Annals of Epidemiology 33 (2019) 84–88

Contents lists available at ScienceDirect

Annals of Epidemiology

Brief communication

Diversity and characterization of HIV-1 subtypes in the United States, 2008–2016

Richard L. Kline, MS [a], Neeraja Saduvala, MS [b], Tianchi Zhang, MPH [b], Alexandra M. Oster, MD [a,*]

[a] Division of HIV/AIDS Prevention, Centers for Disease Control and Prevention, Atlanta, GA
[b] ICF International, Atlanta, GA

ARTICLE INFO

Article history:
Received 28 December 2018
Accepted 28 February 2019
Available online 6 March 2019

Keywords:
Molecular epidemiology
HIV-1
Risk factors
Genetic diversity
Public health surveillance

ABSTRACT

Purpose: This article describes subtype diversity among diagnosed HIV-1 infections in the United States during 2008–2016 by demographic or risk group and over time.
Methods: HIV-1 polymerase sequences reported to the National HIV Surveillance System for persons in 17 U.S. states with HIV infection diagnosed during 2008–2016 were subtyped using COMET, an automated subtyping tool, and National HIV Surveillance System demographic data were analyzed.
Results: Subtype B was identified in 93.6% of 121,793 reported sequences. The most common non-B subtypes and circulating recombinant forms (CRFs) were C, CRF02_AG, A, CRF01_AE, and G. Elevated percentages of non-B subtypes or CRFs were found in persons who were female, aged less than 13 years at diagnosis, Asian, or had transmission attributable to heterosexual contact (females only) or perinatal exposure. Foreign-born persons had a higher percentage of non-B subtypes. The prevalence of non-B subtypes and CRFs increased from 5.0% in 2008 to 8.5% in 2016; among specific subtypes and CRFs, subtype G and CRF01_AE increased.
Conclusions: Subtype B remains the predominant strain in the United States. Non-B subtypes and CRFs were not widespread, but diversity and numbers increased from 2008 through 2016, which could have consequences for clinical management, diagnostic testing, and vaccine development.
Published by Elsevier Inc.

Background

HIV-1 subtype diversity and distribution reflect the complexity of the molecular epidemiology of the virus. Distinct characteristics of HIV-1 subtypes are believed to impact transmission, disease progression, clinical management, diagnostic test results, and vaccine development efforts [1,2]. HIV-1 strains have been phylogenetically classified into four groups (M, N, O, P), which originated from distinct cross-species transmissions to humans. Group M, the cause of the HIV pandemic, has been further classified into nine subtypes (A, B, C, D, F, G, H, J, K) [2,3]. Due to recombination occurring in the setting of co-infection with HIV strains from two different subtypes, intersubtype recombinants have evolved. These recombinants are called circulating recombinant forms (CRFs) if found in three or more epidemiologically unlinked patients and unique recombinant forms otherwise. By 2007, recombinant subtypes were estimated to account for at least 20% of HIV-1 infections worldwide [4]. At this time, approximately 96 CRFs are recognized [5]; and the list will likely continue to grow.

In North America and Central and Western Europe, HIV-1 infections are predominantly subtype B [4]. However, rising prevalence of non-B subtypes is being reported [6–10]. Abecasis et al. analyzed sequence data from persons with diagnosed HIV in Europe from 2003 to 2005 and found that although most subtype trends were stable, there were increases in subtypes F and CRF02_AG, and subtype prevalence was compartmentalized and varied by region. They found that continent and country of origin, country of sampling, and gender were associated with subtype prevalence [10]. Beloukas et al. analyzed more recent data from the Los Alamos HIV Sequence Database and determined that prevalence of non-B subtypes in Central and Western Europe was increasing, varied by individual

The authors report no conflicts of interest.

The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

* Corresponding author. Division of HIV/AIDS Prevention, Centers for Disease Control and Prevention, 1600 Clifton Rd NE, MS E-47, Atlanta, GA 30329. Tel.: +1 404-639-6141.
E-mail address: aoster@cdc.gov (A.M. Oster).

https://doi.org/10.1016/j.annepidem.2019.02.010
1047-2797/Published by Elsevier Inc.

Table 1
Subtype assignment (subtype B vs. non-B), by selected characteristics, 17 U.S. states, 2008–2016

Characteristic	Total		Subtypes/CRF			
			B		Non-B*	
	n	%	*n*	%	*n*	%
Total	121,793	100.0	114,034	93.6	7759	6.4
Sex						
Male	98,162	80.6	93,459	95.2	4703	4.8
Female	23,631	19.4	20,575	87.1	3056	12.9
Age at diagnosis (y)						
<13	380	0.3	234	61.6	146	38.4
13–19	6682	5.5	6417	96.0	265	4.0
20–29	44,494	36.5	42,329	95.1	2165	4.9
30–39	29,839	24.5	27,483	92.1	2356	7.9
40–49	23,266	19.1	21,608	92.9	1658	7.1
50–59	13,035	10.7	12,194	93.5	841	6.5
60+	4097	3.4	3769	92.0	328	8.0
Race/ethnicity						
American Indian/Alaska Native	366	0.3	359	98.1	7	1.9
Asian	2394	2.0	1608	67.2	786	32.8
Black/African American	51,595	42.4	47,511	92.1	4084	7.9
Hispanic/Latino	34,391	28.2	32,981	95.9	1410	4.1
Native Hawaiian/Other Pacific Islander	99	0.1	90	90.9	9	9.1
White	27,806	22.8	26,737	96.2	1069	3.8
Multiple races	5142	4.2	4748	92.3	394	7.7
Unknown	0	0.0	0	0.0	0	0.0
Transmission category						
Male-to-male sexual contact	79,296	65.1	75,966	95.8	3330	4.2
Male-to-male sexual contact and injection drug use	5026	4.1	4888	97.2	138	2.8
Injection drug use—men	4173	3.4	3973	95.2	201	4.8
Injection drug use—women	3523	2.9	3286	93.3	237	6.7
High-risk heterosexual contact—men	9389	7.7	8449	90.0	941	10.0
High-risk heterosexual contact—women	19,813	16.3	17,079	86.2	2734	13.8
Perinatal	316	0.3	209	66.1	107	33.9
Other/unknown	257	0.2	185	72.0	72	28.0
Population of area of residence						
Unknown	256	0.2	236	92.2	20	7.8
Nonmetropolitan areas (<50,000)	2734	2.2	2650	96.9	84	3.1
Metropolitan Areas (50,000–499,999)	12,973	10.7	12,464	96.1	509	3.9
Metropolitan statistical areas (500,000–2,499,999)	32,546	26.7	30,875	94.9	1671	5.1
Metropolitan statistical areas (2,500,000+)	73,284	60.2	67,809	92.5	5475	7.5
State of residence at dx						
Alabama	2431	2.0	2375	97.7	56	2.3
Arizona	2263	1.9	2133	94.3	130	5.7
California	20,156	16.5	19,063	94.6	1093	5.4
Colorado	1776	1.5	1650	92.9	126	7.1
Connecticut	2019	1.7	1862	92.2	157	7.8
District of Columbia	2419	2.0	2213	91.5	206	8.5
Florida	17,332	14.2	16,371	94.5	961	5.5
Illinois	5418	4.4	5126	94.6	292	5.4
Louisiana	3744	3.1	3680	98.3	64	1.7
Maryland	5603	4.6	4990	89.1	613	10.9
Michigan	5439	4.5	5167	95.0	272	5.0
New York	20,648	17.0	18,797	91.0	1851	9.0
Oregon	975	0.8	916	93.9	59	6.1
South Carolina	4000	3.3	3929	98.2	71	1.8
Texas	21,762	17.9	20,591	94.6	1171	5.4
Virginia	2912	2.4	2617	89.9	295	10.1
Washington	2896	2.4	2554	88.2	342	11.8

* Includes non-B subtypes, CRFs, and unassigned.

country, and was linked to migration and later dispersal through transmission networks [8]. Analysis of sequence data from newly identified HIV patients (recent diagnoses and new residents) in Alberta, Canada, found that 16% of patients were infected with non-B subtypes in 2001; however, by 2010, 33% of new patients were infected with non-B subtypes [11].

Exploring subtype diversity and trends in the United States, our group previously analyzed protease and reverse transcriptase sequence data reported to the National HIV Surveillance System (NHSS) during 2006–2013. Although we found that B was the predominant subtype, prevalence of non-B infections increased during 2006–2013 and numerous demographic subgroups had a much higher prevalence of non-B subtypes [9]. However, this analysis only included diagnoses through 2013 and was limited to only seven U.S. states.

In this report, we update this previous analysis by describing the prevalence of subtypes and CRFs among diagnosed cases in 17 U.S. states during 2008–2016. We also describe subtype diversity by demographic and risk group and world region of birth. Finally, we report on changes in subtype diversity over the 9-year time frame.

Methods

HIV-1 polymerase sequences (protease and partial reverse transcriptase) generated through standard genotypic drug

resistance testing and reported to NHSS for persons with HIV-1 infection diagnosed during 2008–2016 were used for this analysis. Details about sequence collection have been previously described [9]. Sequences were linked to demographic, risk, and geographic information that was also reported to the NHSS. Written consent was not required as the data were collected for public health surveillance purposes. No personally identifiable information was reported to NHSS.

For analysis, we only included sequences from the 17 jurisdictions that reported sequencing data for at least 30% of diagnoses during the 9-year period. (Alabama, Arizona, California, Colorado, Connecticut, District of Columbia, Florida, Illinois, Louisiana, Maryland, Michigan, New York, Oregon, South Carolina, Texas, Virginia, Washington.) Sequences shorter than 500 nucleotides (<5% of all sequences) were excluded. When more than one sequence was reported for an individual (as determined using unique identifiers), analysis was limited to the sequence with the earliest sample collection date.

HIV-1 subtype or CRF was assigned using a local installation of COMET (Context-based Modeling for Expeditious Typing) v2.2, an automated and statistically based subtyping tool [12]. Each sequence was assigned to one of eight subtypes, one of 49 CRFs, or "unassigned." An "unassigned" sequence may suggest a possible unique recombinant form [12].

Prevalence of different subtypes and CRFs were analyzed. CRFs with less than 0.05% prevalence were grouped together as "other CRFs." We determined the prevalence of subtypes and CRFs (B vs. non-B) by sex, age at diagnosis, race or ethnicity, transmission category, population size of area of residence, and state of residence at diagnosis. Transmission category was hierarchically assigned as male-to-male sexual contact, injection drug use, male-to-male sexual contact and injection drug use, heterosexual contact, perinatal, or other or unknown. We also assessed the prevalence of specific subtypes and CRFs by place of birth, grouping foreign countries by world region: Africa, Asia, Caribbean, Central and South America, Europe, North America (Mexico), North America (other), and other. Finally, we examined subtype distribution trends over the 9-year analysis period by year of HIV diagnosis.

Results

During 2008–2016, there were 389,052 diagnoses of HIV-1 infection in the United States. The 17 jurisdictions included in this analysis reported 253,306 of those infections, comprising 65.1% of all HIV-1 diagnoses in the United States. Of these reported infections, 121,793 (48%) had HIV-1 polymerase sequences available and met the requirements for inclusion. As expected, subtype B was predominant, identified as the subtype in 114,034 (93.6%) of reported sequences. Non-B sequences were distributed as follows: 3421 (2.8%) were non-B pure subtypes, 2428 (2.0%) were CRFs, and 1910 (1.6%) were unassigned. The five most common non-B subtypes and CRFs were C (1604, 1.3%); CRF02_AG (1087, 0.9%); A (898, 0.7%); CRF01_AE (731, 0.6%); and G (569, 0.5%).

Overall, 7759 (6.4%) sequences were non-B subtypes or CRFs. Analysis of demographic and geographic characteristics showed substantial variation in the percentage of non-B subtypes or CRFs by sex, age at diagnosis, race or ethnicity, transmission category, and area of residence at diagnosis (Table 1). Subgroups with relatively high percentages (≥10%) of non-B subtypes were females (12.9%), those aged less than 13 years at diagnosis (38.4%), Asians (32.8%), females with HIV infection attributed to heterosexual contact (13.8%), males with HIV infection attributed to heterosexual contact (10.0%), and persons with HIV infection attributed to perinatal transmission (33.9%). Of the jurisdictions included in this analysis, the highest percentage of non-B subtypes occurred in Washington

Table 2
Subtype assignment, by world region of birth, 17 U.S. states 2008–2016

Subtype/CRF	Total		USA and dependencies		Foreign country/region																Unknown	
					Total		Africa		Asia		Caribbean		Central and South America		Europe		North America (Mexico)		North America (other)		Other			
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Total	121,793	100	81,949	100	20,137	100	2530	100	1962	100	4716	100	4548	100	700	100	5125	100	85	100	471	100	19,707	100
B	114,034	93.6	79,335	96.8	16,034	79.6	272	10.8	1169	59.6	4220	89.5	4297	94.5	581	83.0	5035	98.2	82	96.5	378	80.3	18,665	94.7
A	898	0.7	241	0.3	543	2.7	355	14.0	90	4.6	19	0.4	7	0.2	50	7.1	1	0.0	0	0.0	21	4.5	114	0.6
C	1604	1.3	421	0.5	992	4.9	745	29.4	130	6.6	34	0.7	40	0.9	7	1.0	12	0.2	2	2.4	22	4.7	191	1.0
D	221	0.2	76	0.1	122	0.6	51	2.0	1	0.1	55	1.2	7	0.2	2	0.3	3	0.1	0	0.0	3	0.6	23	0.1
F	105	0.1	25	0.0	68	0.3	16	0.6	2	0.1	13	0.3	20	0.4	13	1.9	4	0.1	0	0.0	0	0.0	12	0.1
G	569	0.5	206	0.3	287	1.4	204	8.1	7	0.4	59	1.3	7	0.2	5	0.7	2	0.0	0	0.0	3	0.6	76	0.4
H	7	0.0	2	0.0	5	0.0	5	0.2	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
J	17	0.0	6	0.0	10	0.0	10	0.4	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	1	0.0
CRF01_AE	731	0.6	215	0.3	400	2.0	9	0.4	352	17.9	2	0.0	6	0.1	3	0.4	21	0.4	1	1.2	6	1.3	116	0.6
CRF02_AG	1087	0.9	340	0.4	577	2.9	498	19.7	17	0.9	24	0.5	23	0.5	8	1.1	1	0.0	0	0.0	6	1.3	170	0.9
CRF06	90	0.1	12	0.0	60	0.3	55	2.2	0	0.0	0	0.0	1	0.0	1	0.1	0	0.0	0	0.0	3	0.6	18	0.1
CRF07	73	0.1	11	0.0	53	0.3	0	0.0	50	2.5	0	0.0	1	0.0	0	0.0	2	0.0	0	0.0	0	0.0	9	0.0
CRF12	78	0.1	39	0.0	32	0.2	0	0.0	1	0.1	2	0.0	23	0.5	1	0.1	4	0.1	0	0.0	1	0.2	7	0.0
CRF20	99	0.1	22	0.0	73	0.4	0	0.0	1	0.1	58	1.2	9	0.2	2	0.3	3	0.1	0	0.0	0	0.0	4	0.0
CRF24	173	0.1	73	0.1	88	0.4	0	0.0	3	0.2	65	1.4	12	0.3	6	0.9	1	0.0	0	0.0	1	0.2	12	0.1
Other CRFs*	97	0.1	15	0.0	73	0.4	18	0.7	7	0.4	35	0.7	8	0.2	1	0.1	0	0.0	0	0.0	4	0.8	9	0.0
Unassigned	1910	1.6	910	1.1	720	3.6	292	11.5	132	6.7	130	2.8	87	1.9	20	2.9	36	0.7	0	0.0	23	4.9	280	1.4

* All CRFs with prevalence <0.05%.

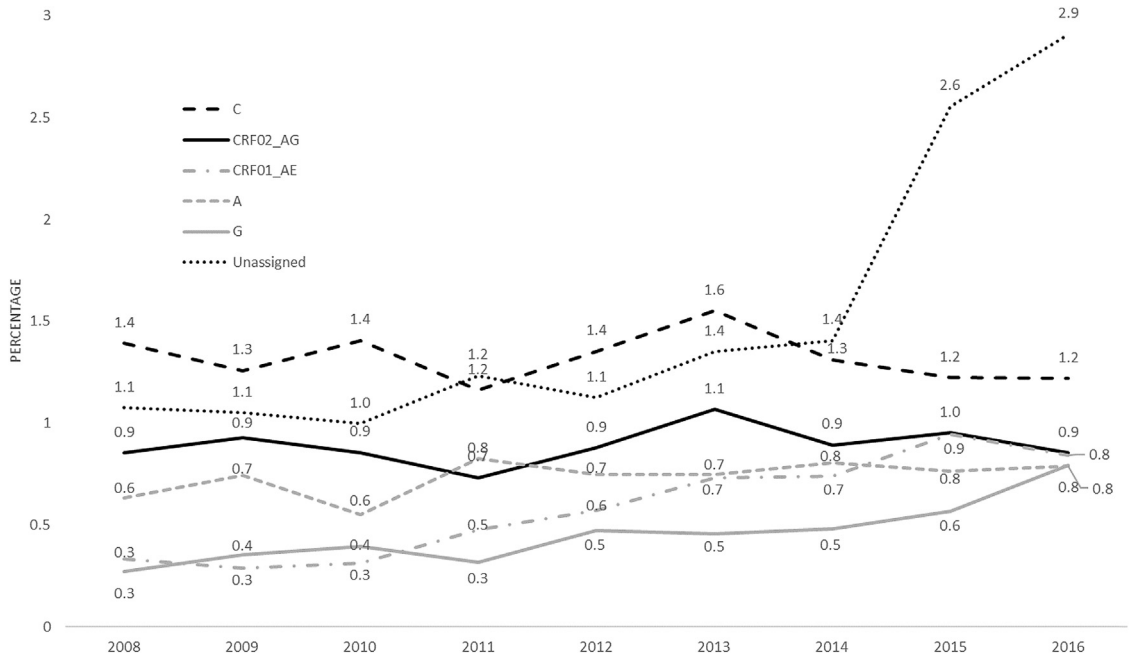

Fig. 1. Prevalence of common non-B subtypes and circulating recombinant forms, by year of diagnosis, 17 U.S. states, 2008–2016.

(11.8%), Maryland (10.9%), and Virginia (10.1%). Multivariable analysis including all variables listed in Table 1 demonstrated that all variables were significantly associated with subtype or CRF (B vs. non-B) at $P < .0001$ except sex ($P = .4$).

Non-B subtypes were present in 20.4% of foreign-born persons compared with 3.2% of U.S.-born persons (Table 2). The prevalence of non-B subtypes varied by geographical region of birth, and the identified subtypes generally matched the expected subtype distribution from the birth region. Subtype B was the dominant subtype among foreign-born persons from every geographical region except Africa. The top four most common non-B subtypes or CRFs among foreign-born persons were as follows: for persons born in Africa: C (29.4%), CRF02_AG (19.7%), A (14.0%), and G (8.1%); for persons born in Asia: CRF01_AE (17.9%), C (6.6%), A (4.6%), and CRF07 (2.5%); for persons born in Europe: A (7.1%), F (1.9%), CRF02_AG (1.1%), and C (1.0%). Persons born in the Caribbean had elevated percentages of CRF 24 (1.4%), G (1.3%), CRF 20 (1.2%), and D (1.2%). Persons born in Central and South America or elsewhere in North America (besides the United States and Mexico) had very high percentages with subtype B infection (≥94.5%).

When looking at subtype or CRF trends during 2008–2016, non-B subtypes as a group increased over time from 5.0% in 2008 to 8.5% in 2016 (data not shown). When separated into individual subtypes, the prevalence of most non-B subtypes was stable (≤100% increase or decrease); however, subtype G and CRF01_AE showed increases over time (Fig. 1). Notably, CRF20 and CRF24 became much more common later in the analysis period; two-thirds of the CRF20 (66/99) and CRF24 (119/173) sequences came from the last 3 years of the 9-year period (data not shown). The percentage of unassigned sequences also increased and may be suggestive of an increase in unique recombinant forms over the study period.

Discussion

Consistent with previous analyses of U.S. surveillance data, subtype B was the most prevalent strain found in the 17 U.S. jurisdictions included in this analysis. Although non-B subtypes were not widespread in these jurisdictions, the percentage of non-B subtypes increased from 2008 through 2016. Most non-B subtype percentages appeared steady over the study period, however, the prevalence of subtype G and CRF01_AE more than doubled, and we detected emergence of CRF20 and CRF24, which were rare previously. When subtypes were linked to demographic data, females, persons aged less than 13 years at diagnosis, Asian, female high-risk heterosexual contact, and perinatal transmission subgroups showed an increased prevalence of non-B subtypes. Again, we also found a higher prevalence of non-B subtypes among foreign-born persons that appeared to be driven by persons born in Africa.

In a local geographic region, an increasing prevalence of a non-B subtype or CRF in diagnosed HIV cases could be an indicator of rapid HIV-1 transmission. Local officials could possibly use subtype prevalence coupled with demographic data and other molecular analysis to identify ongoing HIV-1 transmission and initiate targeted prevention activities.

This analysis was limited in that only 17 jurisdictions met the completeness requirement (>30%) for this analysis. Although the contributing jurisdictions cover 65.1% of diagnoses reported to the NHSS, not all persons had sequence data available, and findings may not be applicable to the entire United States. Still, this is an increase from the seven states included in our previous analysis. Future work will be needed to understand the increase in possible unique recombinant forms.

Conclusions

Although HIV-1 nonsubtype B strains are not prevalent in the United States at present, diversity and numbers appear to be increasing, which could have important consequences for transmission concerns, disease progression, clinical management, diagnostic testing, and vaccine development. Continued monitoring of non-B subtypes and investigation of associations between subtypes and demographic subgroups will give a better understanding of HIV infection in the United States.

Acknowledgment

The authors would like to acknowledge the local and state health department staff who collect and report the HIV sequence

data, the HIV Incidence and Case Surveillance Branch staff who help maintain the NHSS and the Transmission and Molecular Epidemiology Team for guidance and suggestions.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Buonaguro L, Tornesello ML, Buonaguro FM. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. J Virol 2007;81:10209–19.
[2] Hemelaar J. Implications of HIV diversity for the HIV-1 pandemic. J Infect 2013;66:391–400.
[3] Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM. The challenge of HIV-1 subtype diversity. N Engl J Med 2008;358:1590–602.
[4] Hemelaar J, Gouws E, Ghys PD, Osmanov S. WHO-UNAIDS network for HIV isolating and characterization. Global trends in molecular epidemiology of HIV-1 during 2000-2007. AIDS 2011;25(5):679–89.
[5] HIV Circulating Recombinant Forms. https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html. Accessed March 1, 2018.
[6] Germer JJ, Wu P, Soderberg JD, Mandrekar JN, Yao JD. HIV-1 subtype diversity among clinical specimens submitted for routine antiviral drug resistance testing in the United States. Diagn Microbiol Infect Dis 2015;83:257–60.
[7] Pyne MT, Hackett Jr J, Holzmayer V, Hillyard DR. Large-scale analysis of the prevalence and geographic distribution of HIV-1 non-B variants in the United States. J Clin Microbiol 2013;51:2662–9.
[8] Beloukas A, Psarris A, Giannelou P, Kostaki E, Hatzakis A, Paraskevis D. Molecular epidemiology of HIV-1 infection in Europe: an overview. Infect Genet Evol 2016;46:180–9.
[9] Oster AM, Switzer WM, Hernandez AL, Saduvala N, Wertheim JO, Nwangwu-Ike N, et al. Increasing HIV-1 subtype diversity in seven states, United States, 2006-2013. Ann Epidemiol 2017;27:244–51.
[10] Abecasis AB, Wensing AM, Paraskevis D, Vercauteren J, Theys K, Van de Vijver DA, et al. HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. Retrovirology 2013;10:7.
[11] Siemieniuk RA, Beckthold B, Gill MJ. Increasing HIV subtype diversity and its clinical implications in a sentinel North American population. Can J Infect Dis Med Microbiol 2013;24(2):69–73.
[12] Struck D, Lawyer G. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. Nucleic Acids Res 2014;42:e144.