Analysis of HIV Subtype Prevalence and Geographic Distribution in a Large U.S. Population

Michael T. Pyne[1], Vera Holzmayer[2], John Hackett, Jr.[2], and David R. Hillyard[1,3]

ARUP Institute for Clinical and Experimental Pathology®[1], Abbott Diagnostics[2], and University of Utah Department of Pathology[3]

ARUP LABORATORIES
An enterprise of the University of Utah and its Department of Pathology

Abbott
A Promise for Life

Abstract

Background: Although subtype B remains the predominant form of HIV-1 in the United States, limited data is available on the prevalence and distribution of non-subtype B strains or their trends over time. Due to increased international travel and immigration, the proportion of non-B subtypes is likely to increase. Genetic diversity of HIV-1 has significant implications for diagnosis, monitoring and clinical management of patients. Thus, it is important to be aware of the overall diversity of strains in the U.S.

Methods: Sequence data from 7941 samples submitted to ARUP between 2004 and 2008 for HIV antiretroviral resistance genotyping were analyzed to determine subtype. The ViroSeq™ HIV-1 Genotyping System v 2.0 (Celera, Alameda, CA) was used to generate 1302 nucleotide sequences comprising codons 1-99 of the protease gene and 1-335 of the reverse transcriptase gene. Initial subtype characterization was performed with the REGA HIV-1 Subtyping Tool – Version 2.0 (Stanford University website). For 478 sequences not successfully subtyped using the REGA tool, phylogeny was determined using PHYLIP software (v3.5c). Sequences were analyzed for recombination breakpoints with SimPlot (v3.5.1). All non-B sequences were reanalyzed by PHYLIP, which resulted in the reassignment of five samples. Results were correlated to the patient's age, gender, and geographic information.

Updated Results: The data set included samples submitted by clients in 43 states. The age of the patients ranged from newborn to 80, with an average age of 40.2 (median of 42), 69.3% male, 30.4% female, and 0.3% unknown. Of 7941 sequences, 7709 (97.07%) were subtype B. The 232 non-B subtype samples included: 42 subtype A, 76 subtype C, 16 subtype D, three subtype F1, 11 subtype G, 42 CRF02_AG, 26 other CRFs (11 CRF01_AE, three CRF06_cpx, one CRF09_cpx, one CRF10_CD, two CRF12_BF, two CRF15_01B, one CRF16_A2D, two CRF19_cpx, one CRF21_A2D, and two CRF22_01A1), and 16 unique recombinant forms (one URF_A1J, one URF_0206, three URF_02A3, three URF_A1C, two URF_A1D, one URF_A3G, one URF_BC, one URF_BF1, and three URF_BG). Non-B samples were received from 30 states. Subtype B varied from 100% in 2004 (n=54), 98.90% in 2005 (n=546), 97.99% in 2006 (n=745), 97.34% in 2007 (n=2139), to 96.54% in 2008 (n=4457). Subtypes A and C increased most dramatically during this period.

Conclusions: The prevalence of diverse non-B subtypes and recombinants in the U.S. is steadily increasing and their geographic distribution is widespread. This study illustrates the ability to determine subtype from resistance genotyping sequences that are routinely obtained during patient treatment. Due to the potential impact on patient management, it is important to recognize the growing genetic diversity of HIV-1 in the U.S.

Introduction

HIV-1 is classified into three groups: the "major" group M (which consists of subypes A-D, F-H, J-K), the "outlier" group O, and the "non-M, non-O" group N. The vast majority of infections are group M. Due to frequent recombination events, strains referred to as "circulating recombinant forms" (CRFs) or "unique recombinant forms" (URFs) often occur. These recombinant forms consist of viruses with genomic regions derived from two or more strains.

In the United States, subtype B is the most common strain, although subtype C dominates worldwide. Most HIV-1 viral load assays have been developed and optimized to detect subtype B. In addition, most controls (including the WHO International Standard) are based on subtype B viruses. Although assays have been improved to accomodate the detection of other subtypes, genetic diversity can be challenging for molecular assays.

Results

Subtype	2004 (n = 54)	2005 (n = 546)	2006 (n = 745)	2007 (n = 2139)	2008 (n = 4457)	All Samples (n = 7941)
A	0.00%	0.18%	0.00%	0.51%	0.67%	0.53%
B	100.00%	98.90%	97.99%	97.34%	96.54%	97.08%
C	0.00%	0.00%	0.67%	0.65%	1.28%	0.96%
D, F1, G	0.00%	0.00%	0.27%	0.37%	0.45%	0.38%
CRF02_AG	0.00%	0.55%	0.54%	0.51%	0.54%	0.53%
Other CRF	0.00%	0.37%	0.27%	0.42%	0.29%	0.33%
URF	0.00%	0.00%	0.27%	0.19%	0.22%	0.20%

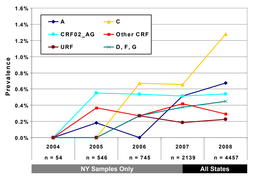


Table 1 and Figure 1: Temporal distribution of subtypes. Sequence from 7941 samples collected between 2004 and 2008 were analyzed. Note that for 2004 to mid-2007, only New York state samples were included. All states were included after mid-2007.

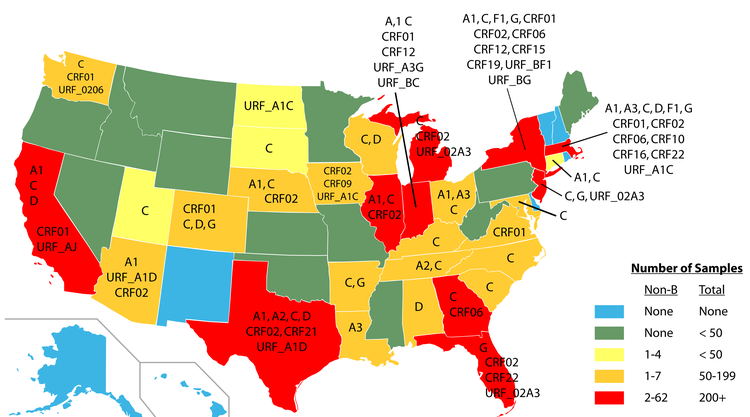


Figure 2: Geographic distribution of non-B subtypes. Samples were received from 43 states, 30 of which included non-B subtype samples. The non-B subtypes are listed for each state. The colors indicate the total number of samples received and the number of non-B samples received from each state.

State	Number of Non-B Samples	Number of All Samples	% of Total from State that are Non-B	Proportion of All Non-B	Proportion of All Samples
AL	2	61	3.3%	0.9%	0.77%
AR	2	189	1.1%	0.9%	2.38%
AZ	3	84	3.6%	1.3%	1.06%
CA	15	464	3.2%	6.5%	5.84%
CO	4	108	3.7%	1.7%	1.36%
CT	2	10	20.0%	0.9%	0.13%
FL	4	420	1.0%	1.7%	5.29%
GA	2	382	0.5%	0.9%	4.81%
IA	3	74	4.1%	1.3%	0.93%
ID	0	2	-	-	0.03%
IL	9	233	3.9%	3.9%	2.93%
IN	26	484	5.4%	11.2%	6.09%
KS	0	20	-	-	0.25%
KY	1	69	1.4%	0.4%	0.87%
LA	1	119	0.8%	0.4%	1.50%
MA	62	636	9.7%	26.7%	8.01%
MD	1	57	1.8%	0.4%	0.72%
ME	0	26	-	-	0.33%
MI	4	209	1.9%	1.7%	2.63%
MN	0	2	-	-	0.03%
MO	0	36	-	-	0.45%
MS	0	2	-	-	0.03%
MT	0	1	-	-	0.01%
NC	2	161	1.2%	0.9%	2.03%
ND	1	2	50.0%	0.4%	0.03%
NE	7	50	14.0%	3.0%	0.63%
NJ	5	200	2.5%	2.2%	2.52%
NV	0	3	-	-	0.04%
NY	40	2325	1.7%	17.2%	29.28%
OH	5	127	3.9%	2.2%	1.60%
OK	0	13	-	-	0.16%
OR	0	18	-	-	0.23%
PA	0	27	-	-	0.34%
SC	1	132	0.8%	0.4%	1.66%
SD	4	33	12.1%	1.7%	0.42%
TN	3	63	4.8%	1.3%	0.79%
TX	13	826	1.6%	5.6%	10.40%
UT	1	36	2.8%	0.4%	0.45%
VA	2	58	3.4%	0.9%	0.73%
WA	3	64	4.7%	1.3%	0.81%
WI	4	112	3.6%	1.7%	1.41%
WV	0	1	-	-	0.01%
WY	0	2	-	-	0.03%
TOTAL	232	7941		100.0%	100.00%

Table 2: Statistics by state. The 43 states from which samples were received are listed. For each state, the number of non-B samples received, the number of all samples received, the percentage of non-B samples from that state, the proportion of non-B samples, and the proportion of all samples is listed.

Non-B Strains

Subtypes	Number
A1	37
A2	2
A3	3
C	76
D	16
F1	3
G	11
TOTAL	148

CRFs	Number
CRF01_AE	11
CRF02_AG	42
CRF06_cpx	3
CRF09_cpx	1
CRF10_CD	1
CRF12_BF	2
CRF15_01B	2
CRF16_A2D	1
CRF19_cpx	2
CRF21_A2D	1
CRF22_01A1	2
TOTAL	68

URFs	Number
URF_0206	1
URF_02A3	3
URF_A1C	3
URF_A1D	2
URF_A1J	1
URF_A3G	1
URF_BC	1
URF_BF1	1
URF_BG	3
TOTAL	16

Table 3: Non-B strain details. The number and description of all 232 non-B strains are listed.

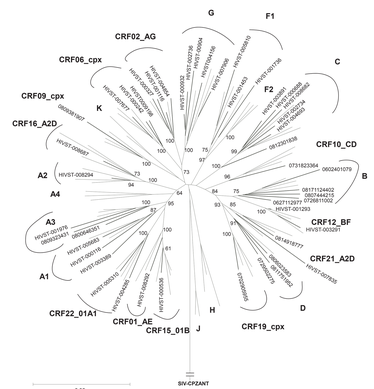


Figure 3: Phylogenetic neighbor-joining tree for protease-RT region of 1302 nucleotides. Representative specimens from each subtype and CRF are shown. Subtype and CRF designations for reference strains from Los Alamos HIV Sequence Database are shown near each branch. Bootstrap values are indicated.

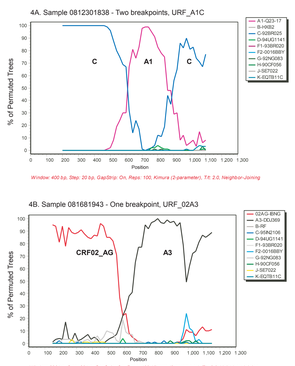
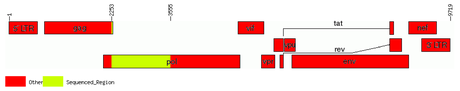


Figure 4: Identification of recombination breakpoints in protease-RT with bootscan analysis. In addition to the phylogenetic analysis, similarity plot and bootscanning were performed to identify recombination breakpoints within protease-RT. Bootscan analysis revealed several unique forms; for example, recombination of two subtypes with 1 breakpoint (not shown) or 2 breakpoints (4A), and recombination of a CRF and a subtype (4B).

Materials and Methods

Sample Sequences: Samples received by ARUP Laboratories for HIV-1 virco®TYPE or ViroSeq™ antiretroviral resistance genotyping were analyzed. HIV-1 sequences were generated using the ViroSeq™ HIV-1 Genotyping System v 2.0 (Celera, Alameda, CA), which produces 1302 nucleotide sequences comprising codons 1-99 of the protease gene and 1-335 of the reverse transcriptase gene (illustrated below).



HIV-1 sequencing region. The 1302 nucleotide region used in the analyses is shown in relationship to the HIV-1 genome.

Database: The database contains specimens received between 2004 and 2008. Prior to mid-2007, only New York state specimens were included; after mid-2007, sequences from all states were included. The database contains demographic information, including patient age, gender, and client geographic information. The sequences were deidentified by a third party by removing all Personal Health Information (PHI) from the database.

Sequence Analysis: All sequences were initially analyzed using the REGA HIV-1 Subtyping Tool, Version 2.0 (http://dbpartners.stanford.edu/RegaSubtyping/). Sequences that were not successfully subtyped using this tool were analyzed using PHYLIP software (version 3.573; J. Felsenstein, University of Washington, Seattle, WA) and SimPlot software (version 3.5.1; S. Ray, Johns Hopkins University, Baltimore, MD; http://sray.med.som.jhmi.edu/RaySoft/SimPlot/). In addition, all sequences that were non-B by REGA were reanalyzed by PHYLIP and SimPlot. In the case of discordant subtype assignments, the PHYLIP/SimPlot results were used. Five samples were reassigned in this manner (two CRF01_AE sequences were reassigned CRF15_01B, two A1 sequences were reassigned A3, and one A1 sequence was reassigned CRF22_01A1).

Phylogenetic analysis: Sequences that were untypable by the REGA Tool (results of "Check the Bootscan" or "Sequence Error") were aligned with the reference sequences representing HIV-1 group M subtypes and CRFs (1-43), O, N and SIV (Los Alamos HIV Sequence Database; http://www.hiv.lanl.gov/) using BioEdit (version 7.0.4.1, Department of Microbiology, North Carolina State University, Raleigh, NC). Phylogenetic analysis on gap-stripped alignments was performed with PHYLIP software. Evolutionary distances were estimated with DNADIST (Kimura two-parameter method; transition-transversion ratio of 2.0). Phylogenetic reconstructions were generated with NEIGHBOR using the neighbor-joining method. Branch reproducibility was evaluated using SEQBOOT on 100 replicates. Phylogenetic trees were displayed with TreeExplorer. Bootstrap values >70 were considered acceptable for subtype assignment.

Bootscanning analysis: Sequences were examined for intersubtype recombination using SimPlot software. Putative recombinant sequences were analyzed in alignments with one reference sequence for each of HIV-1 group M subtypes (A-D, F-H, J-K and the major CRFs). SimPlot and BootScan were run with a sliding window of 300 or 400 nucleotides and 20 nucleotide steps.

Conclusions

- The prevalence of non-B subtypes is steadily increasing in the U.S.
- The overall percentage of non-B strains between 2004 and 2008 was 2.93%.
- Subtypes A and C increased most dramatically, while the prevalence of CRFs was relatively stable.
- Non-B strains were present in 30 of 43 states.
- Subtypes A1, A2, A3, C, D, F1, and G were detected.
- Eleven different circulating recominant forms were detected.
- Nine different unique recombinant forms were detected.
- Reliable subtype results can be obtained with sequences that are routinely available during patient treatment.
- Most sequences could be analyzed using automated online tools. However, some samples required more sophisticated analysis for accurate subtype assignment.
- The increasing genetic diversity of HIV-1 in the U.S. is important to recognize, due to the potential impact on patient care.

Acknowledgements

The authors wish to thank Heather Beagley for performing the database queries and Denise Jones for removing PHI from the data.

Presentation and Contact Info

Abstract B-106 :: Poster 292 :: Session 59
16th Conference on Retroviruses and Opportunistic Infections
Montréal, QC, Canada
February 8-11, 2009

Michael T. Pyne, M.S.
ARUP Laboratories
500 Chipeta Way
Salt Lake City, UT 84108
michael.pyne@aruplab.com
801-583-2787 x3431