Annals of Epidemiology xxx (2017) 1–8

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Original article

Increasing HIV-1 subtype diversity in seven states, United States, 2006–2013

Alexandra M. Oster MD [a,*], William M. Switzer MPH [a], Angela L. Hernandez MD, MPH [a], Neeraja Saduvala MS [b], Joel O. Wertheim PhD [b,c], Ndidi Nwangwu-Ike MPH [a], M. Cheryl Ocfemia MPH [a], Ellsworth Campbell MS [a], H. Irene Hall PhD, MPH [a]

[a] Division of HIV/AIDS Prevention, Centers for Disease Control and Prevention, Atlanta, GA
[b] ICF International, Atlanta, GA
[c] Department of Medicine, University of California, San Diego

ARTICLE INFO

Article history:
Received 16 May 2016
Accepted 7 February 2017
Available online xxx

Keywords:
Molecular epidemiology
HIV-1
Subtypes
Surveillance
Risk factors
Homosexuality
Substance abuse
Intravenous

ABSTRACT

Purpose: The aim of the analysis was to explore HIV-1 subtype diversity in the United States and understand differences in prevalence of non-B subtypes and circulating recombinant forms (CRFs) between demographic/risk groups and over time.
Methods: We included HIV-1 polymerase sequences reported to the National HIV Surveillance System for HIV infections diagnosed during 2006–2013 in seven states. We assigned subtype or CRF using the automated subtyping tool COMET, assessed subtype/CRF prevalence by demographic characteristics and country of birth, and determined changes in subtype/CRF by HIV diagnosis year.
Results: Of 32,968 sequences, 30,757 (93.3%) were subtype B. The most common non-B subtypes and CRFs were C (1.6%), CRF02_AG (1.4%), A (0.6%), CRF01_AE (0.5%), and G (0.3%). Elevated percentages of non-B infections occurred among persons aged <13 years at diagnosis (40.9%), Asians (32.1%), persons born outside the United States (22.6%), and persons with infection attributable to heterosexual contact (12.0%–15.0%). Prevalence of non-B infections increased from 5.9% in 2006 to 8.5% in 2013.
Conclusions: Subtype B continues to predominate in the United States. However, the percentage of non-B infections has grown in recent years, and numerous demographic subgroups have much higher prevalence. Subgroups and areas with high prevalence of non-B infections might represent sub-epidemics meriting further investigation.

Published by Elsevier Inc.

Introduction

Understanding the diversity of HIV strains can give insight into the global spread of HIV. Moreover, HIV subtype can have implications for disease progression, effectiveness of treatment, transmissibility, vaccine development, diagnostic test, and incidence assay performance [1–15]. The HIV-1M group, which causes the vast majority of HIV-1 infections worldwide, consists of subtypes A–D, F–H, J, and K and recombinant forms formed through intersubtype recombination, which can occur when a person is simultaneously infected with HIV from distinct subtypes [16]. Depending on the extent of transmission, recombinant forms are considered circulating recombinant forms (CRFs) or unique recombinant forms (URFs) [16].

During the 1960s–1970s, a subtype B HIV-1 strain was introduced into the United States [17]; since then, the overwhelming majority of HIV-1 infections in the United States have been subtype B. However, subtype C accounts for approximately half of all global HIV infections, and the prevalence of subtypes and CRFs varies dramatically in different regions of the world [5]. Overall, 80–95% of HIV-1 infections in the United States and more than 80% of infections in western and central Europe are subtype B, whereas approximately 80% of infections in eastern Europe are subtype A, with variation between European countries [18]. In Asia, CRF01_AE

Sources of Support: The Centers for Disease Control and Prevention.
Disclaimer: The findings and conclusions in this report 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.: 404-639-6141.
E-mail address: AOster@cdc.gov (A.M. Oster).

http://dx.doi.org/10.1016/j.annepidem.2017.02.002
1047-2797/Published by Elsevier Inc.

predominates, whereas in sub-Saharan Africa, where HIV-1 originated, has the greatest diversity of HIV strains, including predominance of subtype C in Southern Africa and substantial transmission of subtypes A, G, and CRF02_AG in West Africa and A, C, and D in East Africa [5].

Although, historically, approximately 95% of HIV-1 infections in the United States have been subtype B [5], immigration and global travel have the potential to result in importation of other subtypes. Most analyses of U.S. subtype distribution have been limited to a single city or state or to particular foreign-born populations [19–26]. Two multisite studies describing the U.S. subtype distribution have been published, but these studies relied on specimens received at commercial laboratories for HIV drug resistance testing and lacked important demographic and geographic information, including country of birth, particularly important given the HIV-1 genetic diversity in other parts of the world [27,28].

In the United States, HIV-1 nucleotide sequences have been collected since 2001 as part of molecular HIV surveillance, an integral component of the National HIV Surveillance System [29]. These data are accompanied by detailed demographic and geographic information collected through HIV surveillance, allowing for analyses to understand subtype prevalence in a variety of demographic and risk groups (including foreign-born persons) in more detail. A previous analysis of subtype diversity using sequence data reported to HIV surveillance has been published, but this analysis only reported data from 2006 [30].

Here, we report the prevalence of subtypes and CRFs among persons with sequences who had HIV-1 infection diagnosed during an 8-year period in seven U.S. states. We also examine how subtype distribution varies by demographic and geographic characteristics, including location of birth, and examine temporal changes in HIV-1 subtype diversity in the United States.

Material and methods

Source of sequences and inclusion criteria

We included HIV-1 polymerase sequences (protease and reverse transcriptase) generated through routine genotypic drug resistance testing at commercial, private, and public laboratories and reported to the National HIV Surveillance System for persons with diagnosed infection during 2006–2013; details of sequence reporting have been described elsewhere [29,31]. Sequences are linked to demographic, risk, and geographic information also reported to the National HIV Surveillance System. Because the data included in this analysis were collected for the purpose of public health surveillance, consent was not required. Data reported to the National HIV Surveillance System do not include personally identifiable information such as name, address, or phone number and are protected by an assurance of confidentiality.

Because we wished to examine temporal changes in subtype diversity, we limited our analysis to the seven states with sequences available for the highest percentage of persons with newly diagnosed HIV infection in our 8-year time period (Colorado, Connecticut, Michigan, New York, South Carolina, Texas, and Washington); these states collected sequence data consistently throughout our time period of interest. Despite originating from the same population, preliminary analysis of this data set indicated that sequences shorter than 500 nucleotides yielded subtyping results that differed from sequences of $\geq$500 nucleotides, likely due to insufficient genetic signal to distinguish among similar subtypes (e.g., subtypes B and D) in the shorter sequences. We did not find differences in subtype results for sequences of moderate length (e.g., 500–800 nucleotides) compared with longer sequences. Therefore, we excluded sequences shorter than 500 nucleotides, which comprised <5% of all sequences. Because, in the absence of rarely occurring superinfection with a different subtype [32], HIV subtype does not change over the course of an infection, we limited analysis to the sequence with the earliest collection date.

Analysis

We assigned HIV-1 subtype or CRF using a local installation of the automated, statistically based subtyping tool COMET (COntext-based Modeling for Expeditious Typing) v 1.0 (available at http://comet.retrovirology.lu/) [33], which has been independently demonstrated to be an excellent subtyping tool [34]. COMET assigns each sequence to one of eight subtypes, one of 49 CRFs, or as "unassigned," indicating that the sequence is a possible unique recombinant form according to Struck et al. [33]. CRFs are named according to the order in which they were described and the subtypes from which they originate (e.g., CRF02_AG was the second CRF described and originated from subtypes A and G). CRFs originating from more than two strains are labeled "complex" (e.g., CRF11_cpx) [16]. Sequences classified as "unassigned" by COMET were also analyzed using REGA V3 and SCUEAL [13,35].

We described overall prevalence of various subtypes and CRFs. Any CRF assigned to <0.05% of sequences was collapsed into "other CRFs." We then assessed prevalence of subtype B and non-B infections (which includes subtypes other than B, CRFs, and sequences unassigned by COMET) by sex, age at diagnosis, race/ethnicity, transmission category, population of area of residence, and state of residence at diagnosis. Transmission category was hierarchically assigned as male-to-male sexual contact, male-to-male sexual contact and injection drug use, injection drug use, heterosexual contact, perinatal, or other/unknown and was imputed when missing, in accordance with standard Centers for Disease Control and Prevention methods [36,37]. We use χ^2 tests to compare prevalence of non-B infections by each characteristic. Because subtype/CRF prevalence varies geographically, we also examined the prevalence of B and non-B infections by these same characteristics after stratifying by location of birth (i.e., United States and its dependencies, foreign country, or unknown).

We further examined prevalence of specific subtypes and CRFs by location of birth, stratifying foreign countries by world region: Africa, Asia, Caribbean, Central and South America, Europe, North America (Mexico), North America (other), and other. Sample size did not permit further stratification. We then described characteristics of persons with the five most common non-B infections (C, CRF02_AG, A, CRF01_AE, and G).

Finally, we described subtype distribution by year of HIV diagnosis, graphing prevalence for non-B infections overall and also among each of the five most common non-B subtypes and CRFs.

Results

The seven states included in our analysis reported 92,428 new diagnoses of HIV infection occurring during 2006–2013. Of these, 32,968 (36%) had HIV-1 polymerase sequences meeting our inclusion criteria (median length of 1302 nucleotides) and were included in our analysis. Persons with sequences available were similar to all persons with HIV infection diagnosed during this time period in these states with respect to sex, age, race/ethnicity, transmission category, and location of birth. In all, 30,757 (93.3%) sequences were subtype B. The most common non-B strains were: C (533, 1.6%); CRF02_AG (450, 1.4%); A (213, 0.6%); CRF01_AE (152, 0.5%); and G (112, 0.3%). Subtype/CRF was unassigned by COMET for 520 (1.6%) sequences. Analyses of these 520 sequences by REGA V3 and SCUEAL classified 151 and 191 sequences, respectively, as URFs, with the remainder of sequences being assigned to a variety of