology was constructed using the online version of PhyML v 3.0 [51] with the GTR+I+Γ nucleotide-substitution model (using the estimated proportion of invariable sites and four gamma categories). A heuristic tree search was performed using the SPR branch-swapping algorithm. Branch support was determined with aLRT-SH (approximate likelihood ratio test, Shimodaira Hasegawa-like) [52]. Clusters were defined as monophyletic clades with aLRT-SH support ≥0.9. The subtype-resolved ML phylogeny trees were visualized using the FigTree v1.4.0 program. Sequence(s) that formed a cluster with the reference sequences belonging to the same subtype were assigned to that subtype.

2.4. HIV-1 Drug Resistance Analysis

INSTI-associated mutations were identified using the Stanford HIV Drug Resistance Database (HIVdB v9.0) (https://hivdb.stanford.edu/hivdb/by-mutations (accessed on 7 July 2021)). INSTI DRMs were categorized as major resistance mutations, accessory resistance mutations, and other mutations according to the Stanford HIV Drug Resistance Database. Major resistance mutations were primarily nonpolymorphic DRMs that caused a significant reduction in INSTI susceptibility, even when they occurred alone. Accessory mutations were nonpolymorphic or minimally polymorphic mutations that caused only low-level reduction of INSTI susceptibility when they occurred alone, but may have augmented resistance and/or restored the fitness of viral mutants with major resistance mutations. The other mutations included highly polymorphic and/or rare nonpolymorphic mutations that may have been weakly associated (uncertain role) with drug resistance. We further extensively investigated all amino acid positions associated with decreased INSTI susceptibility. Samples harboring resistant and/or a mixture of wild-type and resistant amino acids were considered resistant.

2.5. HIV-1 Subtype C Integrase Polymorphism and Conservation Analysis

For this analysis, only HIV-1 subtype C sequences were used. Briefly, multiple sequence alignment was conducted using MAFFT version 7 [48] and was then manually

edited using BioEdit V7.0.9.0 [49,50] until a perfect codon alignment was obtained. The nucleotide sequences were translated to an amino acid sequence. Then, each amino acid along the 288 IN positions was extensively investigated for the presence of primary mutations and of nonpolymorphic and polymorphic mutations associated with resistance to INSTI. The prevalence of each amino acid at each IN position was determined and compared to the HIV-1 subtype B reference sequence (GenBank accession number: K03455). We defined NOPs as substitutions within the HIV-1 IN that occurred in ≥1% of the sequences for this analysis [6]. The positions with ≥20% substitutions were defined as highly polymorphic, while those with ≤0.5% variability were considered highly conserved.

2.6. Generation of Consensus HIV-1 Integrase Sequence

To comprehensively describe the variability (polymorphism) in the IN sequences, we downloaded global subtype B and C IN sequences that matched the region (HXB2: 4230- 5093 relative to HXB2 clone) from the HIV Los Alamos National Library (LANL) database (https://www.hiv.lanl.gov (accessed on 13 July 2021)). To avoid the overestimation of variant calling and ensure the sequences included in the analysis were from INSTI-naïve patients, only sequences before 2007 (before the FDA approved INSTIs) were used. The quality of all HIV-1 sequences was verified using the online Quality Control program (http://www.hiv.lanl.gov (accessed on 14 July 2021)). Sequences with stop codons and/or frameshifts and/or poor quality were removed from the analysis. Only one sequence per patient was retained. For a patient with multiple sequences, the earliest sequence was selected and used. The consensus amino acid sequence for IN was generated for Ethiopian HIV-1 subtype C, the global HIV-1 subtype B, and the global subtype C sequence using BioEdit V7.0.9.0 [49,50]. For positions where two amino acids occurred at frequencies higher than 30%, both amino acids were represented, and the first letter seen at the consensus represented the most prevalent amino acid.

Furthermore, to assess the impact of previous exposure to ART on IN gene NOPs, the consensus amino acid sequences of IN from the ART-naïve and ART-experienced patients were generated and compared. Similarly, we also compared the consensus amino acid sequences of IN from patients with one or more major HIVDRMs to protease inhibitor (PI), NRTI, and/or NNRTIs (HIVDR group) with those with no major HIVDRMs (no-HIVDR groups) in their corresponding protease/reverse transcriptase (PR/RT) gene.

2.7. Genetic Barrier to Integrase Strand-Transfer Inhibitor Resistance

To assess differences in the genetic barrier for evolution of drug-resistance substitutions between subtypes C and subtype B, we compared Ethiopian HIV-1 subtype C IN sequences obtained from INSTI-naïve patients and global HIV-1 subtype B sequences obtained from LANL (INSTI-naïve, collected before 2007). We calculated the genetic barrier to INSTI resistance for 10 major INSTI resistance amino acid positions (19 substitutions) using a previously published method [53]. Briefly, we first determined the extent of natural diversity at each selected position in our dataset of Ethiopian HIV-1 subtype C IN sequences and global subtype B IN sequences by identifying all wild-type triplets and their prevalence. Next, we compute genetic barrier score for each wild-type triplet to evolve to resistant amino acid at the specific selected position. The genetic barrier was calculated as the sum of transitions and/or transversions required to evolve to any major drug-resistance substitution. We used a score of 1 for transition (A↔G and C↔T), 2.5 for transversion (A↔C, A↔T, G↔C, G↔T), and 0 when no change was needed, as described by Nguyen et al. (2012) [53]. The smallest number (minimal score) of transversion and/or transition required for evolution from wild-type codon to resistant codon were used to calculate the genetic barrier.

2.8. Modeling and In Silico Predictions of HIV-1 Integrase and Dolutegravir Interaction

For in silico predictions, 20 randomly selected (10 from each ART-naïve (PDR) and ART-experienced (ADR)) sequences were used. The ART-naïve IN sequences used in

our analysis had no HIVDRMs against NRTI, NNRTI, and/or PI in their corresponding PR/RT gene, while the ART-experienced group had one or more HIVDRMs against NRTI, NNRTI, and/or PI. A multiple-sequence alignment of amino acid sequences (without any gap) was made using ClustalW (https://www.genome.jp/tools-bin/clustalw (accessed on 1 November 2021)). An amino acid identity matrix was created with Clustal 12.1 (https://www.ebi.ac.uk/Tools/msa/clustalo (accessed on 1 November 2021)) and visualized using GraphPad Prism 8.

The crystallographic structure of full-length HIV-1 IN (accession number: 6u8q.pdb) was obtained from the Protein Data Bank (www.rcsb.org (accessed on 2 November 2021)) [54]. To visualize both the PDR and ADR HIV-1 IN, the 6u8q was modified by using UCSF-Chimera at 12 amino acid positions (see Table S2), and a monomer was used in the docking prediction. The structure (6u8q) originally included a DNA fragment and DTG. After removing all ligands, the DNA fragment and water molecules from the crystal structure, receptor, and ligand–DTG files were separately saved for further analysis. MGL Tools (Version 1.5.7rc1) was used for creating .pdbqt files of the receptor and ligands needed for docking with Autodock Vina (Vina) (Version 1.1.2) [55,56]. Ligands were docked to the binding site cavity using x = 211, 63 Å; y = 205, 453 Å; and z = 171, 895 Å Cartesian coordinates that used the catalytic site in the monomer of HIV-1 IN. The grid box dimensions used for the search space were 50 Å × 40 Å × 40 Å. Docking calculations were performed with an exhaustiveness option of 8 (average accuracy) and an energy range of 3. Validation of the docking method was performed by redocking DTG to the modified crystal structure to the modified above-mentioned structure.

2.9. Statistical Analysis

Fisher's exact test, the Chi-squared test, and the Mann–Whitney U-test were used to evaluate the statistical differences between groups. p-values ≤ 0.05 were considered statistically significant.

3. Results

A total of 460 IN sequences obtained from INSTI-naïve patients were included in the analysis. Among these, 373 sequences were from patients who did not report exposure to any antiretroviral drug at the time of specimen collection (ART-naive), while 87 sequences were from ART-experienced (NNRT-based or PI-based regimens) patients, with virological failure (viral load $\geq$ 1000 copies/mL) while on a first-line (n = 41) or second-line (n = 46) regimen.

3.1. HIV-1 Subtyping

Online subtyping and the subsequent phylogenetic analysis results showed that 98.5% (453/460) of the sequences were subtype C, while 0.43% (2/460), 0.22% (1/460), 0.22% (1/460), 0.22% (1/460), 0.22% (1/460), and 0.22% (1/460) were subtype B, subtype A1, CRF10_CD, CRF02_AG, CRF49_cpx and CRF_A2D, respectively (Figure 1).

The phylogenetic tree in Figure 1 contains a total of 874 sequences, including Ethiopian sequences (n = 460) and (n = 414) integrase reference sequences for HIV-1 subtypes (A–K) and circulating recombinant forms downloaded from the HIV-1 LANL database. An ML tree was constructed using the online version of PhyML v 3.0. The reference sequences from the Los Alamos National Laboratory are in black in the figure. All the Ethiopian sequence's clusters with the HIV-1 subtype C reference sequence are in green, while the non-subtype C Ethiopian sequences are in pink.

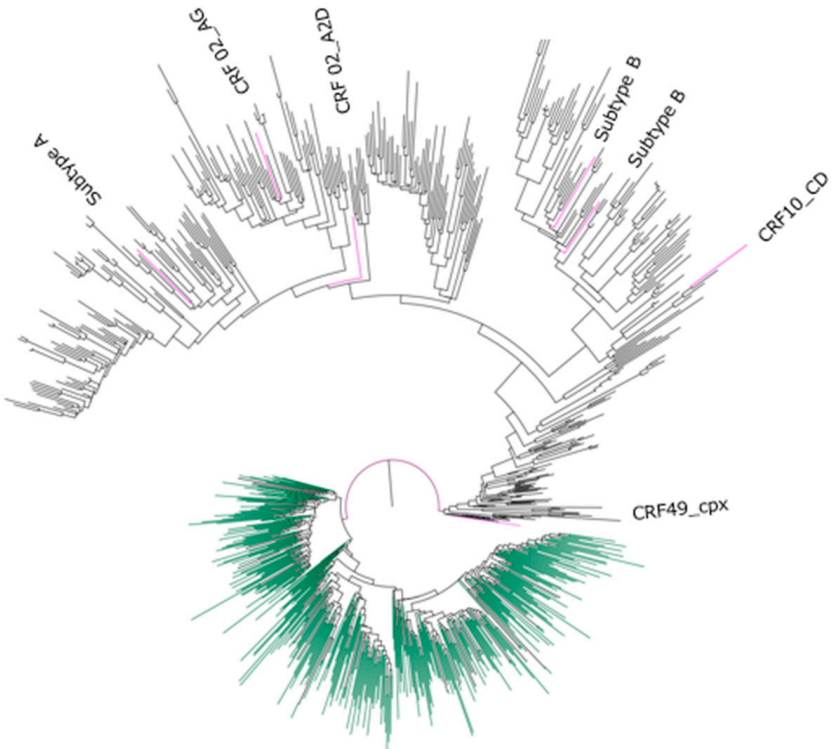


Figure 1. Maximum-likelihood phylogenetic tree of HIV-1 viral strains circulating in Ethiopia using integrase (IN) sequences.

3.2. Prevalence of Major Integrase Strand-Transfer Inhibitor Resistance Mutations

No major DRMs known to be associated with DTG resistance (T66K, E92Q, G118R, E138K/A/T, G140S/A/C, Q148H/R/K, N155H, or R263K) were detected among INSTI-naïve individuals, regardless of previous exposure to ART. However, one (0.22%) sequence from a person without previous ART exposure was found to harbor E92G, a mutation that moderately reduces EVG susceptibility but does not reduce susceptibility to RAL and DTG.

A total of 4.4% (20/460) of the sequences contained five different IN accessory mutations: −E157Q (2.39%), G163R/K (0.65%), Q95K (0.65%), T97A (0.43%), and G149A (0.22%). There was no significant difference in the prevalence of accessory mutations among ART-naïve and ART-experienced patients ($p = 0.9$) (Table 1). Only one accessory mutation per sequence was detected, except for one sequence with two (G149A and E157Q) accessory mutations. In addition, other mutations including M50I (18.5%, 85/460), L74I/M (2.8%, 13/460), S119R, (0.9%,4/460), V151I, (1.3%, 6/460), and D230N (0.4%, 2/460) were also detected.

Table 1. Prevalence of integrase accessory mutations detected and their ART status.

No.	Sequence ID	ART Regimen	Age	Gender	CD4+ T-Cell Count (Cells/mm^3)	Viral Load (Copies/mL)	INSTI Accessory Mutation
1	ETH-0186	Naive	40	M	359	62,118	E157Q
2	ETH-0232	Naive	42	M	83	–	G163R
3	ETH-0343	Naive	40	M	42	26,531	E157Q
4	ETH-0358	Naive	35	M	69	418,611	G149A, E157Q
5	ETH-0366	Naive	27	F	175	62,517	E157Q
6	ETH-0380	Naive	34	M	16	397,306	E157Q
7	ETH-0396	Naive	25	F	341	47,435	G163R
8	ETH-0410	Naive	38	F	538	11,458	E157Q
9	ETH-0493	Naive	45	M	164	7130	Q95K
10	ETH-0508	Naive	39	M	150	–	E157Q
11	ETH-0545	Naive	30	F	–	–	E157Q
12	ETH-0609	Naive	28	F	895	2002	T97A
13	ETH-0622	Naive	21	F	236	155,331	T97A
14	ETH-0631	Naive	35	F	50	295,532	E157Q
15	ETH-0695	Naive	35	F	–	6465	E157Q
16	ETH-0750	TDF+3TC+EFV	46	M	–	18,681	Q95K
17	ETH-0815	TDF+3TC+EFV	50	M	384	–	Q95K
18	ETH-0839	ABC+3TC+ATV/r	40	M	432	1432	G163K
19	ETH-0843	AZT+3TC+LPV/r	39	M	733	4752	G140E
20	ETH-0879	TDF+3TC+ATV/r	50	F	655	2667	E157Q

Abbreviations: age, in years; F, female, M, male; ART, antiretroviral therapy; INSTI, integrase strand-transfer inhibitor; 3TC, lamiduvine; TDF, tenofovir, AZT, zidovudine; EFV, efavirenz; LPV/r, lopinavir/ritonavir; ATV/r, atazanavir/ritonavir; Naïve, ART-naive; "–", missing data. CD4+ T in cells/mm^3; HIV RNA in copies/mL.

3.3. Integrase Strand-Transfer Inhibitor Resistance among Patents on Antiretroviral Therapy

To assess the impact of ART exposure to NRTI, NNRTI, and/or PI on the selection of INSTI-resistance mutations, we further compared the INSTI HIVDRMs from patients with one or more major HIVDR mutations to NRTI, NNRTI, and/or PI (HIVDR group) with those with no HIVDRMs in their corresponding PR/RT genes (no-HIVDR group) (Figure 2).

Briefly, among the total 460 IN sequences used in our analysis, 327 had a corresponding PR/RT gene sequence, of which 234 had no major HIVDRMs (no-HIVDR group), while 93 of the sequences (HIVDR group) had one or more HIVDRMs against the NRTI, NNRTI, and/or PI (see Table S1). No major INSTI HIVDRMs were detected in either of these groups, and there was no significant difference in the presence of accessory mutations with regard to previous ART exposure, nor with regard to DRMs toward other ARVs. Among the HIVDR and no-HIVDR groups, 3.2% (3/93) and 4.7% (11/234) accessory mutations were detected, respectively (p = 0.8); while 4.29% (15/373) and 5.75% (5/87) accessory mutations were detected among ART-naïve and ART-experienced groups, respectively (p = 0.6).

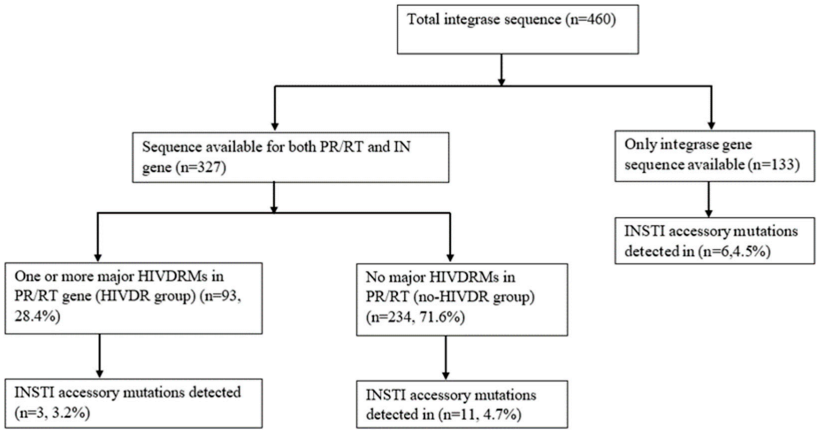


Figure 2. Flow chart of study-participant selection and INSTI drug resistance among INSTI-naïve patients (*n* = 460). Abbreviations: PR/RT, protease and reverse transcriptase gene; IN, integrase; INSTI, integrase strand-transfer inhibitors; HIVDRM, HIV-drug-resistance mutations.

High similarity was also observed when comparing the consensus sequence from ART-naïve and ART-experienced patients, as shown in Figure 3. Similarly, our comparison of the consensus sequences from the HIVDR and non-HIVDR groups also showed high similarity between the two consensus sequences, except at positions K215N, T218L, and R269, where the HIVDR group had one amino acid; while the no-HIVDR group had a mixture of amino acids at positions T215K/N, T218I/L, and R269R/K, respectively (Figure 3).

```
                    10        20        30        40        50        60        70        80        90        100
            ....|....|....|....|... .  | ....|....|....|....|....|....|....|....|....|....|....|....|....|.... |
HXB2_Ref    FLDGIDKAQDEHEKYHSNWRAMA S  D FNLPPVVAKEIVASCDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAY F
ART_Expo    .........E.............N/SE/D.....I.......................................I............................Y/F
ART_Naive   .........E.............N/SE/D.…I.....................................I...........................Y/F
HIVDR       .........E.............N/SE/D.....I.......................................I............................Y/F
No_HIVDR    .........E.............N/SE/D.....I.......................................I............................Y/F

                    110       120       130       140       150       160       170       180       190       200
            ....|....|....|....|... . |....|....| . ...|....|....|....|....|....|. . ..|....|....|....|....|....|
HXB2_Ref    LLKLAGRWPVKTIHTDNGSNFTG A TVRAACWWAGI K QEFGIPYNPQSQGVVESMNKELKKIIGQVR D QAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERI
ART_Expo    I..........V..........ST/AA.K........K/Q..............................E/D.................................
ART_Naive   I..........V..........ST/AA.K........K/Q..............................E/D.................................
HIVDR       I..........V..........ST/AA.K........K/Q.............................. D .................................
No_HIVDR    I..........V..........ST/AA.K........K/Q.............................. D .................................

                    210       220       230       240       250       260       270       280       288
            ....|....|.... | .. . .|....|....|....|....|....|....|....|....|.... | ... . |....|....|....|...
HXB2_AA     VDIIATDIQTKELQ K QI T KIQNFRVYYRDSRNPLWKGPAKLLWKGEGAVVIQDNSDIKVVPRRK A KII R DYGKQMAGDDCVASRQDED
ART_Expo    I.............K/N.. L .............D.I.............................   ...   ........A....G.....
ART_Naive   I.............K/N..I/L.............D.I.............................A/V...K/R........A....G.....
HIVDR       I............. N .. L .............D.I.............................. A ... R ........A....G.....
No_HIVDR    I.............K/N..I/L.............D.I.............................. A ...K/R........A....G.....
```

Figure 3. Alignment of Ethiopian HIV-1 subtype C integrase (IN) consensus sequence. The consensus sequence from the ART-naïve sequences (*n* = 367) is represented as ART_Naive, and that from ART-experienced (*n* = 87) is represented as ART_Expo. The consensus sequence from the sequence with no HIVDR mutation in the protease/reverse transcriptase (PR/RT) gene (*n* = 234) is represented as No_HIVDR, while that with one or more major mutation in PR/RT is represented as HIVDR. Positions with more than one amino acid are both represented. HXB2 represents the consensus HIV-1 subtype B reference sequence from the LANL database (accession number: K03455).

3.4. Prevalence of Naturally Occurring Integrase Polymorphisms in HIV-1 Subtype C

An alignment of the 453 HIV-1 subtype C IN sequences from the INSTI-naïve Ethiopian patients was extensively analyzed and compared to the HIV-1 subtype B reference sequence (GenBank accession number: K03455). Based on our definition of polymorphism (≥1.0% variability), an overall 64.9% (187/288) amino acid positions of the IN were conserved. The conservations of the NTD, CCD, and CTD were 60% (30/60), 66.1% (107/162), and 66.8% (50/76), respectively. The distribution of polymorphisms in 453 HIV-1 subtype C IN sequences is shown in Figure 4.

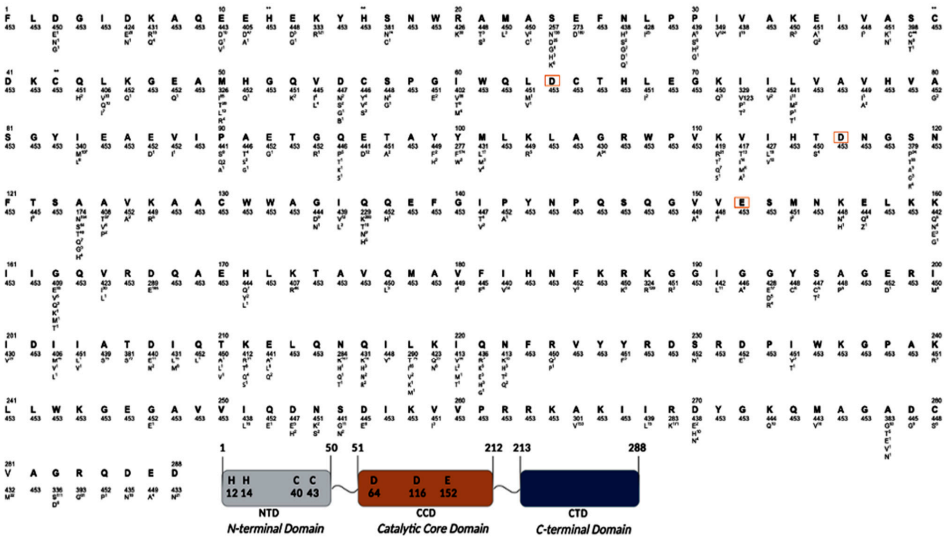


Figure 4. Distribution of variants among HIV-1 subtype C integrase sequences. The number at the top is the amino acid position in the integrase gene (1–288), and the consensus subtype B sequence is indicated below the number. Beneath the consensus, the number indicates the number of sequences containing the amino acid at the indicated position. The variant amino acid at each position is indicated along with number of sequences with that amino acid (superscript). The HHCC zinc-binding motifs are indicated by **, the amino acid of the DDE active sites are indicated by red boxes. The N-terminal domain is indicated in gray at positions 1–50, while the catalytic core domain (CCD) at positions 51–212 is indicated in orange, and the C-terminal domain (CTD) at positions 213–288 is indicated in blue.

3.5. Analysis of the N-Terminal Domain (NTD)

Within the NTD, the Zn-binding motif ($H_{12}H_{16}C_{40}C_{43}$) involved in the multimerization of the IN subunit, stabilization of folding, and interaction with LEDGF/p75 were highly conserved [6]. However, amino acid positions, D10E, S24N, D25E, V31I, and M50I were highly polymorphic (>20.0% variability). We also observed that the residue E10 had been replaced by D (aspartic acid) in 97.8% of sequences, which might be the signature of subtype C (Figure 4).

3.6. Analysis of the Catalytic Core Domain (CCD)

In the CCD, the catalytic triad $D_{64}D_{116}E_{152}$ was highly conserved, and was found within the conserved regions 61–70, 114–118, and 152–155, respectively. The critical positions for DNA-binding HIV-1 integration and replication (Q62, H67, N120, N144, Q148, and N155) [57] and the residue involved in the chemical bond and hydrophobic contact

with the LEDGF/p75 [6] (A128-A129-W131-W132-Q168-E170-T174-M178) were also highly conserved. However, amino acids at codon positions G163, V165, D167, H171, and K173 within the I161-K173 region known to be involved in the noncanonical nuclear localization signal [6,22], and the K188 within the KRK motif (K_{186}, R_{187}, K_{188}), which is vital for the integrase:integrase:oligomerization at the dimer:dimer interface [6,22], showed 28.5% variability.

Among the INSTI-mutation positions in the CCD residues that directly reduced the INTSI susceptibility, H51, T66, E92, F121, G140, Y143, Q146, S147, Q148, S153, N155, and E157Q were highly conserved, except for codon position E157Q, which was a polymorphic position (>1.0% variability). However, a highly polymorphic residue in the CCD including V72I, I84M, F100Y, L101I, T112V, T124A, T125A, R127K, K136Q, D167E, K188R, and V201I was observed.

3.7. Analysis of the C-Terminal Domain (CTD)

Within CTD, the two large consecutive residues, L241-Q252 and I257-K264, which are involved in the binding of viral and cellular DNA, were found to be highly conserved, except for positions I251 and V257, which were mutated to I251L and V259I in 3.5% and 0.7% of the sequences, respectively. However, the important positions for DNA binding and integrase multimerization (K258, V260, R262, R263, and K264) [6] were fully conserved.

Our analysis also showed that 24 amino acid positions were highly polymorphic (>20.0% variability): D10E, K14R, S24N, D25E, V31I, M50I, V72I, I84M, F100Y, L101I, T112V, T124A, T125A, R127K, K136Q, D167E, K188R, V201I, K215N, T218I, A265V, R269K, D278A, and S283G. Six of these (D10E, K14R, S24N, D25E, V31I, and M50I) belonged to the NTD, whereas 12 (V72I, I84M, F100Y, L101I, T112V, T124A, T125A, R127K, K136Q, D167E, K188R, and V201I) belonged to the CCD, and the other 6 (K215N, T218I, A265V, R269K, D278A, and S283G) belonged to the CTD.

Our comparison of the NOPs' distribution with the global subtype B and global subtype C sequences downloaded from LANL showed that the Ethiopian HIV-1 subtype C IN sequences had a high similarity to the global subtype C sequence, but were quite different from the global subtype B, as shown in Figure 5.

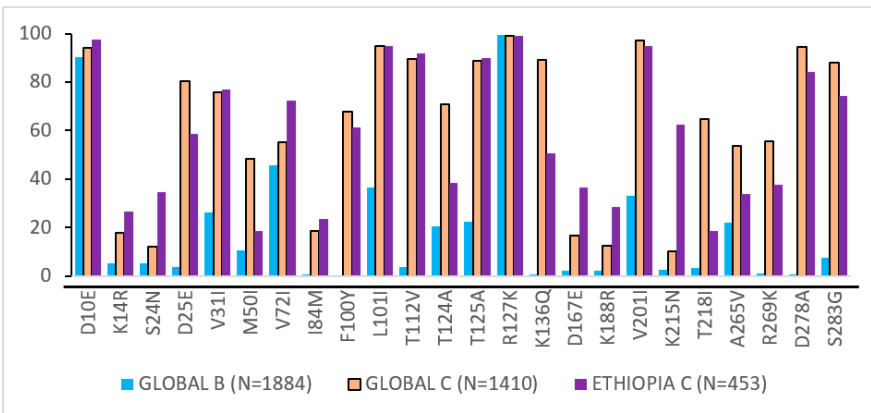


Figure 5. Prevalence of integrase highly polymorphic mutation in INSTI-naïve patients. The comparison was done using sequences downloaded from the Los Alamos National Library (LANL) database for subtype B (*n* = 1884), HIV-1 subtype C (*n* = 1410), and Ethiopian subtype C (*n* = 453).

3.8. Analysis of the Subtype Consensus Integrase Sequences

The consensus IN sequence for the global HIV-1 subtype B and global subtype C were generated using 1884 and 1410 sequences, respectively. Our comparison of the 288 amino acid sequence alignment of the consensus Ethiopian HIV-1 subtype C with the global HIV-1 subtype C showed high similarity, except for the mixture of amino acid sequences at positions 25E/D, 100Y/F,124T/A, 136K/Q, 167E/D, 215K/N, and 218I/L in the Ethiopian consensus; and 50M/I, 72I/V, and 265A/V in the global HIV-1 subtype C consensus sequence. However, it differed from the global subtype B consensus at eight positions with complete amino acid replacement (31, 112, 125, 201, 218, 234, 278, 283), while a mixture of amino acids was detected at positions 11E/D, 72I/V, and 101I/L in the global subtype B consensus sequences and at 24N/S, 25E/D, 100Y/F, 124T/A, 136K/Q, 167E/D, 215K/N, and 269K/R in the Ethiopian subtype C consensus sequence (Figure 6).

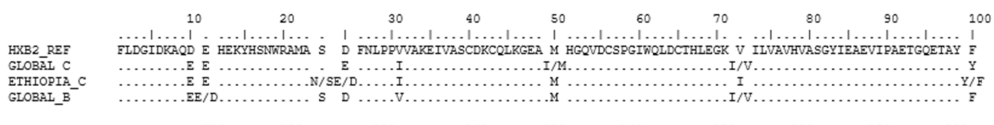


Figure 6. Alignment of integrase (IN) consensus sequence from LANL database (https://www.hiv.lanl.gov (accessed on 25 July 2021)). The consensus sequence from the global HIV-1 subtype C (n = 1410 sequences) is represented as GLOBAL C, the global HIV-1 subtype B (n = 1884 sequences) is represented as GLOBAL_B, and the Ethiopian subtype C sequence (n = 453 sequences) is represented as ETHIOPIA_C. Positions with more than one amino acid are both represented. HXB2 represents the consensus HIV-1 subtype B reference sequence from the LANL database (accession number: K03455).

3.9. Genetic Barrier to Dolutegravir Resistance

In this study,19 substitutions conferring major resistance to DTG at 10 amino acid positions in the IN (T66A/I/K, E92G, G118R, E138K/A/T, G140S/A/C, Y143R/C/H, S147G, Q148H/R/K, N155H, and R263K) were assessed to explore the genetic barrier to DTG. For each codon, the number of transitions and/or transversions required for a IN drug resistance associated substitution were calculated. A total of 1884 global HIV-1 subtype B sequences and 453 Ethiopian subtype C sequences from INSTI-naïve patients were compared for differences in the genetic barrier to INSTI resistance (Table 2).

Overall, the sequence analysis of the two subtypes showed similar predominant codon use at the selected amino acid positions, resulting in a similar minimum score for the genetic barrier to DTG. However, at position 140, the predominant codons in subtype C were GGG (53.6%) and GGA (45.9%). In contrast, in subtype B, GGC (85.0%) was the predominant codon resulting in a difference in the calculated genetic barrier at this position. For subtype C, two transversions (minimum score of 5) were required to mutate to G140C (GGG/A to ATG/C); while for subtype B, one transversion and transition (minimum score: 3.5) were required to mutate to G140C (GGC to TGT). Similarly, a two-point mutation (one transversion and one transition) (minimum score of 3.5) was required to mutate to G140S (GGG/A to AGT/C) for subtype C; while subtype B required a one-step transition (minimum score of 1) (GGC to AGC).

Table 2. Analysis of genetic barrier based on the minimum number of transitions and transversions required to obtain mutation resistance to DTG.

Codon Position	Substitution	Subtype C, n (%) [a]	Subtype B, n (%) [b]	Wild-Type Codon	Mutant Codon	Minimal Score [c]
66	T66A	439 (96.91)	1829 (97.08)	ACA	GTC, GCC/A/G	1
		3 (0.66)	7 (0.37)	ACG		1
		3 (0.66)	6 (0.32)	ACT		2
		8 (1.77)	42 (2.23)	ACC		1
	T66K	439 (96.91)	1829 (97.08)	ACA	AAA/G	2.5
		3 (0.66)	7 (0.37)	ACG		2.5
		3 (0.66)	6 (0.32)	ACT		5
		8 (1.77)	42 (2.23)	ACC		5
	T66I	439 (96.91)	1829 (97.08)	ACA	ATT/C/A	1
		3 (0.66)	7 (0.37)	ACG		3.5
		3 (0.66)	6 (0.32)	ACT		1
		8 (1.77)	42 (2.23)	ACC		1
92	E92Q	441 (97.35)	446 (23.67)	GAA	CAA/G	2.5
		12 (2,65%)	1438 (76.33)	GAG		2.5
118	G118R	394 (86.98)	1750 (92.89)	GGC	CGT/C/A/G, AGA/G	2.5
		19 (4.19)	32 (1.7)	GGA		1
		1 (0.22)	5 (0.27)	GGG		1
		39 (8.61)	91 (4.83)	GGT		2.5
138	E138A	440 (93.16)	1831 (97.19)	GAA	GTC, GCC/A/G	2.5
		13 (2.87)	39 (2.07)	GAG		2.5
	E138K	440 (93.16)	1831 (97.19)	GAA	GTC, GCC/A/G	1
		13 (2.87)	39 (2.07)	GAG		1
	E138T	440 (93.16)	1831 (97.19)	GAA	ACT/C/A/G	3.5
		13 (2.87)	39 (2.07)	GAG		3.5
140	G140A	243 (53.64)	18 (0.96)	GGG	GTC, GCC/A/G	2.5
		208 (45.92)	58 (3.08)	GGA		2.5
		1 (0.22)	201 (10.67)	GGT		3.5
		1 (0.22)	1607 (85.30)	GGC		2.5
	G140S	243 (53.64)	18 (0.96)	GGG	TCT/C/A/G, AGT/C	3.5
		208 (45.92)	58 (3.08)	GGA		3.5
		1 (0.22)	201 (10.67)	GGT		1
		1 (0.22)	1607 (85.30)	GGC		1
	G140C	243 (53.64)	18 (0.96)	GGG	TGT, TTC	5
		208 (45.92)	58 (3.08)	GGA		5
		1 (0.22)	201 (10.67)	GGT		2.5
		1 (0.22)	1607 (85.30)	GGC		3.5

Table 2. *Cont.*

Codon Position	Substitution	Subtype C, n (%) [a]	Subtype B, n (%) [b]	Wild-Type Codon	Mutant Codon	Minimal Score [c]
143	Y143C	436 (96.25)	1877 (99.63)	TAC	TGT, TTC	2.5
		17 (3.75)	7 (0.37)	TAT		1
	Y143H	436 (96.25)	1877 (99.63)	TAC	CAT/C	1
		17 (3.75)	7 (0.37)	TAT		1
	Y143R	436 (96.25)	1877 (99.63)	TAC	CGT/C/A/G. AGA/G	3.5
		17 (3.75)	7 (0.37)	TAT		2
147	S147G	403 (88.96)	1828 (97.03)	AGT	GGT/C/A/G	1
		50 (11.04)	56 (2.97)	AGC		1
148	Q148H	71 (15.67)	1828 (97.03)	CAA	CAT/C	2.5
		382 (84.33)	56 (2.97)	CAG		2.5
	Q148K	71 (15.67)	1828 (97.03)	CAA	AAA/G	2.5
		382 (84.33)	56 (2.97)	CAG		2.5
	Q148R	71 (15.67)	1828 (97.03)	CAA	CGT/C/A/G, AGA/G	1
		382 (84.33)	56 (2.97)	CAG		1
155	N155H	427 (94.26)	1849 (98.14)	AAT	CAT/C	2.5
		26 (5.74)	35 (1.86)	AAC		2.5
263	R263K	62 (13.69)	1833 (97.29)	AGA	AAA/G	1
		389 (85.87)	4 (0.21)	AGG		1

[a] Subtype C: Ethiopian sequence used in the analysis (n = 453). [b] Subtype B: global subtype B sequence deposited before 2007 (before INSTI was used) retrieved from the Los Alamos database (n = 1884). [c] Minimal score calculated by the sum of number of transversions and transitions for each, with transitions scored as 1 and transversions scored as 2.5.

3.10. Impact of Protease and Reverse-Transcriptase Drug-Resistance Mutation on the Structure of HIV-1 Integrase

The effects of HIVDRMs in HIV-1 PR and/or RT on the secondary structure of HIV-1 IN were investigated on 20 sequences: 10 from ART-naïve (PDR) and 10 from ART-experienced (ADR) individuals representative of randomly selected HIV-1 IN sequences. The sequence identity matrix (Figure 7a) showed that all the sequences were more than 92% identical at the amino acid level, and there were no major differences between the two main groups. To study the effects of PR and RT drug-induced resistance on the structure of HIV-1 IN, chain A of the 6u8q structure was modified at 12 positions to represent both the ADR and the PDR sequences (see Table S2). The alignment of the monomers of the PDR and ADR INs did not result in any differences between the two groups. DTG was successfully docked to both the PDR and ADR IN by Autodock Vina (Figure 7c), and the docking score was −6.5 kcal/mol, which was at a similar position as the original DTG ligand.

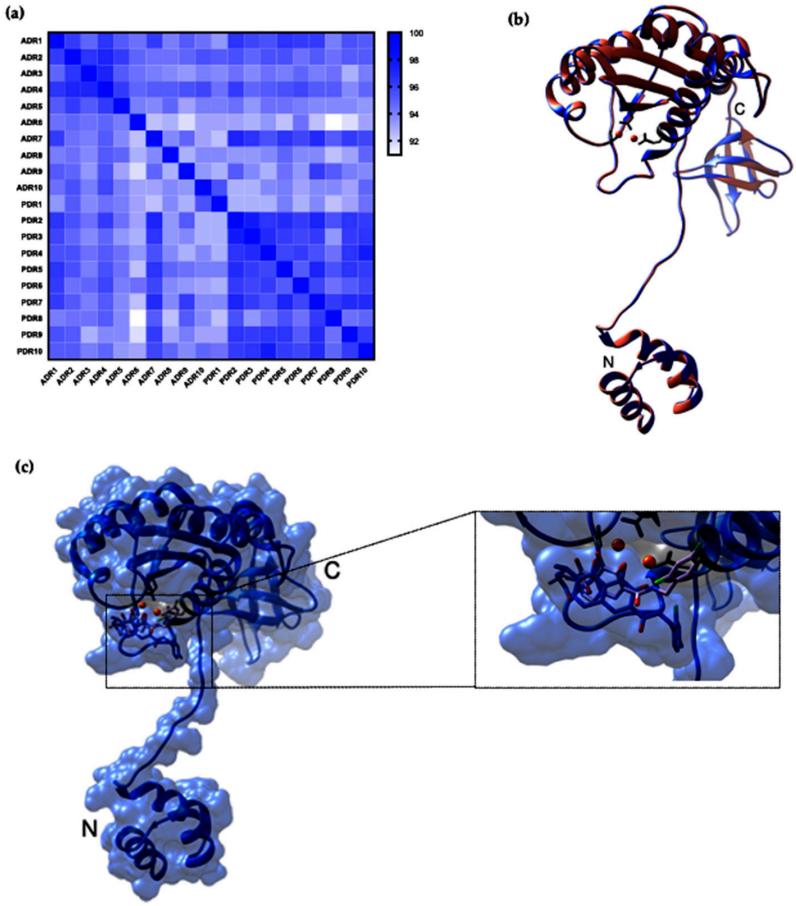


Figure 7. Visualization of HIV-1 IN amino acid sequences from ART-naive (PDR) and ART-experienced drug-resistant (ADR) individuals. (**a**) Heat map of amino acid identities of HIV-1 IN amino acid sequences from PDR and ADR individuals. The heat map was generated using a percent identity matrix table created in Clustal 12.1, and the heat map was visualized in GraphPad Prism 8. (**b**) For in silico predictions, a molecular model of the monomeric HIV-1 integrase structure 6u8q was used and modified based on the multiple-sequence alignment of amino acid sequences of 10 ADR and 10 PDR sequences. Figures were created in Chimera (https://www.cgl.ucsf.edu/chimera https://www.cgl.ucsf.edu/chimera (accessed on 2 November 2021)). PDR (salmon) and ADR (cornflower blue) HIV-1 monomers are represented as ribbons, the catalytic triad is represented with sticks in light grey, and magnesium is represented in orange-red. The PDR structure was moved at 0.01 Å on the x-axis. (**c**) Surface views of the structure and the validation of docking were conducted using Autodock Vina. Dolutegravir is represented with sticks; blue shows the original coordinates, and purple shows the docking mode of dolutegravir in Autodock Vina.

4. Discussion

Overall, our results revealed no major DTG associated HIVDRM mutations among INSTI-naïve individuals, regardless of previous exposure to ART. In one individual, the E92EG mutation was found, which moderately reduced EVG susceptibility, but had no effect on DTG. However, INSTI accessory mutations and NOPs, which could influence INSTI susceptibility and the genetic barrier to INSTI resistance, were detected. Our polymorphism analysis showed that 64.9% (187/288) of amino acid positions of the HIV-1 subtype C IN sequences from INSTI-naïve individuals were conserved (<1.0% variability). The majority of amino acids involved in key functions of the enzyme (the HHCC motif and the DDE motifs [6,22]) were fully conserved. The genetic barriers to DTG resistance were similar at selected amino acid positions for subtypes B and C, except that subtype C had a higher genetic barrier for the G140C and G140S mutations, highlighting that the Q148H/K/R DTG resistance pathway was selected less in subtype C. Docking analysis of the DTG showed that the PR- and RT-associated HIVDRM did not affect the structure of the HIV-1 IN, supporting the use of DTG as a salvage therapy for patients with resistance to drugs targeting these enzymes.

The absence of major INSTI DRMs among INSTI-naïve patients in our study was consistent with other studies from Africa [58–64], Asia [65–67], and Europe [68–70], showing no or highly infrequent major INSTI mutations among INSTI-naïve patients. Our finding was not unexpected, and was in line with studies from other settings based on samples obtained before the rollout of DTG [71–73]. However, following the wide scale-up of DTG, an increase in DTG resistance has been reported, especially in persons receiving DTG monotherapy [15,19,23–28]. Hitherto, the prevalence of transmitted resistance to DTG resistance has been low [20–23]. Similarly, in Ethiopia, after implementing the test-and-treat strategy, an increased number of patients will be on a DTG-based regimen. Thus, the emergence of INSTI resistance is expected, especially in settings with low access to viral load monitoring, delaying the identification of patients with treatment failure and increasing the risk of HIV drug resistance [74].

When present alone, accessory mutations have a minimal effect on INSTI susceptibility, but may serve to augment resistance and/or restore the fitness of viral mutants with major resistance mutations [5,30]. INSTI accessory mutations were detected in 20 (4.4%) of our specimens, and were equally distributed in both ART-naive and ART-experienced patients. Similar to our findings, different studies [67,72,73,75] revealed that NOPs were common among INSTI-naive patients. However, the prevalence differed with HIV-1 subtypes or circulating recombinant forms.

E157Q was the most common nonpolymorphic accessory mutation detected in our analysis. It is a natural polymorphism present in 1–10% of untreated individuals, depending on the subtype. It has no effect on the susceptibility of INSTI. However, it may act as a compensatory substitution for R263K-induced resistance to DTG [76]. Q95K was among the other nonpolymorphic accessory INSTI resistance detected in our study, and it had little, if any, effect on drug susceptibility to INSTI; however, in the presence of a N155H mutation, it increased INSTI resistance and improved the impaired replication of the virus [77].

L74M/I (2.9%) and M50I (18.8%) were the other polymorphic mutations detected in our study. L74M/I has been reported at levels between 0.5–20% in the untreated population, with a high prevalence in subtypes A, G, and A/G recombinants. It does not decrease INSTI susceptibility alone, but it can contribute to a high-level resistance when occurring with major INSTI-resistance mutations, mainly the Q148H/K/R mutation [24,58,78,79]. Studies in South Africa, Brazil and Europe have also confirmed a low frequency of L74M in INSTI-naïve patients [64,68,80]. M50I can be found in 10–25% of INSTI-naïve patients [81]. M50I alone does not negatively impact integrase strand-transfer activity and HIV replication capacity, but in combination with R263K, it increased resistance to DTG by 15.6-fold [81].

The other nonpolymorphic and polymorphic accessory mutations detected were G163R and T97A, which can contribute to a high-level resistance when occurring with Y143 and N155H major INSTI-resistance mutations [30].

In this study, we characterized the distribution of amino acid variants among the 453 HIV-1 subtype C IN sequences from INSTI-naïve individuals. Our results revealed that 64.9% of HIV-1 IN amino acid positions were conserved (<1.0% variability). The conserved position in the NTD, CCD, and CTD were 60%, 66.0%, and 65.8%, respectively. This was comparable to the study by Rhee et al. (2008) that showed 70% (202/288) of IN amino acid positions of the 1500 sequences obtained from INSTI-naïve (ART-naive or ART-experienced) individuals with different subtypes (<1.0% variability) [5]. Similarly, Hackett et al. (2008) also showed that 65% (187/288) of amino acid positions were conserved after analyzing 1304 HIV-1 sequences from groups M, N, and O IN sequences [82].

In general, our results showed that the majority of amino acids involved in key functions of the enzyme, including the zinc-binding HHCC motif, the multimerization of IN subunits, and the binding with the human cellular factor LEDGF/p75 in the catalytic core domain, the catalytic triad DDE [6,22] was highly conserved. The high conservation might have been due to the absence of INSTI pressure. All of our study participants were INSTI-naïve, and INSTI was not used in Ethiopia during our sample collection. However, a highly polymorphic residue in the NTD, CCD, and CTD regions, which might have affected the IN-protein function and interfered with the INSTI binding, were also observed [22,30]. Further long-term treatment follow-up studies are needed to assess the potential impact of NOPs on the evolution of INSTI resistance and viral fitness under the pressure of INSTIs.

It was also interesting to note that 20.5% (93/453) of our study participants were found to harbor a major HIVDR mutation (transmitted and acquired HIVDR) for NRTI, NNRTI, and/or PI in their corresponding PR/RT gene. However, DRM directed toward sites other than IN did not have a significant effect on INSTI susceptibility. In line with our findings, different studies have shown that previous NRTIs mutations appeared to have no impact on the risk of virological failure in patients switched to DTG with NNRTIs [83–86]. However, this was in contrast to other studies that showed previous exposure to NNRTI, PI, and/or NNRTI induced mutations or increase polymorphisms in the IN gene, highlighting the functional cooperation between viral IN and RT, and/or a potential coevolution of some of their mutations [9,87]. For instance, a study by Ceccherini et al. (2009, 2010) showed a higher frequency of I84V, M154I, and V165I among ART-treated subtype B patients compared to ART-naïve patients, implying that nonsuppressive ART treatment based on other antiretroviral drug classes (NRTI and/or NNRTI) might induce IN polymorphisms [6,9].

However, in our study, no significant difference was found in I84V and M154I prevalence between the ART-naïve and ART-experienced patients (22.6% and 0.5% of I84V and M154I among ART-naïve, and 1.15% and 12.6% among ART-experienced patients, respectively ($p = 0.5$ and $p = 0.4$)), while an increased prevalence of V165I was observed among ART-experienced groups (5.45% of V165I and12.64% between the ART-naïve and ART-experienced groups, respectively ($p = 0.03$)). Furthermore, our comparison of the HIVDR and no-HIVDR groups showed no differences (17.4%, 0.4%, and 7.3% of I84V, M154I, and V165I for the no-HIVDR group; and 22.6%, 1.1%, and 6.5% for the HIVDR group; $p = 0.4$, $p = 0.5$, and $p = 1$, respectively).

The observed differences between this and previous studies might be due to the number of sequences, range of major/minor mutations, and subtypes included in the analysis. However, the lack of a major INSTI mutation among sequences with multiple mutations in the PR/RT gene and the high conservation of amino acids involved in key functions of the IN enzyme did not support the impact of previous ART treatment on INSTI susceptibility.

Our docking analysis further supported our results, and showed no differences between the HIVDR and no-HIVDR groups. In both groups, DTG was successfully docked at a similar position to the original DTG ligand with the best docking score of −6.5 kcal/mol.

The genetic barrier, which is a crucial factor in the development of drug resistance, is defined by a cumulative number of resistance-associated mutations (RAMs) required for the virus to escape drug-selective pressure [53]. It is an important factor that contributes to the development of drug resistance. The variability at the nucleotide level in the IN among the different subtypes could influence the genetic barrier of INSTI drugs. In this study, we explored how the variability between subtypes C and B could affect DTG resistance.

Overall, our analysis of the codon distribution of the selected amino acid position of HIV-1 subtype C and subtype B revealed a similar genetic barrier for the development of DTG resistance between subtype C and B, except at codon position 140, where subtype C had a higher genetic barrier to develop the G140C and G140S mutations compared to subtype B, highlighting a higher genetic barrier for the Q148H/R/K resistance pathway in subtype C. The G140S mutation has been shown to rescue the catalytic defect due to the Q148H mutation, enabling the recovery of viral fitness [88]. A similar high genetic barrier to acquire mutations G140S or G140C has also been described in CRF02_AG compared with subtype B [53,89].

This study was comprehensive, and included both treatment-naïve and treatment-experienced (first- and second-line regimens) patients, and will be a benchmark for INSTI DRM monitoring in Ethiopia. However, our analysis was based on the Sanger dideoxy sequencing method, which does not detect drug-resistance minority variants below 20% of the virus population, and might have underestimated the prevalence of INSTI DRMs among our study participants [90].

5. Conclusions

Our results showed no major clinically relevant INSTI-associated mutations among INSTI-naïve patients regardless of exposure to other antiretroviral agents, supporting the implementation of the wide scale-up of DTG-based regimes in Ethiopia. However, the detection of polymorphisms contributing to INSTI resistance and the expected increased use of DTG-based regimens in Ethiopia warrant the need for continuous surveillance of INSTI resistance. The genetic barrier analysis showed that subtype C had a high genetic barrier to acquiring the G140C and G140S mutations, highlighting that the Q148H/K/R mutation DTG resistance pathway was selected less in subtype C. Moreover, the docking analysis of the dolutegravir showed that protease- and reverse-transcriptase-associated HIVDRMs did not affect the native structure of the HIV-1 integrase.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/v14040729/s1, Table S1: Type of HIV-drug-resistance mutations detected in the HIVDR group (patients with one or more major HIVDR mutation to NRTI, NNRTI, and/or PI) (n = 93), Table S2: Modifications of the 6u8q.pdb HIV-1 integrase structure according to the alignment of both the ADR and PDR sequences.

Author Contributions: Conceptualization, D.A.A., P.B. and P.M.; Data curation, D.A.A.; Formal analysis, D.A.A., S.S. and P.M.; Funding acquisition, F.A, P.B. and P.M.; Investigation, D.A.A., Z.I.S., K.Z., A.G., F.A., S.S. and P.M.; Project administration, D.A.A. and P.M. Supervision, D.A.A. and P.M.; Visualization, Z.I.S.; Writing—original draft, D.A.A.; Writing—review and editing, D.A.A., S.S., P.B. and P.M. All authors revised the manuscript, provided important intellectual content. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the WHO Ethiopia Country Office. The Swedish Research Council (grant numbers 2019-05235 and 2020-02344), the Österlund Foundation, and a donation to P.M. and P.B. through the Medical Faculty, Lund University, also supported the research. The computations were enabled by resources from project SNIC 2021-5-69 provided by the Swedish

National Infrastructure for Computing (SNIC) at UPPMAX, partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

Informed Consent Statement: The study was conducted in accordance with the Declaration of Hesinki and approved by Research and Ethical Clearance Committee of EPHI and the National Health Research Ethics Review Committee of the Ministry of Science and Technology of Ethiopia (SERO-60-2016; Version 001, on 11 May 2016). Written informed consent was obtained from all participants for the national HIV-drug-resistance survey.

Data Availability Statement: All the sequences from this study were deposited in the GenBank with accession numbers OM302554–OM303013.

Acknowledgments: We express our deep appreciation to the study participants and all staff involved in the survey. We acknowledge the support given by the WHO and the National Institute of Respiratory Diseases-Mexico (INER) laboratory for HIVDR genotyping. We also acknowledge Mauno Vihinen (Protein Bioinformatics. Lund University) for his assistance with the Chimera software.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. WHO. *Guidelines on the Public Health Response to Pretreatment HIV Drug Resistance*; WHO: Geneva, Switzerland, 2017. Available online: http://apps.who.int/iris/bitstream/10665/255880/1/9789241550055-eng.pdf (accessed on 20 December 2021).
2. WHO. Policy Brief: Update of Recommendations on First- and Second-Line Antiretroviral Regimens. 2019. Available online: https://apps.who.int/iris/handle/10665/325892 (accessed on 20 December 2021).
3. WHO. Dolutegravir (DTG) and the fixed Dose Combination (FDC) of Tenofovir/Lamivudine/Dolutegravir (TLD). 2018. Available online: https://www.who.int/hiv/pub/arv/DTG-TLD-arv_briefing_2018.pdf (accessed on 20 December 2021).
4. The Lancet HIV. End resistance to dolutegravir roll-out. *Lancet HIV* **2020**, *7*, e593. [CrossRef]
5. Rhee, S.Y.; Liu, T.F.; Kiuchi, M.; Zioni, R.; Gifford, R.J.; Holmes, S.P.; Shafer, R.W. Natural variation of HIV-1 group M integrase: Implications for a new class of antiretroviral inhibitors. *Retrovirology* **2008**, *5*, 74. [CrossRef]
6. Ceccherini-Silberstein, F.; Malet, I.; D'Arrigo, R.; Antinori, A.; Marcelin, A.G.; Perno, C.F. Characterization and structural analysis of HIV-1 integrase conservation. *AIDS Rev.* **2009**, *11*, 17–29.
7. Craigie, R. HIV integrase, a brief overview from chemistry to therapeutics. *J. Biol. Chem.* **2001**, *276*, 23213–23216. [CrossRef]
8. Zheng, R.; Jenkins, T.M.; Craigie, R. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 13659–13664. [CrossRef]
9. Ceccherini-Silberstein, F.; Malet, I.; Fabeni, L.; Dimonte, S.; Svicher, V.; D'Arrigo, R.; Artese, A.; Costa, G.; Bono, S.; Alcaro, S.; et al. Specific HIV-1 integrase polymorphisms change their prevalence in untreated versus antiretroviral-treated HIV-1-infected patients, all naive to integrase inhibitors. *J. Antimicrob. Chemother.* **2010**, *65*, 2305–2318. [CrossRef]
10. Brown, P.O.; Bowerman, B.; Varmus, H.E.; Bishop, J.M. Retroviral integration: Structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 2525–2529. [CrossRef]
11. FDA. FDA-Approved HIV Medicines. 2021. Available online: https://hivinfo.nih.gov/understanding-hiv/fact-sheets/fda-approved-hiv-medicines (accessed on 22 December 2021).
12. Yang, L.L.; Li, Q.; Zhou, L.B.; Chen, S.Q. Meta-analysis and systematic review of the efficacy and resistance for human immunodeficiency virus type 1 integrase strand transfer inhibitors. *Int. J. Antimicrob. Agents* **2019**, *54*, 547–555. [CrossRef]
13. Han, Y.S.; Mesplède, T.; Wainberg, M.A. Differences among HIV-1 subtypes in drug resistance against integrase inhibitors. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* **2016**, *46*, 286–291. [CrossRef]
14. Dow, D.E.; Bartlett, J.A. Dolutegravir, the Second-Generation of Integrase Strand Transfer Inhibitors (INSTIs) for the Treatment of HIV. *Infect Dis.* **2014**, *3*, 83–102. [CrossRef]
15. Lepik, K.J.; Harrigan, P.R.; Yip, B.; Wang, L.; Robbins, M.A.; Zhang, W.W.; Toy, J.; Akagi, L.; Lima, V.D.; Guillemi, S.; et al. Emergent drug resistance with integrase strand transfer inhibitor-based regimens. *AIDS* **2017**, *31*, 1425–1434. [CrossRef]
16. McGee, K.S.; Okeke, N.L.; Hurt, C.B.; McKellar, M.S. Canary in the Coal Mine? Transmitted Mutations Conferring Resistance to All Integrase Strand Transfer Inhibitors in a Treatment-Naive Patient. *Open Forum Infect. Dis.* **2018**, *5*, ofy294. [CrossRef]
17. Oliveira, M.; Ibanescu, R.I.; Anstett, K.; Mésplède, T.; Routy, J.P.; Robbins, M.A.; Brenner, B.G. Selective resistance profiles emerging in patient-derived clinical isolates with cabotegravir, bictegravir, dolutegravir, and elvitegravir. *Retrovirology* **2018**, *15*, 56. [CrossRef]
18. Young, B.; Fransen, S.; Greenberg, K.S.; Thomas, A.; Martens, S.; St Clair, M.; Petropoulos, C.J.; Ha, B. Transmission of integrase strand-transfer inhibitor multidrug-resistant HIV-1: Case report and response to raltegravir-containing antiretroviral therapy. *Antivir* **2011**, *16*, 253–256. [CrossRef]
19. Seatla, K.K.; Maruapula, D.; Choga, W.T.; Ntsipe, T.; Mathiba, N.; Mogwele, M.; Kapanda, M.; Nkomo, B.; Ramaabya, D.; Makhema, J.; et al. HIV-1 Subtype C Drug Resistance Mutations in Heavily Treated Patients Failing Integrase Strand Transfer Inhibitor-Based Regimens in Botswana. *Viruses* **2021**, *13*, 594. [CrossRef]

20. Bailey, A.J.; Rhee, S.Y.; Shafer, R.W. Integrase Strand Transfer Inhibitor Resistance in Integrase Strand Transfer Inhibitor-Naive Persons. *AIDS Res. Hum. Retrovir.* **2021**, *37*, 736–743. [CrossRef]
21. Casadellà, M.; Santos, J.R.; Noguera-Julian, M.; Micán-Rivera, R.; Domingo, P.; Antela, A.; Portilla, J.; Sanz, J.; Montero-Alonso, M.; Navarro, J.; et al. Primary resistance to integrase strand transfer inhibitors in Spain using ultrasensitive HIV-1 genotyping. *J. Antimicrob. Chemother.* **2020**, *75*, 3517–3524. [CrossRef]
22. Semengue, E.N.J.; Armenia, D.; Inzaule, S.; Santoro, M.M.; Dambaya, B.; Takou, D.; Teto, G.; Nka, A.D.; Yagai, B.; Fabeni, L.; et al. Baseline integrase drug resistance mutations and conserved regions across HIV-1 clades in Cameroon: Implications for transition to dolutegravir in resource-limited settings. *J. Antimicrob. Chemother.* **2021**, *76*, 1277–1285. [CrossRef]
23. Lübke, N.; Jensen, B.; Hüttig, F.; Feldt, T.; Walker, A.; Thielen, A.; Däumer, M.; Obermeier, M.; Kaiser, R.; Knops, E.; et al. Failure of Dolutegravir First-Line ART with Selection of Virus Carrying R263K and G118R. *N. Engl. J. Med.* **2019**, *381*, 887–889. [CrossRef]
24. Ndashimye, E.; Avino, M.; Olabode, A.S.; Poon, A.F.Y.; Gibson, R.M.; Li, Y.; Meadows, A.; Tan, C.; Reyes, P.S.; Kityo, C.M.; et al. Accumulation of integrase strand transfer inhibitor resistance mutations confers high-level resistance to dolutegravir in non-B subtype HIV-1 strains from patients failing raltegravir in Uganda. *J. Antimicrob. Chemother.* **2020**, *75*, 3525–3533. [CrossRef]
25. Pena, M.J.; Chueca, N.; D'Avolio, A.; Zarzalejos, J.M.; Garcia, F. Virological Failure in HIV to Triple Therapy With Dolutegravir-Based Firstline Treatment: Rare but Possible. *Open Forum Infect. Dis.* **2019**, *6*, ofy332. [CrossRef]
26. Blanco, J.L.; Marcelin, A.G.; Katlama, C.; Martinez, E. Dolutegravir resistance mutations: Lessons from monotherapy studies. *Curr. Opin Infect. Dis.* **2018**, *31*, 237–245. [CrossRef]
27. Brenner, B.G.; Thomas, R.; Blanco, J.L.; Ibanescu, R.I.; Oliveira, M.; Mesplède, T.; Golubkov, O.; Roger, M.; Garcia, F.; Martinez, E.; et al. Development of a G118R mutation in HIV-1 integrase following a switch to dolutegravir monotherapy leading to cross-resistance to integrase inhibitors. *J. Antimicrob. Chemother.* **2016**, *71*, 1948–1953. [CrossRef]
28. Fulcher, J.A.; Du, Y.; Zhang, T.H.; Sun, R.; Landovitz, R.J. Emergence of Integrase Resistance Mutations During Initial Therapy Containing Dolutegravir. *Clin. Infect. Dis.* **2018**, *67*, 791–794. [CrossRef]
29. Theys, K.; Libin, P.J.K.; Van Laethem, K.; Abecasis, A.B. An Evolutionary Model-Based Approach To Quantify the Genetic Barrier to Drug Resistance in Fast-Evolving Viruses and Its Application to HIV-1 Subtypes and Integrase Inhibitors. *Antimicrob. Agents Chemother.* **2019**, *63*, e00539-19. [CrossRef]
30. Mikasi, S.G.; Isaacs, D.; Chitongo, R.; Ikomey, G.M.; Jacobs, G.B.; Cloete, R. Interaction analysis of statistically enriched mutations identified in Cameroon recombinant subtype CRF02_AG that can influence the development of Dolutegravir drug resistance mutations. *BMC Infect. Dis.* **2021**, *21*, 379. [CrossRef]
31. Doyle, T.; Dunn, D.T.; Ceccherini-Silberstein, F.; De Mendoza, C.; Garcia, F.; Smit, E.; Fearnhill, E.; Marcelin, A.G.; Martinez-Picado, J.; Kaiser, R.; et al. Integrase inhibitor (INI) genotypic resistance in treatment-naive and raltegravir-experienced patients infected with diverse HIV-1 clades. *J. Antimicrob. Chemother.* **2015**, *70*, 3080–3086. [CrossRef]
32. Rogers, L.; Obasa, A.E.; Jacobs, G.B.; Sarafianos, S.G.; Sönnerborg, A.; Neogi, U.; Singh, K. Structural Implications of Genotypic Variations in HIV-1 Integrase From Diverse Subtypes. *Front. Microbiol.* **2018**, *9*, 1754. [CrossRef]
33. Bar-Magen, T.; Donahue, D.A.; McDonough, E.I.; Kuhl, B.D.; Faltenbacher, V.H.; Xu, H.; Michaud, V.; Sloan, R.D.; Wainberg, M.A. HIV-1 subtype B and C integrase enzymes exhibit differential patterns of resistance to integrase inhibitors in biochemical assays. *Aids* **2010**, *24*, 2171–2179. [CrossRef]
34. Kirichenko, A.; Lapovok, I.; Baryshev, P.; van de Vijver, D.; van Kampen, J.J.A.; Boucher, C.A.B.; Paraskevis, D.; Kireev, D. Genetic Features of HIV-1 Integrase Sub-Subtype A6 Predominant in Russia and Predicted Susceptibility to INSTIs. *Viruses* **2020**, *12*, 838. [CrossRef]
35. Quashie, P.K.; Mesplède, T.; Han, Y.S.; Oliveira, M.; Singhroy, D.N.; Fujiwara, T.; Underwood, M.R.; Wainberg, M.A. Characterization of the R263K mutation in HIV-1 integrase that confers low-level resistance to the second-generation integrase strand transfer inhibitor dolutegravir. *J. Virol.* **2012**, *86*, 2696–2705. [CrossRef]
36. Rhee, S.Y.; Grant, P.M.; Tzou, P.L.; Barrow, G.; Harrigan, P.R.; Ioannidis, J.P.A.; Shafer, R.W. A systematic review of the genetic mechanisms of dolutegravir resistance. *J. Antimicrob. Chemother.* **2019**, *74*, 3135–3149. [CrossRef]
37. Mouscadet, J.F.; Delelis, O.; Marcelin, A.G.; Tchertanov, L. Resistance to HIV-1 integrase inhibitors: A structural perspective. *Drug Resist. Updates* **2010**, *13*, 139–150. [CrossRef]
38. EPHI. HIV Related Estimats and Projections in Ethiopia for the Year-2019. 2020. Available online: http://repository.iifphc.org/bitstream/handle/123456789/1069/HIV_estimation_and_projection_for_Ethiopia_2019.pdf?sequence=1&isAllowed=y (accessed on 25 December 2021).
39. Arimide, D.A.; Esquivel-Gómez, L.R.; Kebede, Y.; Sasinovich, S.; Balcha, T.; Björkman, P.; Kühnert, D.; Medstrand, P. Molecular Epidemiology and Transmission Dynamics of the HIV-1 Epidemic in Ethiopia: Epidemic Decline Coincided With Behavioral Interventions Before ART Scale-Up. *Front. Microbiol.* **2022**, *13*, 821006. [CrossRef]
40. MOH. National Consolidated Guidelines for Comprehensive HIV Prevention, Care and Treatment. 2018. Available online: https://www.afro.who.int/sites/default/files/2019-04/National%20Comprehensive%20HIV%20Care%20%20Guideline%202018.pdf (accessed on 15 December 2021).
41. WHO. HIV Drug Resistance Surveillance Guidance: 2015 Update. 2015. Available online: https://www.who.int/publications/i/item/978-92-4-151009-7 (accessed on 27 December 2021).

42. WHO. *World Health Organization Global Strategy for the Surveillance and Monitoring of HIV Drug Resistance*; WHO Press: Geneva, Switzerland, 2012.
43. Van Laethem, K.; Schrooten, Y.; Covens, K.; Dekeersmaeker, N.; De Munter, P.; Van Wijngaerden, E.; Van Ranst, M.; Vandamme, A.M. A genotypic assay for the amplification and sequencing of integrase from diverse HIV-1 group M subtypes. *J. Virol. Methods* **2008**, *153*, 176–181. [CrossRef]
44. Woods, C.K.; Brumme, C.J.; Liu, T.F.; Chui, C.K.; Chu, A.L.; Wynhoven, B.; Hall, T.A.; Trevino, C.; Shafer, R.W.; Harrigan, P.R. Automating HIV drug resistance genotyping with RECall, a freely accessible sequence analysis tool. *J. Clin. Microbiol.* **2012**, *50*, 1936–1942. [CrossRef]
45. Pineda-Peña, A.C.; Faria, N.R.; Imbrechts, S.; Libin, P.; Abecasis, A.B.; Deforche, K.; Gómez-López, A.; Camacho, R.J.; de Oliveira, T.; Vandamme, A.M. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of the new REGA version 3 and seven other tools. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* **2013**, *19*, 337–348. [CrossRef]
46. Struck, D.; Lawyer, G.; Ternes, A.M.; Schmit, J.C.; Bercoff, D.P. COMET: Adaptive context-based modeling for ultrafast HIV-1 subtype identification. *Nucleic. Acids Res.* **2014**, *42*, e144. [CrossRef]
47. Martin, D.P.; Lemey, P.; Lott, M.; Moulton, V.; Posada, D.; Lefeuvre, P. RDP3: A flexible and fast computer program for analyzing recombination. *Bioinformatics* **2010**, *26*, 2462–2463. [CrossRef]
48. Katoh, K.; Kuma, K.; Miyata, T.; Toh, H. Improvement in the accuracy of multiple sequence alignment program MAFFT. *Genome Inf.* **2005**, *16*, 22–33.
49. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.
50. Tippmann, H.F. Analysis for free: Comparing programs for sequence analysis. *Brief. Bioinform.* **2004**, *5*, 82–87. [CrossRef]
51. Guindon, S.; Dufayard, J.F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [CrossRef]
52. Guindon, S.; Lethiec, F.; Duroux, P.; Gascuel, O. PHYML Online—A web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* **2005**, *33*, W557-9. [CrossRef]
53. Nguyen, H.L.; Ruxrungtham, K.; Delaugerre, C. Genetic barrier to the development of resistance to integrase inhibitors in HIV-1 subtypes CRF01_AE and B. *Intervirology* **2012**, *55*, 287–295. [CrossRef]
54. Li, M.; Chen, X.; Wang, H.; Jurado, K.A.; Engelman, A.N.; Craigie, R. A Peptide Derived from Lens Epithelium-Derived Growth Factor Stimulates HIV-1 DNA Integration and Facilitates Intasome Structural Studies. *J. Mol. Biol.* **2020**, *432*, 2055–2066. [CrossRef]
55. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [CrossRef]
56. Eberhardt, J.; Santos-Martins, D.; Tillack, A.F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61*, 3891–3898. [CrossRef]
57. Lu, R.; Limón, A.; Ghory, H.Z.; Engelman, A. Genetic analyses of DNA-binding mutants in the catalytic core domain of human immunodeficiency virus type 1 integrase. *J. Virol.* **2005**, *79*, 2493–2505. [CrossRef]
58. Inzaule, S.C.; Hamers, R.L.; Noguera-Julian, M.; Casadellà, M.; Parera, M.; Rinke de Wit, T.F.; Paredes, R. Primary resistance to integrase strand transfer inhibitors in patients infected with diverse HIV-1 subtypes in sub-Saharan Africa. *J. Antimicrob. Chemother.* **2018**, *73*, 1167–1172. [CrossRef]
59. Obasa, A.E.; Mikasi, S.G.; Brado, D.; Cloete, R.; Singh, K.; Neogi, U.; Jacobs, G.B. Drug Resistance Mutations Against Protease, Reverse Transcriptase and Integrase Inhibitors in People Living With HIV-1 Receiving Boosted Protease Inhibitors in South Africa. *Front. Microbiol.* **2020**, *11*, 438. [CrossRef] [PubMed]
60. Fish, M.Q.; Hewer, R.; Wallis, C.L.; Venter, W.D.; Stevens, W.S.; Papathanasopoulos, M.A. Natural polymorphisms of integrase among HIV type 1-infected South African patients. *AIDS Res. Hum. Retrovir.* **2010**, *26*, 489–493. [CrossRef] [PubMed]
61. Oliveira, M.F.; Ramalho, D.B.; Abreu, C.M.; Vubil, A.; Mabunda, N.; Ismael, N.; Francisco, C.; Jani, I.V.; Tanuri, A. Genetic diversity and naturally polymorphisms in HIV type 1 integrase isolates from Maputo, Mozambique: Implications for integrase inhibitors. *AIDS Res. Hum. Retrovir.* **2012**, *28*, 1788–1792. [CrossRef]
62. Mulu, A.; Maier, M.; Liebert, U.G. Lack of integrase inhibitors associated resistance mutations among HIV-1C isolates. *J. Transl. Med.* **2015**, *13*, 377. [CrossRef]
63. Rangel, H.R.; Garzaro, D.; Fabbro, R.; Martinez, N.; Ossenkop, J.; Torres, J.R.; Gutiérrez, C.R.; Pujol, F.H. Absence of primary integrase resistance mutations in HIV type 1-infected patients in Venezuela. *AIDS Res. Hum. Retrovir.* **2010**, *26*, 923–926. [CrossRef]
64. Brado, D.; Obasa, A.E.; Ikomey, G.M.; Cloete, R.; Singh, K.; Engelbrecht, S.; Neogi, U.; Jacobs, G.B. Analyses of HIV-1 integrase sequences prior to South African national HIV-treatment program and available of integrase inhibitors in Cape Town, South Africa. *Sci. Rep.* **2018**, *8*, 4709. [CrossRef] [PubMed]
65. Kim, J.Y.; Kim, E.J.; Choi, J.Y.; Kwon, O.K.; Kim, G.J.; Choi, S.Y.; Kim, S.S. Genetic variation of the HIV-1 integrase region in newly diagnosed anti-retroviral drug-naïve patients with HIV/AIDS in Korea. *Clin. Microbiol. Infect.* **2011**, *17*, 1155–1159. [CrossRef]

66. Kotaki, T.; Khairunisa, S.Q.; Sukartiningrum, S.D.; Witaningrum, A.M.; Rusli, M.; Diansyah, M.N.; Arfijanto, M.V.; Rahayu, R.P.; Nasronudin; Kameoka, M. Detection of drug resistance-associated mutations in human immunodeficiency virus type 1 integrase derived from drug-naive individuals in Surabaya, Indonesia. *AIDS Res. Hum. Retrovir.* **2014**, *30*, 489–492. [CrossRef]
67. Arruda, L.B.; Fonseca, L.A.; Duarte, A.J.; Casseb, J. Genetic diversity on the integrase region of the pol gene among HIV type 1-infected patients naive for integrase inhibitors in São Paulo City, Brazil. *AIDS Res. Hum. Retrovir.* **2010**, *26*, 105–107. [CrossRef]
68. Casadellà, M.; van Ham, P.M.; Noguera-Julian, M.; van Kessel, A.; Pou, C.; Hofstra, L.M.; Santos, J.R.; Garcia, F.; Struck, D.; Alexiev, I.; et al. Primary resistance to integrase strand-transfer inhibitors in Europe. *J. Antimicrob. Chemother.* **2015**, *70*, 2885–2888.
69. Meixenberger, K.; Yousef, K.P.; Smith, M.R.; Somogyi, S.; Fiedler, S.; Bartmeyer, B.; Hamouda, O.; Bannert, N.; von Kleist, M.; Kücherer, C. Molecular evolution of HIV-1 integrase during the 20 years prior to the first approval of integrase inhibitors. *Virol. J.* **2017**, *14*, 223. [CrossRef]
70. Tostevin, A.; White, E.; Dunn, D.; Croxford, S.; Delpech, V.; Williams, I.; Asboe, D.; Pozniak, A.; Churchill, D.; Geretti, A.M.; et al. Recent trends and patterns in HIV-1 transmitted drug resistance in the United Kingdom. *HIV Med.* **2017**, *18*, 204–213. [CrossRef] [PubMed]
71. Inzaule, S.C.; Hamers, R.L.; Noguera-Julian, M.; Casadellà, M.; Parera, M.; Kityo, C.; Steegen, K.; Naniche, D.; Clotet, B.; Rinke de Wit, T.F.; et al. Clinically relevant thresholds for ultrasensitive HIV drug resistance testing: A multi-country nested case-control study. *Lancet HIV* **2018**, *5*, e638–e646. [CrossRef]
72. Ndashimye, E.; Avino, M.; Kyeyune, F.; Nankya, I.; Gibson, R.M.; Nabulime, E.; Poon, A.F.Y.; Kityo, C.; Mugyenyi, P.; Quiñones-Mateu, M.E.; et al. Absence of HIV-1 Drug Resistance Mutations Supports the Use of Dolutegravir in Uganda. *AIDS Res. Hum. Retrovir.* **2018**, *34*, 404–414. [CrossRef] [PubMed]
73. Karade, S.; Sen, S.; Sashindran, V.K. Absence of Integrase Strand Transfer Inhibitor Associated Resistance in Antiretroviral Therapy Naïve and Experienced Individuals from Western India. *AIDS Res. Hum. Retrovir.* **2019**, *35*, 567–571. [CrossRef]
74. Kityo, C.; Thompson, J.; Nankya, I.; Hoppe, A.; Ndashimye, E.; Warambwa, C.; Mambule, I.; van Oosterhout, J.J.; Wools-Kaloustian, K.; Bertagnolio, S.; et al. HIV Drug Resistance Mutations in Non-B Subtypes After Prolonged Virological Failure on NNRTI-Based First-Line Regimens in Sub-Saharan Africa. *J. Acquir. Immune Defic. Syndr.* **2017**, *75*, e45–e54. [CrossRef] [PubMed]
75. Lan, Y.; Li, L.; Chen, W.; Deng, X.; Li, J.; Fan, Q.; Cai, X.; Cai, W.; Hu, F. Absence of Integrase Inhibitor-Associated Resistance Among Antiretroviral Therapy-Naïve HIV-1-Infected Adults in Guangdong Province, China, in 2018. *Infect. Drug Resist.* **2020**, *13*, 4389–4394. [CrossRef]
76. Anstett, K.; Cutillas, V.; Fusco, R.; Mesplède, T.; Wainberg, M.A. Polymorphic substitution E157Q in HIV-1 integrase increases R263K-mediated dolutegravir resistance and decreases DNA binding activity. *J. Antimicrob. Chemother.* **2016**, *71*, 2083–2088. [CrossRef] [PubMed]
77. Fun, A.; Van Baelen, K.; van Lelyveld, S.F.; Schipper, P.J.; Stuyver, L.J.; Wensing, A.M.; Nijhuis, M. Mutation Q95K enhances N155H-mediated integrase inhibitor resistance and improves viral replication capacity. *J. Antimicrob. Chemother.* **2010**, *65*, 2300–2304. [CrossRef]
78. Garrido, C.; Villacian, J.; Zahonero, N.; Pattery, T.; Garcia, F.; Gutierrez, F.; Caballero, E.; Van Houtte, M.; Soriano, V.; de Mendoza, C. Broad phenotypic cross-resistance to elvitegravir in HIV-infected patients failing on raltegravir-containing regimens. *Antimicrob. Agents Chemother.* **2012**, *56*, 2873–2878. [CrossRef]
79. Temesgen, Z.; Siraj, D.S. Raltegravir: First in class HIV integrase inhibitor. *Clin. Risk Manag.* **2008**, *4*, 493–500. [CrossRef]
80. Passaes, C.B.; Guimarães, M.L.; Fernandez, S.L.; Lorete Rdos, S.; Teixeira, S.L.; Fernandez, J.C.; Morgado, M.G. Lack of primary mutations associated with integrase inhibitors among HIV-1 subtypes B, C, and F circulating in Brazil. *J. Acquir. Immune Defic. Syndr.* **2009**, *51*, 7–12. [CrossRef] [PubMed]
81. Wares, M.; Mesplède, T.; Quashie, P.K.; Osman, N.; Han, Y.; Wainberg, M.A. The M50I polymorphic substitution in association with the R263K mutation in HIV-1 subtype B integrase increases drug resistance but does not restore viral replicative fitness. *Retrovirology* **2014**, *11*, 7. [CrossRef]
82. Hackett, J.; Harris, B.; Holzmayer, V.; Yamaguchi, J.; Luk, K.C.; Brennan, C.; Schochetman, G.; Devare, S.; Swanson, P. Naturally occurring polymorphisms in HIV-1 Group M, N, and O Integrase: Implications for integrase inhibitors. In Proceedings of the Fifteenth Conference on Retroviruses and Opportunistic Infections, Boston, MA, USA, 2–6 February 2008.
83. Giacomelli, A.; Lai, A.; Franzetti, M.; Maggiolo, F.; Di Giambenedetto, S.; Borghi, V.; Francisci, D.; Magnani, G.; Pecorari, M.; Monno, L.; et al. No impact of previous NRTIs resistance in HIV positive patients switched to DTG+2NRTIs under virological control: Time of viral suppression makes the difference. *Antivir. Res.* **2019**, *172*, 104635. [CrossRef] [PubMed]
84. Olearo, F.; Nguyen, H.; Bonnet, F.; Yerly, S.; Wandeler, G.; Stoeckle, M.; Cavassini, M.; Scherrer, A.; Costagiola, D.; Schmid, P.; et al. Impact of the M184V/I Mutation on the Efficacy of Abacavir/Lamivudine/Dolutegravir Therapy in HIV Treatment-Experienced Patients. *Open Forum Infect. Dis.* **2019**, *6*, ofz330. [CrossRef]
85. Andreatta, K.; Willkom, M.; Martin, R.; Chang, S.; Wei, L.; Liu, H.; Liu, Y.-P.; Graham, H.; Quirk, E.; Martin, H.; et al. Switching to bictegravir/emtricitabine/tenofovir alafenamide maintained HIV-1 RNA suppression in participants with archived antiretroviral resistance including M184V/I. *J. Antimicrob. Chemother.* **2019**, *74*, 3555–3564. [CrossRef]

86. Chen, G.J.; Sun, H.Y.; Chang, S.Y.; Cheng, A.; Huang, Y.S.; Lin, K.Y.; Huang, Y.C.; Su, Y.C.; Liu, W.C.; Hung, C.C. Effectiveness of switching from protease inhibitors to dolutegravir in combination with nucleoside reverse transcriptase inhibitors as maintenance antiretroviral therapy among HIV-positive patients. *Int. J. Antimicrob. Agents* **2019**, *54*, 35–42. [CrossRef]
87. Van Hal, S.J.; Herring, B.; Deris, Z.; Wang, B.; Saksena, N.K.; Dwyer, D.E. HIV-1 integrase polymorphisms are associated with prior antiretroviral drug exposure. *Retrovirology* **2009**, *6*, 12. [CrossRef]
88. Brenner, B.G.; Lowe, M.; Moisi, D.; Hardy, I.; Gagnon, S.; Charest, H.; Baril, J.G.; Wainberg, M.A.; Roger, M. Subtype diversity associated with the development of HIV-1 resistance to integrase inhibitors. *J. Med. Virol.* **2011**, *83*, 751–759. [CrossRef]
89. Maïga, A.I.; Malet, I.; Soulie, C.; Derache, A.; Koita, V.; Amellal, B.; Tchertanov, L.; Delelis, O.; Morand-Joubert, L.; Mouscadet, J.F.; et al. Genetic barriers for integrase inhibitor drug resistance in HIV type-1 B and CRF02_AG subtypes. *Antivir* **2009**, *14*, 123–129.
90. Telele, N.F.; Kalu, A.W.; Gebre-Selassie, S.; Fekade, D.; Abdurahman, S.; Marrone, G.; Neogi, U.; Tegbaru, B.; Sönnerborg, A. Pretreatment drug resistance in a large countrywide Ethiopian HIV-1C cohort: A comparison of Sanger and high-throughput sequencing. *Sci. Rep.* **2018**, *8*, 7556. [CrossRef] [PubMed]

Printed by Media-Tryck, Lund 2022 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Translational Medicine

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2022:105
ISBN 978-91-8021-266-3
ISSN 1652-8220