Early Presence of HIV-1 Subtype C in Washington, D.C.

Zehava Grossman,^{1,2} Sheryl-vi Rico,³ Katherine Cone,^{3,4} Wei Shao,⁵ Catherine Rehm,³ Sara Jones,⁶ Giorgio Bozzi,¹ Sophie Dean,¹ Robin Dewar,⁵ Tauseef Rehman,⁵ Julia Purdy,⁷ Colleen Hadigan,³ Alice K. Pau,⁸ and Frank Maldarelli¹

Abstract

The presence of non-B HIV subtypes in the USA has been documented during the epidemic, although the timing of early introductions of different subtypes remains uncertain. Subtype C, the most common HIV variant worldwide, was first reported in the USA in 1996–97, after subtype C had expanded greatly in sub-Saharan Africa. In this study, we report a patient with subtype C infection acquired by mother-to-child transmission, born in the USA in 1990 to a Washington, D.C. resident who never traveled outside the USA, demonstrating that subtype C was present in the USA much earlier. Comparative analysis of the sequence from this patient and subtype C sequences in the USA and elsewhere suggest multiple independent introductions of this subtype into the USA have taken place, many of which are traced to sub-Saharan or East Africa. These data indicate subtype C HIV was already present in the USA years earlier than previously reported, and during the early period of subtype C expansion.

Keywords: HIV-1 subtype C, HIV epidemiology, transmission in Washington, D.C.

THE HIV-1 GROUP M EPIDEMIC is genetically diverse, composed of 14 distinct subtypes, over 90 circulating recombinant forms, and a number of less well-defined unique recombinant forms (Los Alamos, www.hiv.lanl.gov). Early spread of HIV out of Africa was characterized by strong founder effects, with subtype B predominating in many developed countries, including the USA. Non-B subtypes have gradually accumulated; by 2013, less than 5% of HIV infections in the USA were of non-B origin,^{1,2} but frequency of non-B subtypes has been reported to be increasing.^{3–5}

Subtype C, which is the most common HIV variant worldwide, representing over 50% of all HIV infections, is present mainly in sub-Saharan Africa and South Asia.⁶ Although the most common non-B subtype in the USA, subtype C is still relatively infrequent,¹ representing <2% of the overall U.S. HIV epidemic. Subtype C was first reported in the USA in samples obtained after 1996,^{7–9} after the subtype C epidemic had undergone a large expansion into sub-Saharan countries, including South Africa^{10,11}; these sampling times suggested

that subtype C introduction to the USA occurred during a later part of its expansion.

In this study, we report that subtype C HIV was already being transmitted in the USA before 1990, as documented in a patient with subtype C infection acquired by mother-to-child transmission in 1990. We investigated the HIV sequence from this subtype C-infected child to determine the origin and relationship with other reports of subtype C infection diagnosed in the USA and elsewhere.

The index patient, an African American female, was born in Washington, D.C. in 1990 by normal spontaneous vaginal delivery. Neither the mother nor the child received transfusions and no perinatal HIV testing was performed. Five years later, the mother was diagnosed with HIV when she developed complications of AIDS. The child was first tested for HIV at that time and found to be HIV positive. The mother died shortly after her diagnosis; she had no history of intravenous drug use and presumably acquired HIV via heterosexual transmission. The mother served in the military, but was not

¹HIV Dynamics and Replication Program, NCI, NIH, Frederick, Maryland.

²Sackler Faculty of Medicine, School of Public Health, Tel Aviv University, Tel Aviv, Israel.

³Intramural Clinical Management & Operations Branch, NIAID, NIH, Bethesda, Maryland.

⁴Department of Social Work, NIH Clinical Center, Bethesda, Maryland.

⁵Leidos Biomedical Research, Inc., Frederick, National Laboratory for Cancer Research, Frederick, Maryland.

⁶Clinical Research Directorate/Clinical Monitoring Research Program, Frederick National Laboratory for Cancer Research sponsored by the National Cancer Institute, Frederick, Maryland.

⁷Critical Care Medicine Department, NIH Clinical Center, NIH, Bethesda, Maryland.

⁸Intramural Clinical Management & Operations Branch, NIAID, NIH, Bethesda, Maryland.

been stationed outside Washington, D.C. and never traveled outside of the USA. The identity of the father is unknown.

After diagnosis, the child was treated with multiple mono- and dual- antiretroviral therapy regimens, and a number of combination antiretroviral regimens, but seldom achieved viral suppression. Her clinic notes (without laboratory documentation available) at age 9, report evidence of lamivudine resistance from HIV genotyping; clinic evaluations noted no injecting drug use or other HIV exposure risks except mother-to-child transmission. By age 12, she had acquired numerous drug resistance conferring mutations (DRMs) to nucleoside reverse transcriptase inhibitor (NRTI), including the multi-drug resistance mutation, Q151M, as well as mutations conferring resistance to non-nucleoside reverse transcriptase inhibitors (NNRTI) and protease inhibitors (PI, see Supplementary Table S1; Supplementary Data are available online at www.liebertpub.com/aid).

In 2014, she enrolled in a study of evaluation and management of treatment-experienced patients with virologic failure at the NIH Clinical Center (NCT01976715). At enrollment, her HIV RNA level was 117,951 copies/mL plasma, and CD4 T cell count was 14 cells/ μ L (2%). HIV genotyping (TRUGENE) in the protease (amino acids 3–99), reverse transcriptase (amino acids 31–247), and the entire integrase genes was performed and the virus was noted as subtype C¹² with extensive genotypic resistance, including NRTI resistance mutations D67N, T69D, V75I, V77L, T116Y, Q151M, and a complex of mutations at position 215, including 215T, A, I, V; NNRTI resistance mutations 101Q, 181C, and 190A, and PI mutation 90M; all identified previously in this patient.

The participant reported no other potential transmission risks and had no history of injection drug use. In investigating potential therapeutic options, we retrieved sequential genotypes obtained from previous rebound viremia periods. Most of these DRM were already present by 2002 (see Supplementary Table S1), indicating longstanding persistence of drug resistance.

We studied additional subtype C sequences to investigate potential phylogenetic relationships of this sequence. The *pro-pol* sequence of the child's virus, designated NIH HIV infected pediatric individual (NIH-C), was compared with other HIV-1 *pro-pol* sequences using phylogenetic and Bayesian techniques. We constructed a dataset of subtype C sequences from the USA and worldwide (Total $n = 392$, sequence length 910 nucleotides (nt), trimmed to correspond to commercial genotyping positions nt 2,262–2,549 of protease, and nt 2,661–3,282 of reverse transcriptase). The set consisted of three subsets: (1) all unique subtype C sequences reported to GenBank with the USA as country of origin and confirmed dates of isolation ($n = 280$; www.hiv.lanl.gov/; last accessed on January 26, 2016). (2) sequences most closely related to the child's sequence using a BLAST search, using the child's sequence as the query sequence (https://www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html; last accessed on January 26, 2016).

From the top 200 sequences closely related to the patient's sequence, all unique sequences of 910 nt in length with isolation dates reported in the GenBank entry were selected and considered closely related to the patient sequence ($n = 80$); all of these sequences had identity values in the BLAST algorithm ranging from 84% to 93% and E value less than 92%. 3)

Other well-curated subtype C sequences reported worldwide with known dates and locations of isolation, including a well-described set of sequences from Ethiopia and Israel ($n = 29$) and a number of subtype C reference sequences were retrieved (www.hiv.lanl.gov and <http://dbpartners.stanford.edu:8080/RegaSubtyping/stanford-hiv/typingtool/subtyping-process>).

All these sequences, and the HIV-1 reference sequence HXB2 that was used as an outgroup (GenBank accession number K03455), were aligned and subjected to (1) a maximum likelihood phylogenetic analysis (MEGA 6)¹³; trees reconstructions were performed using the entire available sequence, as well as a reduced sequence with all resistance sites removed; and (2) Bayesian evolutionary analysis by sampling trees (BEAST) after removing resistance mutations; the HKY + G model with uncorrelated lognormal relaxed molecular clock was used, and tree coalescent of constant size was used with chain length of 600,000,000 steps. The outcome posterior's effective sample size was >500 . All sequence identification numbers are provided in Supplementary Table S2.

Maximum likelihood analyses revealed (Fig. 1A) that the phylogeny of subtype C was diverse, with only a few branches having substantial bootstrap support ($>70\%$, Fig. 1A, red circle). The child's sequence, denoted NIH-C (indicated by arrow) was located within a cluster of 44 sequences originating from sub-Saharan Africa (most from South Africa and Malawi), and Europe as well as 21 of the 280 U.S. sequences studied (7.5% of U.S. sequences). The pairwise distance between the NIH-C sequence and the closest sequence in the cluster (from South Africa) was 6.7%. Identical phylogenetic branching was obtained in the presence or absence of DRM codons in the sequences.

Bayesian analysis (BEAST, Fig. 1B) was conducted, and the estimated time of recent common ancestor of the NIH-C sequence (18.6 years) was c. 1973; only a few years after the estimated emergence of subtype C in the 1960's and early 1970's.^{14,15} In addition, none of the branches of the other early U.S. subtype C sequences^{2–4} appeared to share a recent common ancestor with the NIH-C sequence, suggesting each branch was the product of an independent subtype C introduction to the USA.

In general, the spread of HIV out of Africa in the early years of the epidemic was characterized by the existence of distinct founder viruses that established specific subtypes in various parts of the world, such as subtype B in the USA and western Europe, subtype AE in Southeast Asia, and subtype A in Eastern Europe and countries of the former Soviet Union.¹⁶ The introduction and spread of other variants within these regions after the founder viruses arrived have not been well characterized, and additional analyses provide useful insight into the spread of HIV within populations. Subtype C, the most common subtype worldwide, is diverse^{11,12} and chiefly concentrated in sub-Saharan Africa,¹⁷ where it underwent a broad and complex expansion in the 1990's with multiple introductions into South Africa.^{10,11} Subtype C was reported in the USA in mid-to-late 1990's,^{2–4,14} years after other non-B subtypes had been introduced into the USA, which were already present in 1988–1992.¹⁸

In this study, we report that subtype C was actually present in the USA in the late 1980s before the large expansion of subtype C in sub-Saharan Africa; the variant we identified

FIG. 1. NIH-C sequence is related to subtype C sequences from Southern Africa, rather than to most of those reported from the USA. Sequences ($n=393$) were assembled as described and subjected to phylogenetic and BEAST analyses. **(A)** Maximum likelihood Phylogenetic analysis of 393 sequences identifies NIH-C as most closely related to sequences from sub-Saharan Africa. Several U.S. sequences are present in the same branch, but all are more distantly related to NIH-C and were isolated a number of years later. The nodes are coded by color and size according to bootstrap value; *larger red*: 70–80; smaller remainder: 50–70. **(B)** Bayesian analysis to investigate the source of NIH-C. The sequence set was analyzed with BEAST, noting the relationship between NIH-C (*arrow*) and African sequences; NIH-C has common ancestors with sequences from South Africa, Malawi, and Tanzania, including sequences from 1992. Branches leading to sequences isolated in USA are marked in *turquoise*, and those leading to sequences from South Africa, Israel (Ethiopian origin), or other locations [Malawi (9), Bulgaria (6), Belgium, Botswana, Zambia (3 each), Sweden (2), Congo, Mozambique, Portugal, Tanzania, and Zimbabwe (1 each)] in *brown*, *blue*, and *black*, respectively. The figure shows four distinct lineages. A lineage color corresponds to the color of the majority of branches it contains. Three of the four lineages are marked by the colors of African-origin isolates. The lineage, including only a small minority of sub-Saharan sequences is “American” (*turquoise*). The time scale in years from origin is indicated below each panel; branch length from the patient sequence to recent ancestral node was 18.6 years and is consistent with the early origin of this subtype C sequence. The nodes are coded by size and color according to posterior probability; *large red*: ≥ 0.95 ; *small blue*: 0.80–0.90; smaller remainder: 0.50–0.70. IL, Israel; NIH-C, NIH HIV infected pediatric individual; ZA, South Africa.

was genetically most closely related to other variants in sub-Saharan Africa. The individual reported here was diagnosed at age 5 and the most likely route of infection was mother-to-child transmission. It is, in a sense, not surprising that subtype C was identified in the District of Columbia, as the Washington metropolitan area has been a popular destination for African immigrants.

Phylogenetic analyses of HIV subtype C sequences identified in the USA revealed several distinct branches, but no strong genetic support to link these variants, suggesting multiple independent subtype C infections occurred. Detailed characterization of these events is not described, therefore it is not clear whether these were sporadic events or continuous transmissions of these individual variants.

There has been, in general, a lack of extensive viral sampling for analysis from this early period; new efforts to analyze large numbers of contemporary sequences made available through commercial genotyping sources are now in progress,¹⁹ which will yield fine structure data on the current spread of HIV subtypes and can be analyzed with these early sequences. Such efforts have critical epidemiologic value to quickly characterize new outbreaks, including point source outbreaks,^{20,21} essential for tracking the contemporary spread of HIV in populations to target prevention efforts to key populations in the setting of limited public health resources.²² Subtype analysis is also essential to characterize strains of HIV circulating in regions in preparation for vaccine efforts. An abundance of circulating viruses may have critical impact on vaccine design; subtype C viruses elicit cross-clade responses,^{23,24} which may be useful in design and composition of vaccines where multiple subtypes circulate.^{25–27} Subtypes may also influence response to antiretroviral therapy, and detailed understanding of subtype distribution in the context of prevalence of transmitted drug resistance will be useful in informing therapeutic choices for individuals initiating antiretroviral therapy.^{28,29} Continuous molecular surveillance of HIV remains a clear public health imperative.

Acknowledgments

We are grateful to Ed Tramont, Donald S. Burke, Nelson Michael, and Sheila Peel for insightful discussions.

This project has been funded, in part, with federal funds from the National Cancer Institute, National Institutes of Health, under contract no. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Author Disclosure Statement

No competing financial interests exist.

References

- Pyne MT, Hackett J, Jr, 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;51:2662–2669.
- Bennett DE, Camacho RJ, Otelea D, *et al.*: Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. *PLoS One* 2009;4:e4724.
- Chan PA, Reitsma MB, DeLong A, *et al.*: Phylogenetic and geospatial evaluation of HIV-1 subtype diversity at the largest HIV center in Rhode Island. *Infect Genet Evol* 2014;28:358–366.
- Dennis AM, Hue S, Learner E, Sebastian J, Miller WC, Eron JJ: Rising prevalence of non-B HIV-1 subtypes in North Carolina and evidence for local onward transmission. *Virus Evol* 2017;3:vex013.
- 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;24:69–73.
- Tebit DM, Arts EJ: Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infect Dis* 2011;11:45–56.
- Brodine SK, Starkey MJ, Shaffer RA, *et al.*: Diverse HIV-1 subtypes and clinical, laboratory and behavioral factors in a recently infected US military cohort. *AIDS* 2003;17:2521–2527.
- de Oliveira CF, Diaz RS, Machado DM, *et al.*: Surveillance of HIV-1 genetic subtypes and diversity in the US blood supply. *Transfusion* 2000;40:1399–1406.
- Tovanabutra S, Brodine SK, Mascola JR, *et al.*: Characterization of complete HIV type 1 genomes from non-B subtype infections in U.S. military personnel. *AIDS Res Hum Retroviruses* 2005;21:424–429.
- Wilkinson E, Engelbrecht S, de Oliveira T: History and origin of the HIV-1 subtype C epidemic in South Africa and the greater southern African region. *Sci Rep* 2015;5:16897.
- Wilkinson E, Rasmussen D, Ratmann O, Stadler T, Engelbrecht S, de Oliveira T: Origin, imports and exports of HIV-1 subtype C in South Africa: A historical perspective. *Infect Genet Evol* 2016;46:200–208.
- de Oliveira T, Deforche K, Cassol S, *et al.*: An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* 2005;21:3797–3800.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S: MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 2013;30:2725–2729.
- Roy CN, Khandaker I, Oshitani H: Evolutionary dynamics of Tat in HIV-1 subtypes B and C. *PLoS One* 2015;10:e0129896.
- Walker PR, Pybus OG, Rambaut A, Holmes EC: Comparative population dynamics of HIV-1 subtypes B and C: Subtype-specific differences in patterns of epidemic growth. *Infect Genet Evol* 2005;5:199–208.
- Hemelaar J, Gouws E, Ghys PD, Osmanov S: Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS* 2006;20:W13–W23.
- Hemelaar J, Gouws E, Ghys PD, Osmanov S; WHO-UNAIDS Network for HIV Isolation and Characterisation: Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS* 2011;25:679–689.
- Brodine SK, Mascola JR, Weiss PJ, *et al.*: Detection of diverse HIV-1 genetic subtypes in the USA. *Lancet* 1995;346:1198–1199.
- Perez-Losada M, Castel AD, Lewis B, *et al.*: Characterization of HIV diversity, phylogenetics and drug resistance in Washington, DC. *PLoS One* 2017;12:e0185644.
- Peters PJ, Pontones P, Hoover KW, *et al.*: HIV infection linked to injection use of oxycodone in Indiana, 2014–2015. *N Engl J Med* 2016;375:229–239.
- Conrad C, Bradley HM, Broz D, *et al.*: Community outbreak of HIV infection linked to injection drug use of

- oxymorphone—Indiana, 2015. *MMWR Morb Mortal Wkly Rep* 2015;64:443–444.
22. Drummond AJ, Suchard MA, Xie D, Rambaut A: Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 2012;29:1969–1973.
 23. Betts MR, Krowka J, Santamaria C, *et al.*: Cross-clade human immunodeficiency virus (HIV)-specific cytotoxic T-lymphocyte responses in HIV-infected Zambians. *J Virol* 1997;71:8908–8911.
 24. Cheedarla N, Hemalatha B, Anangi B, *et al.*: Evolution of neutralization response in HIV-1 subtype C-infected individuals exhibiting broad cross-clade neutralization of HIV-1 strains. *Front Immunol* 2018;9:618.
 25. Chen M, Yang L, Ma Y, *et al.*: Emerging variability in HIV-1 genetics among recently infected individuals in Yunnan, China. *PLoS One* 2013;8:e60101.
 26. Graham BS: Advances in antiviral vaccine development. *Immunol Rev* 2013;255:230–242.
 27. Leelawiwat W, Rutvisuttinunt W, Arroyo M, *et al.*: Increasing HIV-1 molecular complexity among men who have sex with men in Bangkok. *AIDS Res Hum Retroviruses* 2015;31:393–400.
 28. Chan PA, Kantor R: Transmitted drug resistance in nonsubtype B HIV-1 infection. *HIV Ther* 2009;3:447–465.
 29. Zhukova A, Cutino-Moguel T, Gascuel O, Pillay D: The role of phylogenetics as a tool to predict the spread of resistance. *J Infect Dis* 2017;216(Suppl_9):S820–S823.

Address correspondence to:

Frank Maldarelli

HIV Dynamics and Replication Program

NCI

NIH

Building 535, Room 108E

Frederick, MD 21702

E-mail: fmalli@mail.nih.gov