

RESEARCH ARTICLE

Genetic Characterization of a Panel of Diverse HIV-1 Isolates at Seven International Sites

Bhavna Hora¹✉, Sheila M. Keating^{2,3}✉, Yue Chen¹, Ana M. Sanchez¹, Ester Sabino⁴, Gillian Hunt⁵, Johanna Ledwaba⁵, John Hackett, Jr⁶, Priscilla Swanson⁶, Indira Hewlett⁷, Viswanath Ragupathy⁷, Sai Vikram Vemula⁷, Peibin Zeng⁸, Kok-Keng Tee⁹, Wei Zhen Chow⁹, Hezhao Ji¹⁰, Paul Sandstrom¹⁰, Thomas N. Denny¹, Michael P. Busch^{2,3*}, Feng Gao^{1*}, on behalf of the REDS-III and EQAPOL programs[¶]

1 Duke Human Vaccine Institute and Department of Medicine, Duke University Medical Center, Durham, North Carolina, United States of America, **2** Blood Systems Research Institute, San Francisco, California, United States of America, **3** Department of Laboratory Medicine, University of California, San Francisco, California, United States of America, **4** Instituto de Medicina Tropical, Sao Paulo Brazil, **5** National Institute of Communicable Diseases, Johannesburg, South Africa, **6** Abbott Laboratories, Infectious Diseases Research, Abbott Park, Illinois, United States of America, **7** Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Springs, Maryland, United States of America, **8** Institute of Blood Transfusion, Chinese Academy of Medical Sciences, Chengdu, China, **9** Centre of Excellence for Research in AIDS, Department of Medical Microbiology, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia, **10** National HIV & Retrovirology Laboratories at JC Wilt Infectious Diseases Research Center, Public Health Agency of Canada, Winnipeg, Canada

✉ These authors contributed equally to this work.

¶ Membership of the REDS-III and EQAPOL is provided in the acknowledgments.

* fgao@duke.edu (FG); mbsusch@bloodsystems.org (MPB)



OPEN ACCESS

Citation: Hora B, Keating SM, Chen Y, Sanchez AM, Sabino E, Hunt G, et al. (2016) Genetic Characterization of a Panel of Diverse HIV-1 Isolates at Seven International Sites. PLoS ONE 11(6): e0157340. doi:10.1371/journal.pone.0157340

Editor: Dimitrios Paraskevis, University of Athens, Medical School, GREECE

Received: February 15, 2016

Accepted: May 29, 2016

Published: June 17, 2016

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](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All data are available from the paper or GenBank (Accession numbers: JX140645-JX140653, JX140655-X140663, JX140665, JX140668, JX140670-JX140672, JX140675, JX140677-JX140679, KC473825, KC473827, KC473836-KC473840, KC473842-KC473846, KC596061, KC596066-KC596069, KC596071-KC596073, KF859742-KF859744).

Funding: This study was supported by the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH) for External Quality Assurance Program Oversight Laboratory (EQAPOL) (HHSN272201000045C and HHSN272200800014C),

Abstract

HIV-1 subtypes and drug resistance are routinely tested by many international surveillance groups. However, results from different sites often vary. A systematic comparison of results from multiple sites is needed to determine whether a standardized protocol is required for consistent and accurate data analysis. A panel of well-characterized HIV-1 isolates (N = 50) from the External Quality Assurance Program Oversight Laboratory (EQAPOL) was assembled for evaluation at seven international sites. This virus panel included seven subtypes, six circulating recombinant forms (CRFs), nine unique recombinant forms (URFs) and three group O viruses. Seven viruses contained 10 major drug resistance mutations (DRMs). HIV-1 isolates were prepared at a concentration of 10⁷ copies/ml and compiled into blinded panels. Subtypes and DRMs were determined with partial or full *pol* gene sequences by conventional Sanger sequencing and/or Next Generation Sequencing (NGS). Subtype and DRM results were reported and decoded for comparison with full-length genome sequences generated by EQAPOL. The partial *pol* gene was amplified by RT-PCR and sequenced for 89.4%-100% of group M viruses at six sites. Subtyping results of majority of the viruses (83%-97.9%) were correctly determined for the partial *pol* sequences. All 10 major DRMs in seven isolates were detected at these six sites. The complete *pol* gene sequence was also obtained by NGS at one site. However, this method missed six group M viruses and sequences contained host chromosome fragments. Three group O viruses were only

the National Heart, Lung, and Blood Institute (NHLBI), NIH for funding for the Recipient Epidemiology and Donor Evaluation Study-III (REDS-III) (HHSN2682011000011, HHSN2682011000071, HSN2682011000081 and HHSN2682011000091), Global HIV panel project (DAIDS IAA - Y1-HB-5026-01) and the High Impact Research grant for KKT (UM.C/625/1/HIR/MOE/CHAN/02/02). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Abbott Laboratories did not play a role in the study design, decision to publish or preparation of the manuscript and only provided financial support in the form of authors' salaries and research materials.

Competing Interests: John Hackett and Priscilla Swanson are employees and shareholders of Abbott Laboratories. These affiliations do not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

characterized with additional group O-specific RT-PCR primers employed by one site. These results indicate that PCR protocols and subtyping tools should be standardized to efficiently amplify diverse viruses and more consistently assign virus genotypes, which is critical for accurate global subtype and drug resistance surveillance. Targeted NGS analysis of partial *pol* sequences can serve as an alternative approach, especially for detection of low-abundance DRMs.

Introduction

One hallmark of HIV-1 characteristics is its extraordinary genetic variability. HIV-1 is classified into four groups (M, N, O and P) [1–4]. HIV-1 group M viruses, which are responsible for most infections in the world can be further classified into subtypes (A, B, C, D, F, G, H, J and K), sub-subtypes (A1–A4 and F1–F2), circulating recombinant forms (CRFs) and countless unique recombinant forms (URFs) [5, 6]. These recombinant viruses account for at least 20% of HIV infections in recent studies [7]. Therefore, the high level genetic diversity of human immunodeficiency virus type I (HIV-1) poses significant public-health and clinical challenges, including implications for performance of blood donor screening, diagnostic testing and patient monitoring by viral load and drug resistance testing [1, 8–10]. Determination of viral loads in HIV-1 infected individuals is critical for prognosis and monitoring the efficacy of anti-viral therapy. The cross-subtype reactivity of nucleic acid tests (NATs) for measuring viral loads of different subtypes has been significantly improved in recent years. Although some assays show better detection efficiency for different HIV-1 subtypes and even among different HIV-1 groups [11, 12] the sensitivity of some NATs is still affected by different subtypes [12, 13]

Distinct prevalence of drug resistance mutations (DRMs) and kinetics of DRM emergence were observed among HIV-1 subtypes [14–18], although other studies indicated that outcomes of antiretroviral therapies (ART) may not be affected by subtypes and differences in virological and immunologic responses to ART might be contributed by ethnicity and/or adherences [19–21]. Drug resistance mutations (DRMs) are a major cause of antiretroviral therapy failure, and hence surveillance for the emergence of drug resistance among all subtypes and recombinants has been implemented in ART programs. The worldwide effort to improve treatment outcomes and reduce transmission of HIV-1 through early and optimal delivery of ART and HIV prevention programs should be a coordinated effort that includes national, regional, and global evaluations of HIV-1 drug resistance [22, 23].

The Recipient Epidemiology and Donor Evaluation Study-III (REDS-III) is a multicenter transfusion safety research program, launched in 2011 by the National Institute of Heart, Lung, and Blood of the National Institutes of Health, USA [24]. It is intended that REDS-III serve as an impetus for more widespread recipient and linked donor-recipient research in the United States and other international locations. It includes a domestic component and three distinct international programs in Brazil, China, and South Africa. The REDS-III portfolio includes molecular surveillance of HIV-1 infections in donors [25–27] which necessitated comparison, training and adoption of improved sequencing methods for the partial *pol* gene and/or HIV-1 whole genome to enable high yield of sequence data, better drug resistance classification, and subtype assignment.

Many international surveillance groups, including the REDS-III network, are involved in HIV-1 genetic analyses but in the absence of standardized protocols and external quality

assurance these studies may report inconsistent results [28, 29]. The REDS-III and the External Quality Assurance Program Oversight Laboratory (EQAPOL) programs determined that a systematic comparison of results from multiple sites was needed to determine if standardized protocols are required for consistent reports and accurate data analysis. Therefore, the REDS-III and EQAPOL networks planned a multi-center comparison study to assess the molecular methods for accurately determining genotype and testing for DRMs in international reference laboratories. The aim of this study was to compare results of the partial or full *pol* gene sequences generated with various RT-PCR conditions and sequencing methods at seven different international laboratories.

Materials and Methods

Assembly of a panel of diverse HIV-1 isolates

A large panel of HIV-1 isolates representing HIV-1 groups, subtypes, CRFs and URFs from different countries was assembled at EQAPOL. The HIV-1 strains were derived from either plasma samples or established isolates and expanded to high titer stocks by short-term culturing with peripheral blood mononuclear cells (PBMC) from HIV-1 negative donors [30]. All virus donations to EQAPOL were collected with ethical review and approval of the collection protocol and informed consent as previously established for the donor organization's collection activities. This study was approved by the Institutional Review Board (IRB) at Duke University under the IRB exemption protocol from 45 CFR 46. Before previously acquired samples were delivered to Duke University, all samples were de-identified under IRB exemption and no identifiable subject information was provided. The same de-identified virus isolates were sent to evaluation sites for further analysis. Near full-length genome sequences were obtained for all viruses by single genome amplification (SGA) at EQAPOL [30]. The subtype or recombination pattern of each virus was determined by phylogenetic tree analysis (Neighbor-joining method with Kimura two-parameter model), the REGA subtyping tool version 2 (<http://jose.med.kuleuven.ac.be/genotypetool/html/subtypinghiv.html>) and the jumping profile Hidden Markov Model (jpHMM) tool (<http://jphmm.gobics.de/>) [31]. If discordant subtyping results were observed for any regions in a sequence among different algorithms, phylogenetic trees were constructed for this particular region to determine the subtype origin of this region of sequence.

Fifty well-characterized diverse HIV-1 isolates from 20 countries were selected to establish the EQAPOL Genetic Diversity Panel. This 50-member panel consists of 28 viruses from six subtypes and two sub-subtypes (A, B, C, D, F1, F2 and G), 10 from six CRFs (CRF01_AE, CRF02_AG, CRF04_cpx, CRF14_BG, CRF24_BG, and CRF47_BF), 9 URFs and three group O viruses. Each virus was diluted into HIV-1 negative defibrinated plasma (Gemini Bio-Product, Sacramento, CA) to a concentration of 10^7 viral RNA copies/ml and 1 mL cryopreserved aliquots were coded before they were sent out to test sites.

International evaluation sites

Seven sites participated in this study. Three of these sites are a part of the international component of REDS-III initiative: the South African National Blood Services (SANBS) in collaboration with the National Institute of Communicable Diseases, South Africa; the Instituto de Medicina Tropical, Brazil; and the Institute of Blood Transfusion, Chinese Academy of Medical Sciences, China. The other four sites were Abbott Laboratories, USA (USA_Abbott); the Laboratory of Molecular Virology at the US Food and Drug Administration, USA (USA_FDA); the University of Malaya, Malaysia; and the National HIV & Retrovirology Laboratories, Public Health Agency of Canada, Canada.

Subtype determination and detection of drug resistance mutations

Each test site received the same blinded panel of 50 viruses for evaluation and used their own protocol for RT-PCR amplification, sequencing the PCR amplicons, determination of subtypes and detection of major and minor DRMs (S1 Table). Genotypes and DRMs were determined with partial *pol* sequences (protease and partial reverse transcriptase) at six sites (Brazil, South Africa, USA_Abbott, China, Malaysia and Canada) or with complete *pol* sequences generated by NGS at the USA_FDA site. All 50 viruses were analyzed by both Sanger sequencing and NGS (MiSeq, Illumina, San Diego, CA) methods at the Canada site. Raw reads were analyzed using HyDRA, a proprietary data analytical pipeline developed for NGS-based HIV drug resistance (DR) analysis (<http://hydra.canada.ca>) [32]. To generate *pol* gene consensus sequences at the USA_FDA site, raw reads for the *pol* gene were imported to High-performance Integrated Virtual Environment (HIVE; <https://hive.biochemistry.gwu.edu/>) and an algorithmic pipeline was used to generate consensus sequences obtained by NGS. Using genotype reference sets from the Los Alamos HIV-1 sequence database (<http://www.hiv.lanl.gov>) and the hexagon tool in HIVE, reference mapping was performed. Raw reads were assembled with the clonal assembly tool in HIVE. Consensus sequences were generated by linking quasispecies clones for further downstream analysis.

Comparison of results from all sites

Results (sequences, genotypes and DRMs) for the blinded panels from each test site were reported back to EQAPOL. The results were decoded and compared to near full-length reference sequences generated at EQAPOL. Sequences from each site were aligned to the EQAPOL reference sequences using CLUSTAL W and manually adjusted for optimal alignments with Seaview [33]. If the site subtype assignment and sequence matched the EQAPOL reference sequence, no further analysis was performed. When the site sequences differed from the EQAPOL references, further analysis was carried out to identify the causes for the differences. Subtyping results from each site were categorized as concordant or discordant relative to EQAPOL results. The major and minor DRMs from each site were also compared with those identified by analysis of the EQAPOL sequences. If the reported DRMs were discordant, sequences were further examined for mutations or ambiguous bases at the affected codons.

GenBank accession numbers

GenBank accession numbers of near full length HIV sequences are: JX140645—JX140653, JX140655—JX140663, JX140665, JX140668, JX140670—JX140672, JX140675, JX140677—JX140679, KC473825, KC473827, KC473836 -KC473840, KC473842—KC473846, KC596061, KC596066—KC596069, KC596071—KC596073, KF859742—KF859744.

Results

Majority of group M viruses were amplified at all sites

The partial *pol* gene was successfully amplified for all 47 group M viruses at the South Africa, USA_Abbott, Malaysia and Canada sites. It was amplified from 42 and 44 group M viruses at the Brazil and China sites, respectively. All failed PCR amplifications were from independent HIV-1 isolates among all sites except DE00400GR002 that was not amplified at both Brazil and China sites (Table 1), suggesting the PCR negative results were not strain specific. Complete *pol* gene sequences were obtained from 41 samples generated by NGS at the USA_FDA site. All three group O viruses were amplified only at the Malaysia site when an additional PCR primer set specific for group O viruses was used, according to their standard protocol for samples that

Table 1. Comparison of subtyping results from all sites.

Sample Name	EQAPOL		Brazil	South Africa	USA_Abbott	China	Malaysia	Canada	USA_FDA #
	Full length genome	Partial pol							
DEMA105TZ001	A1	A1	CRF01	A1	A1	A1	CRF02	A1	Neg
DEMA106ES002	A1	A1	A1	A1	A1	A1	A1	A1	A1 (AF1)
DEMB03JP004	B	B	B	B	B	B	B	B	B
DEMB05FR001	B	B	B	B	B	B	B	B	B
DEMB08ES001	B	B	B	B	B	B	B	B	B
DEMB08UY001	B	B	B	B	B	B	B	B	B
DEMB09BO001	B	B	Neg	B	B	B	B	B	B
DEMB09CN002	B	B	B	B	B	B	B	B	B
DEMB09US002	B	B	B	B	B	B	B	B	B
DEMB10CN002	B	B	B	B	B	B	B	B	B
DEMB10ES002	B	B	B	B	B	B	B	B	B
DEMB10ES003	B	B	B	B	B	B	B	B	Neg
DEMB10US001	B	B	B	B	B	B	B	B	Neg
DEMB10US004	B	B	B	B	B	B	B	B/CRF01	B (BD)
DEMB10VE001	B	B	B	B,D	B	B	B	B	B
DEMBXXDE001	B	B	B	B	B	B	B	B	B
DEMBXXPL001	B	B	Neg	B	B	B	B	B	B
DEMC06ES003	C	C	C	C	C	C	C	C	C (CB)
DEMC07AO001	C	C	C	C	C	C	C	C	C (CF1)
DEMC07BR003	C	C	C	C	C	C	C	C	C (CBF1)
DEMC08NG001	C	C	C	C	C	C	C	C	Neg
DEMC09ZA009	C	C	C	C	C	Neg	C	C	C
DEMD07UG002	D	D	D	D	D	D	CRF10	D	D
DEMD08UG001	D	D	D	D	D	D	D	D	D
DEMD10CM009	D	D	D	D	D	D	D	D	D
DEMF110ES001	F1	F1	F1	F1	F1	F1	F1	F1	F1 (F1BD)
DEMF210CM001	F2	F2	B/F	F2	F2	F2	F2	F2	F2 (F2A)
DEMG09ES002	G	G	G	G	CRF14	G	G	G	Neg
DE00109CN003	CRF01	A1	CRF01	CRF01	CRF01	CRF01	CRF01	CRF01	CRF01
DE00208CM004	CRF02	A1G	CRF02	A1,G	CRF02	CRF02	DG	CRF02	CRF02
DE00208CM001	CRF02	A1G	CRF02	A1,G	CRF02	CRF02	CRF02	CRF02	CRF02
DE00206AO001	CRF02	A1G	CRF02	A1,G	CRF02	CRF02	CRF02	CRF02	CRF02 (02B)
DE00400GR002	CRF04	A1FGK	Neg	CRF04	CRF04	Neg	CRF04	CRF04	A1FK
DE01405BR001	CRF14	G	G	G	<i>CRF14</i>	<i>CRF14</i>	<i>CRF14</i>	G	BF (BFGDA1F1)
DE01405ES002	CRF14	G	G	G	<i>CRF14</i>	<i>CRF14</i>	<i>CRF14</i>	<i>CRF14</i>	CRF14_BG (BGD)
DE02408ES002	CRF24	BG	Neg	<i>CRF24</i>	<i>CRF24</i>	<i>CRF24</i>	<i>CRF24</i>	<i>CRF24</i>	BG
DE04708ES003	CRF47	BF1	B	B,F	<i>CRF47</i>	<i>CRF47</i>	<i>CRF47</i>	<i>CRF47</i>	BF1
DE04708ES004	CRF47	BF1	B	B,F	<i>CRF47</i>	<i>CRF47</i>	<i>CRF47</i>	<i>CRF47</i>	BF1
DEURF09ES005	URF_A1B	B	B	B	B	<u>A1B</u>	B	B	B
DEURF10US008	URF_A1B	B	B	B	B	<u>A1B</u>	B	B	B
DEURF07UG006	URF_A1D	D	D	D	D	<u>A1D</u>	D	D	D
DEURF09GQ001	URF_A1DG	G	G	G,A*	URF_G/CRF02	<u>A1DG</u>	G	G	Neg

(Continued)

Table 1. (Continued)

Sample Name	EQAPOL		Brazil	South Africa	USA_Abbott	China	Malaysia	Canada	USA_FDA #
	Full length genome	Partial <i>pol</i>							
DEURF07BR002	URF_BC	C	C	C	C	Neg	CD	C	BC
DEMBF09ES003	URF_BF1	F1	F1	F1	F1	<u>BF1</u>	CRF72	F1	BF1
DEMBF09ES006	URF_BF1	BF1	D/B	B,F1	BF1	BF1	<i>CRF12</i>	<i>CRF12</i>	BF1
DEURF07ES002	URF_02/A3	A3	Neg	A1	A3	URF_02/A1	A1	B	URF_02/A1
DEURF10DZ001	URF_02/06/01	CRF06	CRF02	CRF06	CRF06	U	CRF02	CRF02	URF_AGKD
DEOXXDE004	O	O	Neg	Neg	Neg	Neg	O	Neg	Neg
DEOXXES001	O	O	Neg	Neg	Neg	Neg	O	Neg	Neg
DEOXXUS001	O	O	Neg	Neg	Neg	Neg	O	Neg	Neg

Subtype results provided by sites were compared to those based on the near full length HIV-1 genome sequences generated by EQAPOL. Negative PCR is indicated as “Neg”, discordant subtypes from sites are indicated in bold, and subtype results that are correctly obtained for CRFs based on the partial *pol* sequences are indicated in italics. Subtype results that are discordant for the partial *pol* gene but concordant for near full-length genome are underlined. Additional subtype fragments are shown in parentheses.

#: Complete *pol* gene sequence.

*: longer 3’ end sequence that contains subtype A region.

doi:10.1371/journal.pone.0157340.t001

fail to amplify with group M primers (Table 1). The overall successful RT-PCR amplification rate among six sites that amplified the partial *pol* gene averaged 95.7% (range from 89.4% to 100%) for group M viruses (Fig 1).

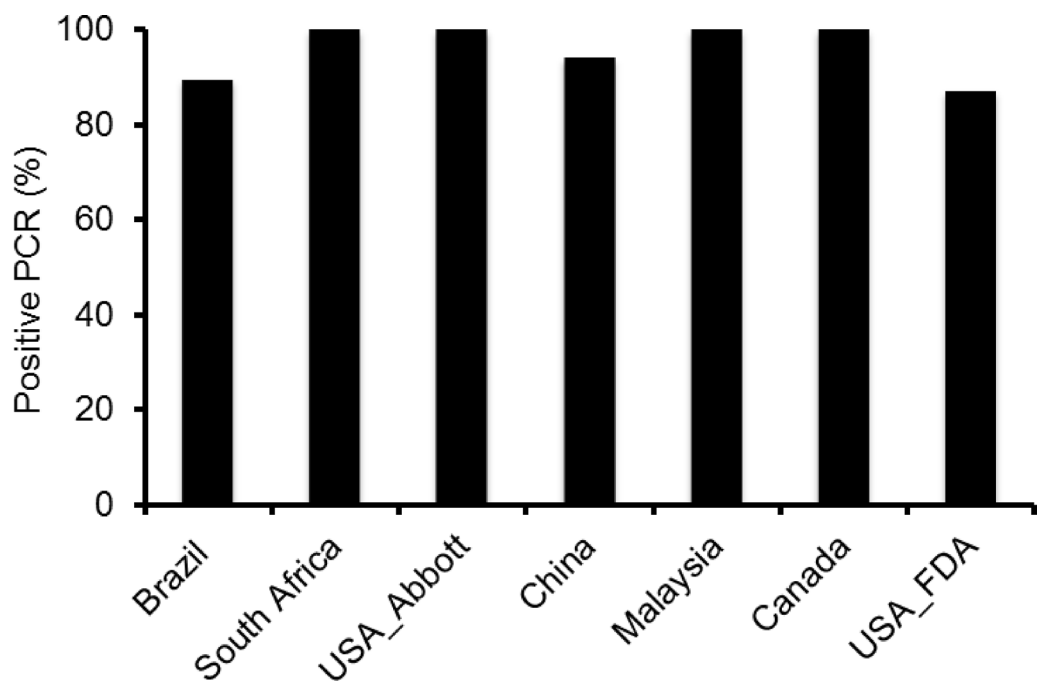


Fig 1. Comparison of PCR success rates. Partial *pol* genes were amplified for 47 group M viruses at six sites. The whole *pol* gene was amplified at the USA_FDA site.

doi:10.1371/journal.pone.0157340.g001

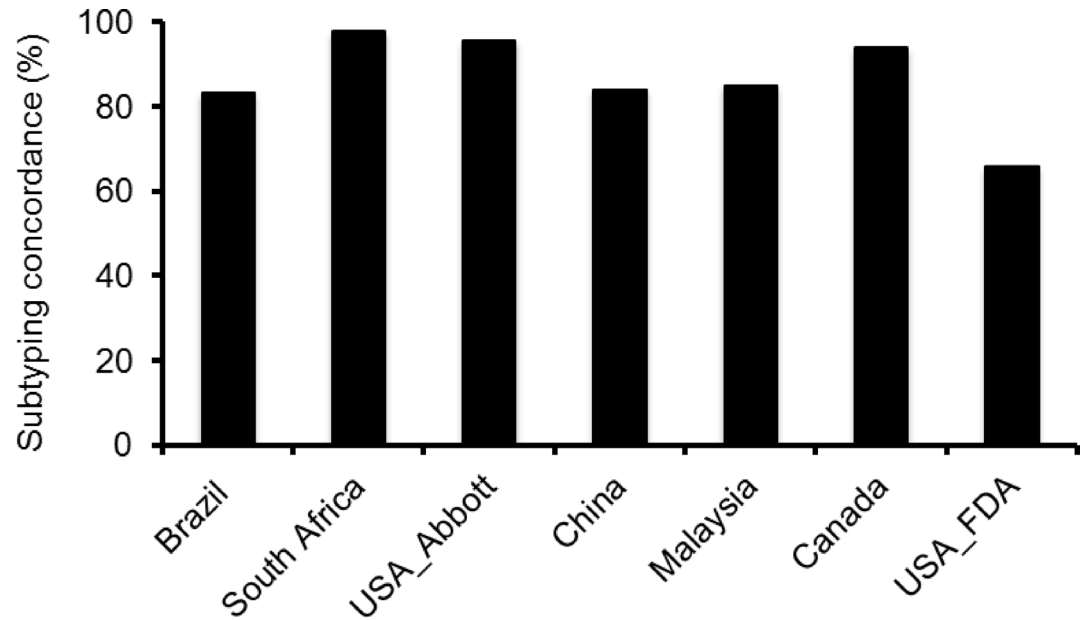


Fig 2. Comparison of subtype results across all sites. The subtype results from each site were compared to those determined by analyzing the 47 group M whole genome sequences at EQAPOL.

doi:10.1371/journal.pone.0157340.g002

Majority of subtyping results were in agreement with the EQAPOL reference sequences

Each site used their own genetic analysis tools to determine subtypes of partial or full *pol* gene sequences. Between 1 and 7 (2.1%-14.9%) partial *pol* sequences had discordant subtype results with the EQAPOL reference sequences at South Africa, USA_Abbott, Malaysia, Canada, Brazil and China sites (Fig 2 and Table 1). All sites used one subtyping method, except the South Africa site that used three different subtyping methods (RIP, REGA and SIMPLOT) among which the results from only the REGA method (the highest concordant rate) were used for comparison. Fewer discordant subtyping results were reported for pure subtypes (6 of 168 comparisons) than recombinants (20 of 114 comparisons) in the partial *pol* gene sequences ($p = 0.00032$).

To investigate the causes for the discordant subtyping results, we performed additional in-depth analysis by aligning the site sequences, EQAPOL references sequences, and subtype/CRF references from the Los Alamos HIV-1 sequence database. Of the six discordant subtyping results for pure subtype viruses based on the partial *pol* sequences, two viruses had discordant results at the Brazil and Malaysia sites and one virus was discordant at the USA_Abbott and Canada sites. No discordant results were reported by the China site. Discordant result (B/D recombinant) was reported for DEMB10VE001 by the South Africa site. Analysis with two different tools (RIP and SIMPLOT) confirmed that the virus was subtype B (Table 1).

When the partial *pol* sequences from CRFs and URFs containing recombination break-points were analyzed, 20 discordant subtyping results were observed for 8 viruses (bold and underlined letters in Table 1 and Fig 3). Discordant results with 1 to 7 viruses were reported from each site. Only one discordant result was found at the South Africa and USA_Abbott sites. Four to seven discordant results were reported at the Brazil, Malaysia and China sites. Two viruses (DEURF10DZ001 and DEURF07ES002) were classified differently by five sites (Table 1), indicating that subtyping analysis can be very challenging for complex recombinants. DEURF07ES002 is an A3 sub-subtype in the sequenced region but it was classified as A1 at

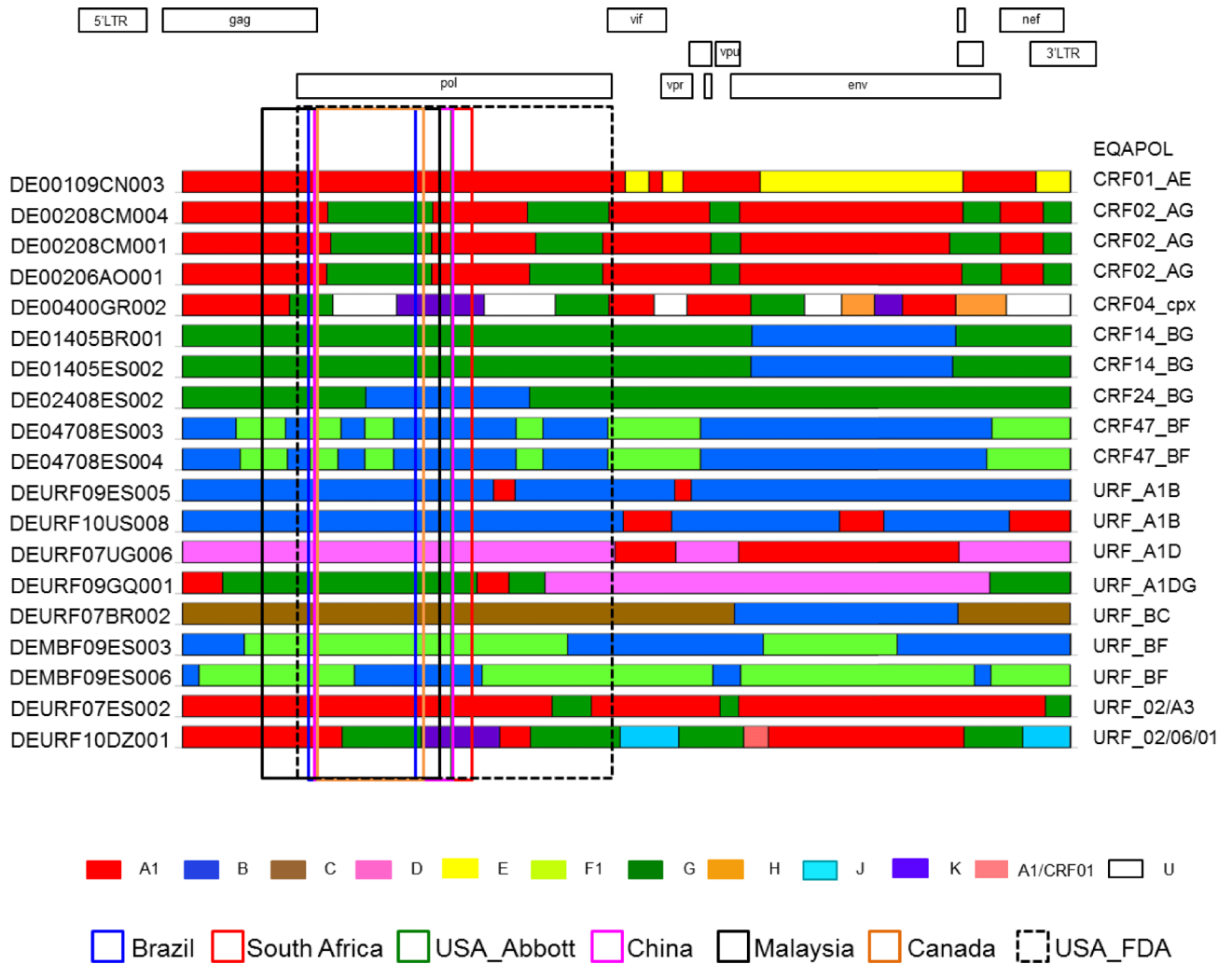


Fig 3. Recombination patterns in the *pol* gene and the near full length genome. The recombination patterns of near full-length sequences determined at EQAPOL are shown and indicated on the far right. The *pol* gene region amplified at each site is indicated by the colored box. Subtypes are color coded (shown at the bottom).

doi:10.1371/journal.pone.0157340.g003

the South Africa and Malaysia sites. DEURF10DZ001 is a recombinant between two CRFs, hence making the subtyping very challenging when only partial *pol* gene was sequenced (Fig 3). Five recombinants were reported as concordant with near full-length EQAPOL reference sequences by the China site, although no recombination breakpoints were present in the *pol* region sequenced for this analysis (Table 1; underlined).

For CRF01 and CRF02, the unique recombination breakpoints and the pure subtype regions that have been derived from the same common ancestor can clearly distinguish them from other viruses in phylogenetic analysis. The same was observed for five sequences of CRF24_BG, CRF14_BG and CRF47_BF (Fig 4). The partial *pol* sequences from both CRF24_BG and CRF47_BF contain unique recombination breakpoints, resulting in independent clusters in the phylogenetic tree (Fig 4). One CRF24_BG and two CRF47_BF sequences in the EQAPOL diversity panel were derived from the same reference viruses used to define those two CRFs [34–36]. Consequently, they were correctly defined by the majority of the test sites (italic in Table 1) while remaining test sites correctly identified them as recombinants between

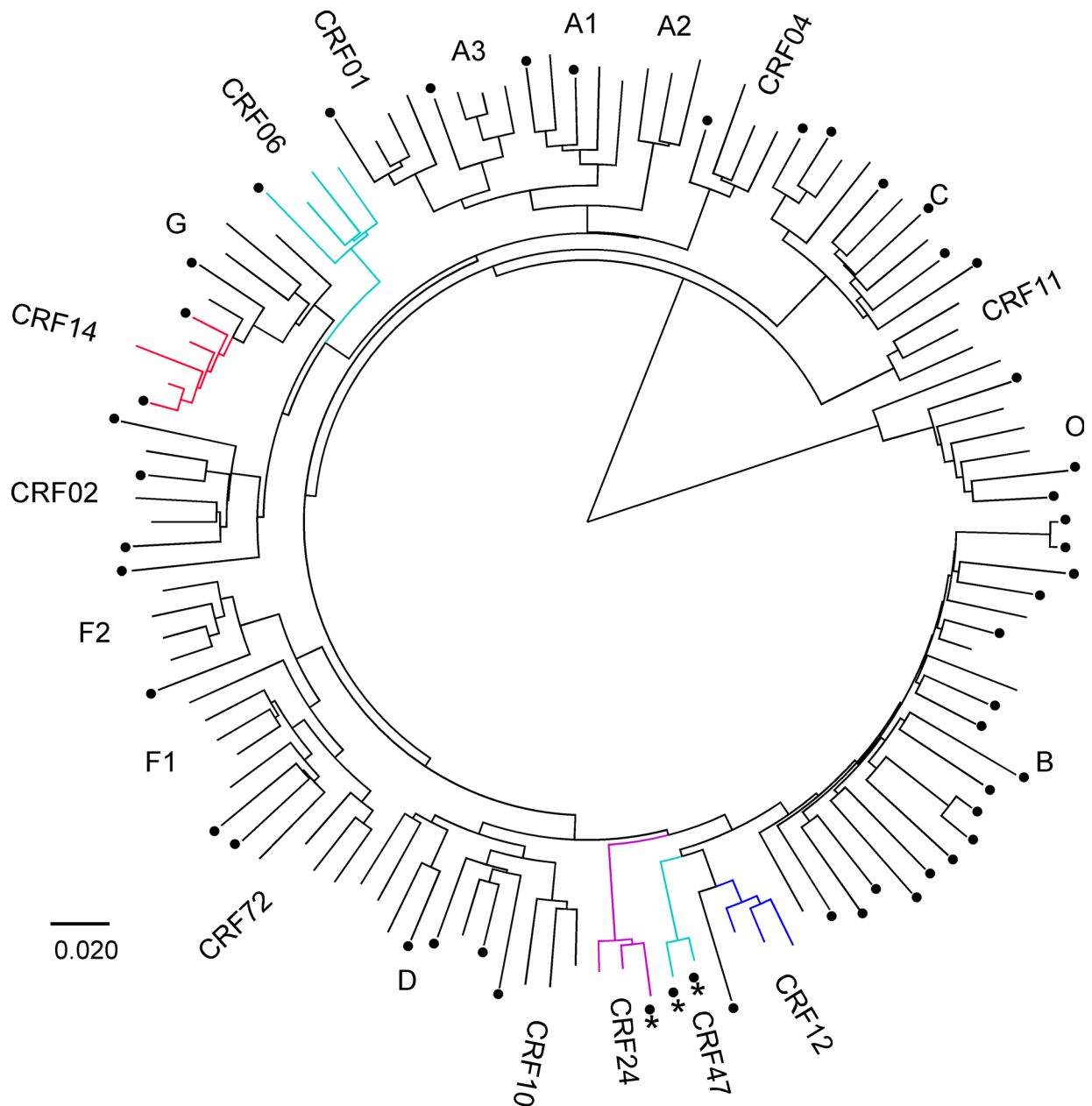


Fig 4. Phylogenetic clusters of CRF sequences. Partial *pol* gene sequences were aligned with all subtype references from the HIV sequence database. The phylogenetic tree was constructed using the neighbor-joining (NJ) method with Kimura two-parameter model. CRFs (06, 12, 14, 24 and 47) form distinct clusters and are depicted in different colors. The 50 sequences in the EQAPOL diversity panel are indicated by black dots. Asterisks indicate the EQAPOL sequences that are derived from the same strains used to define CRF24 and CRF47.

doi:10.1371/journal.pone.0157340.g004

the corresponding subtypes. The partial *pol* sequence of CRF14_BG viruses was pure subtype G without any recombinant breakpoints. However, phylogenetic analysis with reference sequences showed that they formed a tight cluster within the main subtype G clade (Fig 4). Thus, phylogenetic analysis with CRF14_BG reference sequences could identify this CRF in the *pol* region. Indeed, two CRF14_BG viruses were correctly identified at the USA_Abbott, China and Malaysia sites (Table 1; italic letters), while other sites correctly reported them as

subtype G. These results showed that although the small partial *pol* sequences were generally reliable for subtyping pure subtypes, it was very challenging to accurately classify viruses that contain recombinant breakpoints in this region.

Analysis of the partial *pol* gene sequences by next generation sequencing

The cost and throughput of next-generation sequencing (NGS) methods have significantly improved over the last few years, and barcoded PCR amplicons derived from different samples can be batched for NGS to further reduce costs and increase the throughput [37, 38]. This makes NGS a potentially attractive alternative approach to characterize HIV-1 sequences. To investigate this possibility, the same partial *pol* PCR amplicons used for conventional Sanger sequencing were subjected to NGS analysis at the Canada site. Phylogenetic analysis of NGS consensus, generated with mixed based identification threshold at 20%, and Sanger sequences in conjunction with EQAPOL sequences from each virus showed that 37 of 47 group M viruses clustered tightly together. Since both Sanger and NGS consensus sequences reflect the dominant genotypes of HIV-1 quasispecies present in the samples, ambiguous bases were observed for both methods in nine viruses. These ambiguous bases resulted in longer branches for some viruses, or even a separate branch (DEURF07ES002) and an independent branch (DEURF07BR002) (Fig 5). A large number of ambiguous bases were also reported in DEURF07BR002 by the Brazil site and it, together with both Sanger and NGS sequences from the Canada site, formed an independent branch (Fig 5). Comparing those sequences to the EQAPOL reference sequences showed that nearly all unambiguous sites were identical to the EQAPOL references, confirming that they were indeed derived from DEURF07ES002 and DEURF07BR002, respectively.

Analysis of the full *pol* gene sequences by next generation sequencing

The complete *pol* gene sequence was obtained by NGS at the USA_FDA site. Among all 47 group M HIV-1 isolates, 41 (87%) were successfully amplified (Fig 1 and Table 1). No PCR products were obtained for six group M viruses and all three group O viruses. Among 37 *pol* sequences that were reported to have concordant subtyping results as EQAPOL, 10 sequences from the USA_FDA site showed different subtype results (Table 1). Four other sequences showed discordant subtyping results from the EQAPOL reference sequences. Examination of *pol* sequences also showed the presence of human DNA sequences, indels, and sequence fragments that belonged to different viruses.

All major DRMs were identified in partial *pol* sequences

Ten major DRMs were present in seven viruses (Table 2). They were all correctly identified in the partial *pol* sequences at six sites (South Africa, Brazil, China, USA_Abbott, Malaysia and Canada). Eight major DRMs were detected in 6 viruses at the USA_FDA site: Y188L in DEMB10CN002 and M46I in DEMBF09ES003 viruses were not reported. DEMB10ES003 that had major DRMs was not amplified at the USA_FDA site.

In addition to the 10 major DRMs, all sites also reported some minor DRMs although they were not originally requested. While the majority of those minor DRMs were the same as identified in EQAPOL reference sequences, six (L10, K20, L33, V75, I84 and V108) were different in eight viruses (Table 3). Examination of the site sequences revealed that differences at 5 of 6 mutation sites (K20, L33, V75, I84 and V108) were due to the mixed viral population. In all five cases, mixed bases at each position were reported by the PCR product population sequencing while only single bases were reported in the SGA sequences generated at EQAPOL.

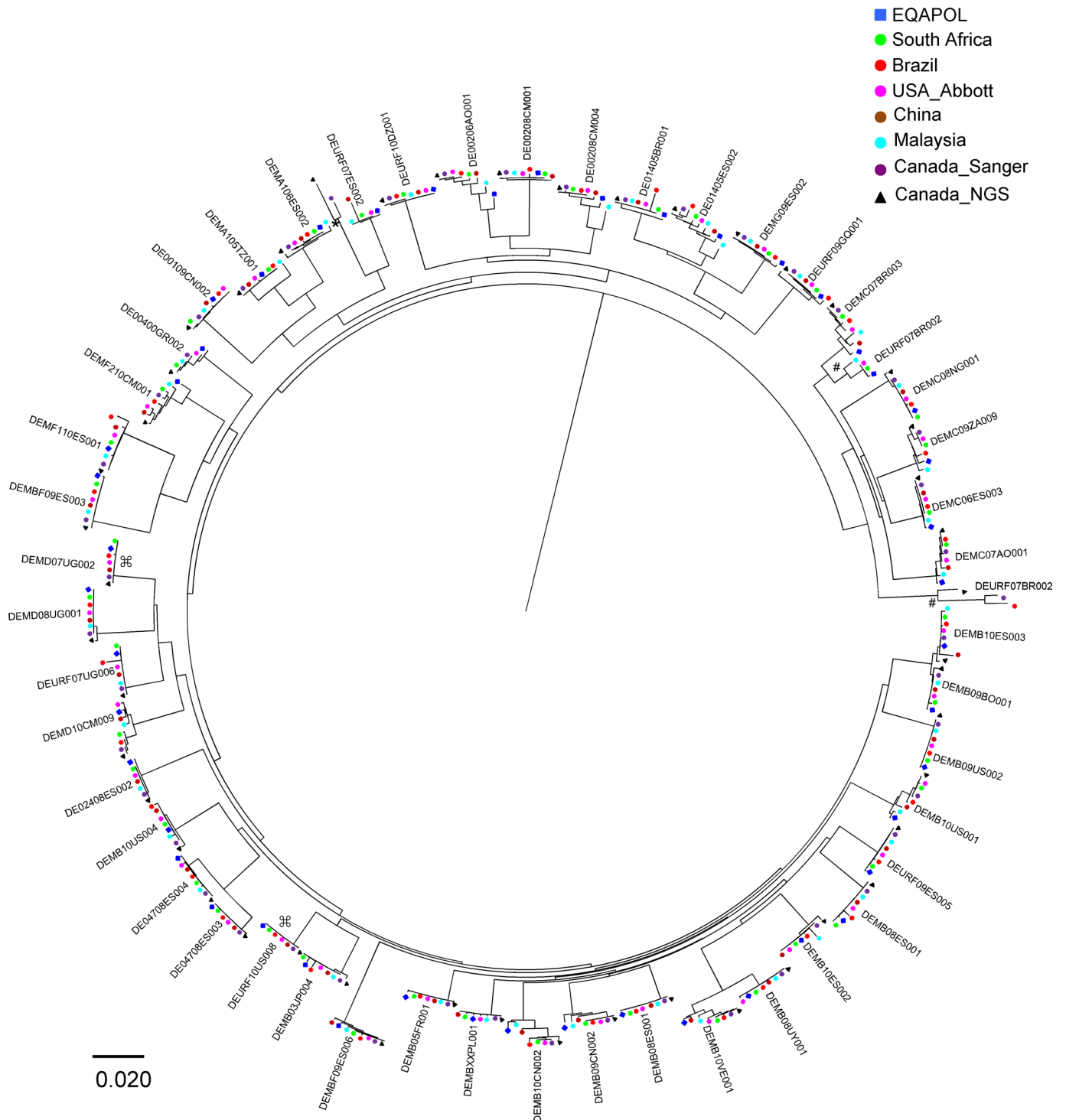


Fig 5. Phylogenetic analysis of partial *pol* gene sequences from all sites. All group M virus sequences (47) from six sites were analyzed together with the EQAPOL reference sequences. The phylogenetic tree was constructed using the neighbor-joining (NJ) method with Kimura two-parameter model. Sequences from the USA_FDA site were not included due to divergent sequence fragments. Ambiguous bases in DEURF07ES002 (*) and DEURF07BR002 (#) from Canada and Brazil sites resulted in independent branches. The short sequences (§) of DEMD07UG002 and DEURF10US008 from the Malaysia site were excluded from the analysis.

doi:10.1371/journal.pone.0157340.g005

Table 2. Major drug resistance mutations in seven viruses.

Virus ID	PI	NRTI	NNRTI
DEMB09US002			K103N
DEMBF09ES006	M46I, I54V, L76V, V82A		K103S
DEMB10CN002	L90M	T215C	Y188L
DEMB03JP004		T215D	
DEMB10ES002	L90M		
DEMB10ES003		D67N, K219Q	
DEMB09BO001		D67N, K219Q	

doi:10.1371/journal.pone.0157340.t002

(Table 3). NGS methods have the potential to more accurately detect percentages of different viral species in a quasispecies population. To further validate this, we analyzed NGS sequences from the Canada site that had an average coverage of 11,987 for the polymorphism positions. A threshold of 1% was used for low abundance variant detection and reporting (Table 3). Mixtures of amino acids at V108 in DEMC07AO001 and at L33 in DEMA105TZ001 were detected and the predominant virus species were identical to the SGA sequence obtained by EQAPOL. Mixtures of amino acids at L10 in DEMBXXPL001 and also at K20 in DE0145ES002 were detected by NGS, but the EQAPOL reference sequence represented the minority species while population sequences from other test sites represented the majority virus species. No mixed populations were detected at V75 in DEMBF09ES006 and I84 in DEMF110ES001 by NGS or Sanger sequencing at any test site except Brazil, suggesting the polymorphisms at both positions may be due to *in vitro* PCR artifacts. A mixture of V and L at the L10 position in DEMBF09ES006 was detected by NGS while all other test sites only detected valine, although isoleucine was present in the EQAPOL reference sequence. Notably, the 10L phenotype in this sample was detected by NGS at frequency of 0.4%, which is considered un-reportable as 1% threshold being applied. These results confirmed that NGS analysis of amplified partial *pol* gene products could more accurately determine mixed HIV-1 populations than Sanger population sequencing, while a SGA sequence occasionally did not represent the major viral species in the HIV-1 quasispecies population of some virus stocks.

Analysis of complete *pol* sequences generated by NGS at the FDA site showed different patterns of DRMs. In addition to D67N and K219Q in DEMB09BO001 and K103N in DEMB09US002 that were detected by all sites, a minor A71I mutation was also reported for both viruses, although examination of the sequences did not show the A71I mutation in either sequence. Among five major DRMs in DEMBF09ES006 (Table 2), four were reported while the M46I mutation was not detected in the sequence due to A to G substitution in the sequence from the site.

Discussion

Determination of subtypes and recombinants of HIV-1 and monitoring of DRMs are routinely performed in many research and national reference labs globally [14, 39, 40]. The partial *pol* gene is an advantageous target region for molecular surveillance since the sequences can be used for both subtyping and drug resistance analysis [41]. However, the target region, RT-PCR conditions utilized, genetic analysis tools and choice of reference sequences vary from site to site. Our analysis of results from seven international sites showed that the majority of 50 genetically diverse HIV-1 isolates were successfully amplified and correctly subtyped. However, optimization and standardization of HIV-1 genotyping methods would improve success rates of RT-PCR amplification and achieve more accurate subtype assignment.

Table 3. Summary of discordant minor DRMs across test sites relative to the EQAPOL sequences.

Sample ID	Site	Mutations					
		PI			NRTI		
		L10I/V	K20R/I/V/L	L33I	I84I/M	V75I	V108I
DEMA105TZ001	EQAPOL	I	K	L			
	Brazil	I	K	L			
	South Africa	I	K	I/L [#]			
	USA_Abbott	I	K	I			
	China	I	K	L			
	Malaysia	V	R	L			
	Canada (Sanger)	I	K	I			
	Canada (NGS)	I (97.2%), L (2.8%)	K (100%)	I (19.7%), L (80.3%)			
DEMB10CN002	EQAPOL		V				
	Brazil		V/L ^{\$}				
	South Africa		V/L ^{\$}				
	USA_Abbott		V/L ^{\$}				
	China		L				
	Malaysia		L				
	Canada (Sanger)		V/L ^{\$}				
	Canada (NGS)		V (35.8%) L (64.2%)				
DEMBXXPL001	EQAPOL	L					
	Brazil	Neg PCR					
	South Africa	V					
	USA_Abbott	V					
	China	L					
	Malaysia	L					
	Canada (Sanger)	V					
	Canada (NGS)	V (65.5%), L (33.5%)					
DEMC07AO001	EQAPOL						V
	Brazil						I/V*
	South Africa						I
	USA_Abbott						V
	China						V
	Malaysia						V
	Canada (Sanger)						I/V*
	Canada (NGS)						I (23.0%), V (77.0%)
DEMF110ES001	EQAPOL				I		
	Brazil				I/M*		
	South Africa				I		
	USA_Abbott				I		
	China				I		
	Malaysia				I		
	Canada (Sanger)				I		
	Canada (NGS)				I (100%)		
DE01405ES002	EQAPOL		V				
	Brazil		I				
	South Africa		I				
	USA_Abbott		I				

(Continued)

Table 3. (Continued)

Sample ID	Site	Mutations					
		PI				NRTI	
		L10I/V	K20R/I/V/L	L33I	I84I/M	V75I	V108I
	China		I				
	Malaysia		I				
	Canada (Sanger)		I				
	Canada (NGS)		I (96.3%), V (3.7%)				
DEMBF09ES006	EQAPOL	I				V	
	Brazil	V				I/V*	
	South Africa	V				V	
	USA_Abbott	V				V	
	China	V				V	
	Malaysia	V				V	
	Canada (Sanger)	V				V	
	Canada (NGS)	V (99.6%), L(0.4%)				V (100%)	
DEURF07BR002	EQAPOL	L					
	Brazil	L					
	South Africa	L					
	USA_Abbott	L					
	China	L					
	Malaysia	F					
	Canada (Sanger)	L					
	Canada (NGS)	L (100%)					

Discordant DRMs between each site and EQAPOL sequences are indicated in bold. Percentages of mutations determined by NGS at the Canada site are indicated in parenthesis. The positions are shown relative to those in the HXB2 genome.

*: a mixture of G and A

#: a mixture of T and A

\$: a mixture of G and C

doi:10.1371/journal.pone.0157340.t003

All 47 group M viruses were amplified by RT-PCR at the South Africa, USA_Abbott, Malaysia and Canada sites, while 89.4%-94.0% of these viruses were amplified at the Brazil and China sites. Since all group M viruses were successfully amplified in a single amplicon at South Africa, USA_Abbott and Canada sites, the primers and conditions used at these sites with a 100% PCR amplification rate are recommended for global subtype and drug resistance surveillance. The same commercial kit was used at both the USA_Abbott and China sites (S1 Table). However, 100% PCR success rate was reported at the USA_Abbott site but only 84.1% at the China site. This suggests the need for optimization and standardization of operation procedures as well as standard personnel training among different sites. In this study only high viral genome copy samples (10^7 /ml) were used to ensure that the failure of amplification of HIV-1 *pol* gene or whole genome was not due to the low copy numbers in the samples. Thus, this study did not address the issues that might occur with samples that had low copy viral genomes during screening, diagnosis and follow-up. The PCR sensitivity for samples with low viral genome numbers among all test sites will be addressed in a separated study. When the complete *pol* gene sequence was analyzed by NGS, a lower success rate (87%) was observed, indicating that a larger genome amplification (~3000bp) and subsequent NGS analysis by this method may not be practical. However, adoption of newer methods for library preparation such as

HIV-SMART and metagenomics NGS sequencing will advance the field and likely supplant traditional sub-genomic PCR and standard sequencing in the future [42, 43].

While the majority of the subtype classifications based on partial *pol* gene sequences were accurate for pure subtypes, determination of recombination patterns and hence classification of CRFs and URFs were considerably more challenging. Among 168 comparisons for the pure subtype sequences in the partial *pol* gene, only 6 had discordant results compared to the EQAPOL references. When the same partial *pol* gene sequences with recombination breakpoints were analyzed, the recombinant sequence regions for each subtype were smaller, making subtype assignment more difficult. However, pure subtype sequences represented in CRFs are generally derived from one common ancestor or have evolved into closely related sequences to form tight clusters in phylogenetic analysis. In such cases, they could be accurately classified even with relatively small partial *pol* gene sequences (for example, CRF12, CRF14, CRF24 and CRF47 in Fig 4). Since the subtyping results based on partial *pol* gene sequences could not predict the complex recombinant nature of viruses in the non-sequenced regions, the recombination patterns of such recombinants could only be fully resolved by analyzing whole genome sequences.

Group O viruses, as expected, were not amplified with group M virus-specific primers. Only the Malaysia site analyzed the panel viruses with group O specific primers and successfully amplified all three group O viruses, according to their standard protocol for samples that fail to amplify with group M primers. The conventional global subtyping and drug resistance survey with group M specific primers will therefore miss viruses of other groups. While this might not significantly affect the survey results in areas where group M viruses are predominant, non-group M primers should be included in areas where epidemics may include other group viruses, especially when PCR targeting group M viruses is unsuccessful.

All 10 DRMs in seven isolates were detected by all six sites where the partial *pol* sequences were analyzed. This is most likely because these mutations were present in the predominant virus populations (>97% of NGS reads as reported by the Canada site). However, the ability to detect mixed bases at the minority DRM sites varied among samples. This indicates that if the major DRMs are present as minority populations in samples, their presence might not be reliably detected by SGA or by conventional Sanger sequencing of PCR products. However, the percentages of variants at those positions could be more accurately determined by NGS analysis of the same partial *pol* PCR amplicons. With significant reduction of cost using the primer ID technology, NGS methods could be a good alternative tool for global surveillance when the resources are available. Importantly, the higher sensitivity of NGS for detection of minor DRMs relative to Sanger population sequencing will allow surveillance of minority DRMs in the population.

More diverse HIV-1 isolates continue to be characterized by EQAPOL, including other HIV-1 group viruses and HIV-2 isolates. A more comprehensive panel that fully represents HIV-1 groups, subtypes, major CRFs as well as HIV-2 can be assembled in the future to more rigorously evaluate accuracy and efficiency of global surveillance tools. NGS analysis of partial *pol* sequences showed promise in subtyping and detection of DRMs, whereas analysis of larger sequences by NGS demonstrated poorer performance. With continuous advances and optimization, it is likely that NGS analysis of whole HIV-1 genome from clinical samples will be refined to more accurately assign subtypes and detect DRMs for all classes of antiretroviral drugs including those that do not target the protease and reverse transcriptase genes.

Supporting Information

S1 Table. Summary of the PCR conditions and sequence analysis methods in all seven international sites.

(PDF)

Acknowledgments

We thank Jim Lane and Cesar Boggiano for their supports for the project and Brian Foley for assistance in sequence analysis. The members in the NHLBI Recipient Epidemiology Donor Evaluation Study—III (REDS-III) are M.P. Busch (Leader; Blood Systems Research Institute, USA; mbusch@bloodsystems.org), A.E. Mast (Blood Center of Wisconsin, USA), J.L. Gottschall (Blood Center of Wisconsin, USA), D.J. Triulzi (The Institute For Transfusion Medicine, USA), J.E. Kiss (The Institute For Transfusion Medicine, USA), E.L. Murphy (University of California at San Francisco, USA), E.W. Fiebig (University of California at San Francisco, USA), E.L. Snyder (Yale University USA), R.G. Cable (American Red Cross Blood Services, USA). B. Custer (Blood Systems Research Institute and University of California at San Francisco, USA), E.C. Sabino (University of Sao Paulo, Brazil), C. Ingram (South Africa National Blood Service, South Africa), H. Shan, (Johns Hopkins University, USA), J. Wang (Institute of Blood Transfusion, China), D. J. Brambilla (RTI International, USA) and M. T. Sullivan (RTI International, USA), P.J. Norris (Blood Systems Research Institute, USA), R.Y. Dodd (American Red Cross, USA), S.H. Kleinman (University of British Columbia, Canada), S.A. Glynn (National Heart, Lung, and Blood Institute, NIH, USA) and A.M. Cristman (National Heart, Lung, and Blood Institute, NIH, USA). The members in the External Quality Assurance Program Oversight Laboratory (EQAPOL) are T.N. Denny (Leader; Duke University, USA; thomas.denny@duke.edu), F. Gao (Duke University, USA), B. Hora (Duke University, USA), Y. Chen (Duke University, USA), T. Demarco (Duke University, USA), M. Carter (Duke University, USA), S. Keinonen (Duke University, USA), A. Sanchez (Duke University, USA), M.P. Busch (Blood Systems Research Institute, USA), S. Keating (Blood Systems Research Institute, USA) and M. Stone (Blood Systems Research Institute, USA).

Author Contributions

Conceived and designed the experiments: AMS TND MPB FG. Performed the experiments: BH SMK YC JL ES GH P. Swanson VR SVV PZ WZC HJ. Analyzed the data: BH YC AMS JL ES GH JH P. Swanson IH VR SVV PZ KKT WZC HJ P. Sandstrom TND MPB FG. Wrote the paper: BH MPB FG. Edited the manuscript: TND AMS SMK JH P. Swanson KKT IH VR.

References

1. Hemelaar J. Implications of HIV diversity for the HIV-1 pandemic. *The Journal of infection*. 2013; 66(5):391–400. Epub 2012/10/30. doi: [10.1016/j.jinf.2012.10.026](https://doi.org/10.1016/j.jinf.2012.10.026) PMID: [23103289](https://pubmed.ncbi.nlm.nih.gov/23103289/).
2. Plantier JC, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Leme V, et al. A new human immunodeficiency virus derived from gorillas. *Nature medicine*. 2009; 15(8):871–2. doi: [10.1038/nm.2016](https://doi.org/10.1038/nm.2016) PMID: [19648927](https://pubmed.ncbi.nlm.nih.gov/19648927/).
3. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. *Cold Spring Harbor perspectives in medicine*. 2011; 1(1):a006841. Epub 2012/01/10. doi: [10.1101/cshperspect.a006841](https://doi.org/10.1101/cshperspect.a006841) PMID: [22229120](https://pubmed.ncbi.nlm.nih.gov/22229120/); PubMed Central PMCID: PMC3234451.
4. Vallari A, Holzmayer V, Harris B, Yamaguchi J, Ngansop C, Makamche F, et al. Confirmation of putative HIV-1 group P in Cameroon. *Journal of virology*. 2011; 85(3):1403–7. doi: [10.1128/JVI.02005-10](https://doi.org/10.1128/JVI.02005-10) PMID: [21084486](https://pubmed.ncbi.nlm.nih.gov/21084486/); PubMed Central PMCID: PMC3020498.
5. Foley B LT, Apetrei C, Hahn B, Mizrachi I, Mullins J, Rambaut A, et al. HIV Sequence Compendium. 2013.
6. Vidal N, Mulanga C, Bazepeo SE, Lepira F, Delaporte E, Peeters M. Identification and molecular characterization of subtype A4 in central Africa. *AIDS research and human retroviruses*. 2006; 22(2):182–7. doi: [10.1089/aid.2006.22.182](https://doi.org/10.1089/aid.2006.22.182) PMID: [16478401](https://pubmed.ncbi.nlm.nih.gov/16478401/).
7. Hemelaar J, Gouws E, Ghys PD, Osmanov S, Isolation W-UNfH, Characterisation. Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS*. 2011; 25(5):679–89. Epub 2011/02/08. doi: [10.1097/QAD.0b013e328342ff93](https://doi.org/10.1097/QAD.0b013e328342ff93) PMID: [21297424](https://pubmed.ncbi.nlm.nih.gov/21297424/); PubMed Central PMCID: PMC3755761.

8. Aldrich C, Hemelaar J. Global HIV-1 diversity surveillance. *Trends in molecular medicine*. 2012; 18(12):691–4. Epub 2012/07/17. doi: [10.1016/j.molmed.2012.06.004](https://doi.org/10.1016/j.molmed.2012.06.004) PMID: [22796205](https://pubmed.ncbi.nlm.nih.gov/22796205/).
9. Tatem AJ, Hemelaar J, Gray RR, Salemi M. Spatial accessibility and the spread of HIV-1 subtypes and recombinants. *Aids*. 2012; 26(18):2351–60. Epub 2012/09/07. doi: [10.1097/QAD.0b013e328359a904](https://doi.org/10.1097/QAD.0b013e328359a904) PMID: [22951637](https://pubmed.ncbi.nlm.nih.gov/22951637/).
10. Peeters M, Aghokeng AF, Delaporte E. Genetic diversity among human immunodeficiency virus-1 non-B subtypes in viral load and drug resistance assays. *Clinical microbiology and infection*. 2010; 16(10):1525–31. Epub 2010/07/24. doi: [10.1111/j.1469-0691.2010.03300.x](https://doi.org/10.1111/j.1469-0691.2010.03300.x) PMID: [20649800](https://pubmed.ncbi.nlm.nih.gov/20649800/).
11. Nair SV, Kim HC, Fortunko J, Foote T, Peling T, Tran C, et al. Aptima HIV-1 Quant Dx-A fully automated assay for both diagnosis and quantification of HIV-1. *Journal of clinical virology*. 2016; 77:46–54. doi: [10.1016/j.jcv.2016.02.002](https://doi.org/10.1016/j.jcv.2016.02.002) PMID: [26896710](https://pubmed.ncbi.nlm.nih.gov/26896710/).
12. Mor O, Gozlan Y, Wax M, Mileguir F, Rakovsky A, Noy B, et al. Evaluation of the RealTime HIV-1, Xpert HIV-1, and Aptima HIV-1 Quant Dx Assays in Comparison to the NucliSens EasyQ HIV-1 v2.0 Assay for Quantification of HIV-1 Viral Load. *Journal of clinical microbiology*. 2015; 53(11):3458–65. doi: [10.1128/JCM.01806-15](https://doi.org/10.1128/JCM.01806-15) PMID: [26292298](https://pubmed.ncbi.nlm.nih.gov/26292298/); PubMed Central PMCID: [PMC4609691](https://pubmed.ncbi.nlm.nih.gov/PMC4609691/).
13. Bourlet T, Signori-Schmuck A, Roche L, Icard V, Saoudin H, Trabaud MA, et al. HIV-1 load comparison using four commercial real-time assays. *Journal of clinical microbiology*. 2011; 49(1):292–7. doi: [10.1128/JCM.01688-10](https://doi.org/10.1128/JCM.01688-10) PMID: [21068276](https://pubmed.ncbi.nlm.nih.gov/21068276/); PubMed Central PMCID: [PMC3020484](https://pubmed.ncbi.nlm.nih.gov/PMC3020484/).
14. Kantor R. Impact of HIV-1 *pol* diversity on drug resistance and its clinical implications. *Current opinion in infectious diseases*. 2006; 19(6):594–606. doi: [10.1097/QCO.0b013e3280109122](https://doi.org/10.1097/QCO.0b013e3280109122) PMID: [17075337](https://pubmed.ncbi.nlm.nih.gov/17075337/).
15. Santos AF, Soares MA. HIV Genetic Diversity and Drug Resistance. *Viruses*. 2010; 2(2):503–31. doi: [10.3390/v2020503](https://doi.org/10.3390/v2020503) PMID: [21994646](https://pubmed.ncbi.nlm.nih.gov/21994646/); PubMed Central PMCID: [PMC3185604](https://pubmed.ncbi.nlm.nih.gov/PMC3185604/).
16. Paraskevis D, Touloumi G, Bakoyannis G, Papanizos V, Lazanas M, Gargalianos P, et al. Effect of HIV type 1 subtype on virological and immunological response to combination antiretroviral therapy: evidence for a more rapid viral suppression for subtype A than subtype B-infected Greek individuals. *AIDS research and human retroviruses*. 2013; 29(3):461–9. doi: [10.1089/AID.2012.0143](https://doi.org/10.1089/AID.2012.0143) PMID: [23034083](https://pubmed.ncbi.nlm.nih.gov/23034083/).
17. Singh K, Flores JA, Kirby KA, Neogi U, Sonnerborg A, Hachiya A, et al. Drug resistance in non-B subtype HIV-1: impact of HIV-1 reverse transcriptase inhibitors. *Viruses*. 2014; 6(9):3535–62. doi: [10.3390/v6093535](https://doi.org/10.3390/v6093535) PMID: [25254383](https://pubmed.ncbi.nlm.nih.gov/25254383/); PubMed Central PMCID: [PMC4189038](https://pubmed.ncbi.nlm.nih.gov/PMC4189038/).
18. Sutherland KA, Ghosn J, Gregson J, Mbisa JL, Chaix ML, Cohen Codar I, et al. HIV-1 subtype influences susceptibility and response to monotherapy with the protease inhibitor lopinavir/ritonavir. *The Journal of antimicrobial chemotherapy*. 2015; 70(1):243–8. doi: [10.1093/jac/dku365](https://doi.org/10.1093/jac/dku365) PMID: [25228587](https://pubmed.ncbi.nlm.nih.gov/25228587/); PubMed Central PMCID: [PMC4267506](https://pubmed.ncbi.nlm.nih.gov/PMC4267506/).
19. Geretti AM, Harrison L, Green H, Sabin C, Hill T, Fearnhill E, et al. Effect of HIV-1 subtype on virologic and immunologic response to starting highly active antiretroviral therapy. *Clinical infectious diseases*. 2009; 48(9):1296–305. doi: [10.1086/598502](https://doi.org/10.1086/598502) PMID: [19331585](https://pubmed.ncbi.nlm.nih.gov/19331585/).
20. Klein MB, Young J, Dunn D, Ledergerber B, Sabin C, Cozzi-Lepri A, et al. The effects of HIV-1 subtype and ethnicity on the rate of CD4 cell count decline in patients naive to antiretroviral therapy: a Canadian-European collaborative retrospective cohort study. *CMAJ open*. 2014; 2(4):E318–29. doi: [10.9778/cmajo.20140017](https://doi.org/10.9778/cmajo.20140017) PMID: [25485259](https://pubmed.ncbi.nlm.nih.gov/25485259/); PubMed Central PMCID: [PMC4251518](https://pubmed.ncbi.nlm.nih.gov/PMC4251518/).
21. Touloumi G, Pantazis N, Chaix ML, Bucher HC, Zangerle R, Kran AM, et al. Virologic and immunologic response to cART by HIV-1 subtype in the CASCADE collaboration. *PLoS one*. 2013; 8(7):e71174. doi: [10.1371/journal.pone.0071174](https://doi.org/10.1371/journal.pone.0071174) PMID: [23936260](https://pubmed.ncbi.nlm.nih.gov/23936260/); PubMed Central PMCID: [PMC3728088](https://pubmed.ncbi.nlm.nih.gov/PMC3728088/).
22. 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. doi: [10.1371/journal.pone.0004724](https://doi.org/10.1371/journal.pone.0004724) PMID: [19266092](https://pubmed.ncbi.nlm.nih.gov/19266092/); PubMed Central PMCID: [PMC2648874](https://pubmed.ncbi.nlm.nih.gov/PMC2648874/).
23. 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. doi: [10.1371/journal.pmed.1001810](https://doi.org/10.1371/journal.pmed.1001810) PMID: [25849352](https://pubmed.ncbi.nlm.nih.gov/25849352/); PubMed Central PMCID: [PMC4388826](https://pubmed.ncbi.nlm.nih.gov/PMC4388826/).
24. Kleinman S, Busch MP, Murphy EL, Shan H, Ness P, Glynn SA, et al. The National Heart, Lung, and Blood Institute Recipient Epidemiology and Donor Evaluation Study (REDS-III): a research program striving to improve blood donor and transfusion recipient outcomes. *Transfusion*. 2014; 54(3 Pt 2):942–55. doi: [10.1111/trf.12468](https://doi.org/10.1111/trf.12468) PMID: [24188564](https://pubmed.ncbi.nlm.nih.gov/24188564/); PubMed Central PMCID: [PMC4383641](https://pubmed.ncbi.nlm.nih.gov/PMC4383641/).
25. Alencar CS, Sabino EC, Carvalho SM, Leao SC, Carneiro-Proietti AB, Capuani L, et al. HIV genotypes and primary drug resistance among HIV-seropositive blood donors in Brazil: role of infected blood donors as sentinel populations for molecular surveillance of HIV. *Journal of acquired immune deficiency syndromes*. 2013; 63(3):387–92. doi: [10.1097/QAI.0b013e31828ff979](https://doi.org/10.1097/QAI.0b013e31828ff979) PMID: [23507660](https://pubmed.ncbi.nlm.nih.gov/23507660/); PubMed Central PMCID: [PMC4136640](https://pubmed.ncbi.nlm.nih.gov/PMC4136640/).

26. Pessoa R, Watanabe JT, Calabria P, Alencar CS, Loureiro P, Lopes ME, et al. Enhanced detection of viral diversity using partial and near full-length genomes of human immunodeficiency virus Type 1 provirus deep sequencing data from recently infected donors at four blood centers in Brazil. *Transfusion*. 2015; 55(5):980–90. doi: [10.1111/trf.12936](https://doi.org/10.1111/trf.12936) PMID: [25413141](https://pubmed.ncbi.nlm.nih.gov/25413141/); PubMed Central PMCID: PMC4428985.
27. Delwart E, Slikas E, Stramer SL, Kamel H, Kessler D, Krysztof D, et al. Genetic diversity of recently acquired and prevalent HIV, hepatitis B virus, and hepatitis C virus infections in US blood donors. *The Journal of infectious diseases*. 2012; 205(6):875–85. doi: [10.1093/infdis/jir862](https://doi.org/10.1093/infdis/jir862) PMID: [22293432](https://pubmed.ncbi.nlm.nih.gov/22293432/); PubMed Central PMCID: PMC3282564.
28. Chudy M, Weber-Schehl M, Pichl L, Jork C, Kress J, Heiden M, et al. Blood screening nucleic acid amplification tests for human immunodeficiency virus Type 1 may require two different amplification targets. *Transfusion*. 2012; 52(2):431–9. doi: [10.1111/j.1537-2995.2011.03281.x](https://doi.org/10.1111/j.1537-2995.2011.03281.x) PMID: [21810100](https://pubmed.ncbi.nlm.nih.gov/21810100/).
29. Müller B, Nübling CM, Kress J, Roth WK, De Zolt S, Pichl L. How safe is safe: new human immunodeficiency virus Type 1 variants missed by nucleic acid testing. *Transfusion*. 2013; 53(10pt2):2422–30. doi: [10.1111/trf.12298](https://doi.org/10.1111/trf.12298)
30. Sanchez AM, DeMarco CT, Hora B, Keinonen S, Chen Y, Brinkley C, et al. Development of a contemporary globally diverse HIV viral panel by the EQAPOL program. *Journal of Immunological Methods*. 2014; 409(0):117–30. <http://dx.doi.org/10.1016/j.jim.2014.01.004>.
31. 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. doi: [10.1093/nar/gkp371](https://doi.org/10.1093/nar/gkp371) PMID: [19443440](https://pubmed.ncbi.nlm.nih.gov/19443440/); PubMed Central PMCID: PMC2703979.
32. Ji H TT, Enns E, Gauthier MK., Capina R, Liang B, Van Domselaar G, Sandstrom P, and Brooks J. Establishment of an Illumina MiSeq-based HIV drug resistance testing platform. Presented at: the 8th International AIDS Society Conference on HIV Pathogenesis, Treatment and Prevention, Vancouver, Canada. 2015;IAS(July, 2015.).
33. Galtier N, Gouy M, Gautier C. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Computer applications in the biosciences: CABIOS*. 1996; 12(6):543–8. PMID: [9021275](https://pubmed.ncbi.nlm.nih.gov/9021275/).
34. Perez L, Thomson MM, Bleda MJ, Aragonés C, González Z, Perez J, et al. HIV Type 1 molecular epidemiology in Cuba: high genetic diversity, frequent mosaicism, and recent expansion of BG intersubtype recombinant forms. *AIDS research and human retroviruses*. 2006; 22(8):724–33. doi: [10.1089/aid.2006.22.724](https://doi.org/10.1089/aid.2006.22.724) PMID: [16910827](https://pubmed.ncbi.nlm.nih.gov/16910827/).
35. Sierra M, Thomson MM, Posada D, Perez L, Aragonés C, González Z, et al. Identification of 3 phylogenetically related HIV-1 BG intersubtype circulating recombinant forms in Cuba. *Journal of acquired immune deficiency syndromes*. 2007; 45(2):151–60. doi: [10.1097/QAI.0b013e318046ea47](https://doi.org/10.1097/QAI.0b013e318046ea47) PMID: [17356464](https://pubmed.ncbi.nlm.nih.gov/17356464/).
36. Fernández-García A, Pérez-Alvarez L, Cuevas MT, Delgado E, Muñoz-Nieto M, Cilla G, et al. Identification of a new HIV type 1 circulating BF intersubtype recombinant form (CRF47_BF) in Spain. *AIDS research and human retroviruses*. 2010; 26(7):827–32. doi: [10.1089/aid.2009.0311](https://doi.org/10.1089/aid.2009.0311) PMID: [20618102](https://pubmed.ncbi.nlm.nih.gov/20618102/).
37. Shapland EB, Holmes V, Reeves CD, Sorokin E, Durot M, Platt D, et al. Low-Cost, High-Throughput Sequencing of DNA Assemblies Using a Highly Multiplexed Nextera Process. *ACS synthetic biology*. 2015; 4(7):860–6. doi: [10.1021/sb500362n](https://doi.org/10.1021/sb500362n) PMID: [25913499](https://pubmed.ncbi.nlm.nih.gov/25913499/).
38. Rohland N, Reich D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome research*. 2012; 22(5):939–46. doi: [10.1101/gr.128124.111](https://doi.org/10.1101/gr.128124.111) PMID: [22267522](https://pubmed.ncbi.nlm.nih.gov/22267522/); PubMed Central PMCID: PMC3337438.
39. Martínez-Cajas JL, Pai NP, Klein MB, Wainberg MA. Differences in resistance mutations among HIV-1 non-subtype B infections: a systematic review of evidence (1996–2008). *Journal of the International AIDS Society*. 2009; 12:11. doi: [10.1186/1758-2652-12-11](https://doi.org/10.1186/1758-2652-12-11) PMID: [19566959](https://pubmed.ncbi.nlm.nih.gov/19566959/); PubMed Central PMCID: PMC2713201.
40. Martínez-Cajas JL, Pant-Pai N, Klein MB, Wainberg MA. Role of genetic diversity amongst HIV-1 non-B subtypes in drug resistance: a systematic review of virologic and biochemical evidence. *AIDS reviews*. 2008; 10(4):212–23. PMID: [19092977](https://pubmed.ncbi.nlm.nih.gov/19092977/).
41. Pasquier C, Millot N, Njouom R, Sandres K, Cazabat M, Puel J, et al. HIV-1 subtyping using phylogenetic analysis of *pol* gene sequences. *Journal of virological methods*. 2001; 94(1–2):45–54. PMID: [11337039](https://pubmed.ncbi.nlm.nih.gov/11337039/).
42. Luk KC, Berg MG, Naccache SN, Kabre B, Federman S, Mbanya D, et al. Utility of Metagenomic Next-Generation Sequencing for Characterization of HIV and Human Pegivirus Diversity. *PloS one*. 2015; 10(11):e0141723. doi: [10.1371/journal.pone.0141723](https://doi.org/10.1371/journal.pone.0141723) PMID: [26599538](https://pubmed.ncbi.nlm.nih.gov/26599538/); PubMed Central PMCID: PMC4658132.
43. Berg MG, Yamaguchi J, Alessandri-Gradt E, Tell RW, Plantier JC, Brennan CA. A Pan-HIV Strategy for Complete Genome Sequencing. *Journal of clinical microbiology*. 2016; 54(4):868–82. doi: [10.1128/JCM.02479-15](https://doi.org/10.1128/JCM.02479-15) PMID: [26699702](https://pubmed.ncbi.nlm.nih.gov/26699702/); PubMed Central PMCID: PMC4809961.