Contact Information:
Mary Rodgers
Abbott Diagnostics
100 Abbott Park Rd
Abbott Park IL 60064-6015
Email: mary.rodgers@abbott.com

A THIRD COMPLETE GENOME ESTABLISHES HIV-1 SUBTYPE L

0935

Julie Yamaguchi[1], Ana Vallari[1], Carole McArthur[2], Larry Sthreshley[3], Gavin Cloherty[1], Michael G. Berg[1], and Mary A. Rodgers[1]

[1]Abbott Diagnostics, Abbott Park, IL, USA, [2]School of Dentistry, University of Missouri – Kansas City, Kansas City, MO, USA, [3]Presbyterian Church (USA), Kinshasa, Democratic Republic of the Congo

Background

As part of the region where HIV-1 initially expanded early in the global pandemic, the Democratic Republic of Congo (DRC) is where the most diverse HIV isolates have been found, including all recognized Group M subtypes and many unclassifiable sequences. Two divergent non-recombinant sequences, 83CD003 and 90CD121E12, collected in 1983 and 1990 in DRC were previously proposed as a new subtype, L. However, HIV nomenclature standards require three epidemiologically distinct genomes for a new classification[1]. A new Group M subtype has not been established since the guidelines were published in 2000.

Methods

Specimen CG-0018a-01 was collected in DRC in 2001 as part of an HIV diversity study. Previous subgenomic HIV-1 sequences branched closely with the proposed subtype L references, but small sample volume and low viral load limited efforts to expand genome coverage. In the present study, the complete genome was assembled through application of metagenomic Next-Generation-Sequencing (mNGS) and target enrichment (HIV-xGen) methods. Neighbor-joining phylogenetic and recombinant analyses were completed to classify the genome using Phylip v3.5 and Simplot v3.5.1.

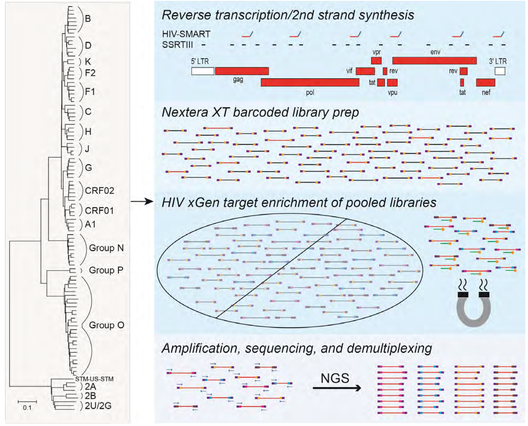


Figure 1. HIV-xGen target enrichment method. CG-0018a-01 plasma is extracted on an *m*2000 instrument, HIV-1 RNA is transcribed to ds cDNA, adapters+barcodes are added with Nextera XT, and HIV sequences are captured by custom non-redundant biotin-labeled HIV-xGen probes, amplified on streptavidin beads and sequenced on a MiSeq.[2]

Results

Next generation sequencing (NGS) The combined mNGS and HIV-xGen approach yielded a 9681 nucleotide HIV genome for specimen CG-0018a-01. This sequence was deposited into Genbank and assigned accession number MN271384.

Table 1. NGS metrics show that 39% of the combined mNGS and HIV-xGen library reads mapped to CG-0018a-01.

Total reads	11,046,542
Mapped HIV reads	4,363,301
% Mapped HIV reads	39%
Average coverage depth	47,783x

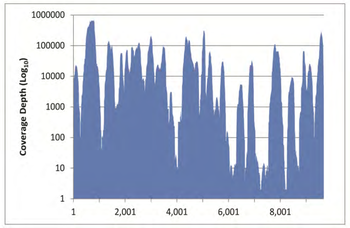


Figure 2. Coverage plot of NGS data for the CG-0018a-01 HIV-xGen library illustrates the average coverage depth of 47,783x across the length of the genome.[3]

Phylogenetic Inference The CG-0018a-01 genome branched with the putative subtype L references with a bootstrap value of 100 in a phylogenetic tree. Notably, the CG-0018a-01 branch was basal to the junction of 83CD003 and 90CD121E12, which suggests CG-0018a-01 may be more closely related to an ancestral strain.

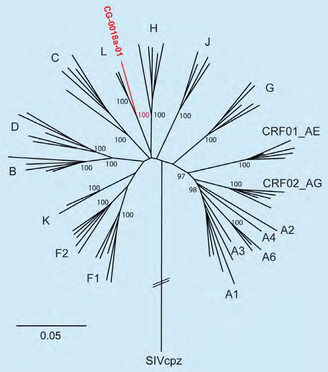


Figure 3. Phylogenetic tree. HIV-1 group M maximum likelihood phylogenetic tree (7578 bp) shows the sequence of CG-0018a-01 groups with the two putative subtype L sequences with a bootstrap of 100.[3]

Recombinant analysis Recombinant analysis did not present evidence of other subtype sequences as the putative subtype L references had the highest percent identity to CG-0018a-01 across the genome except in the well-conserved *pol* region. Subgenomic phylogenetic analysis of the *pol* region confirmed that CG-0018a-01 branched with L references with bootstrap support of 97.

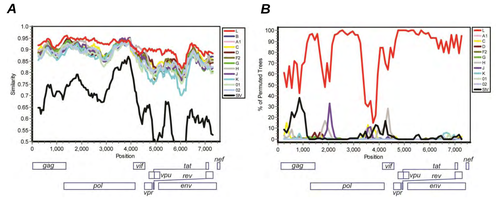


Figure 4a. Recombinant analysis with Simplot. A) SimPlot (Window: 500 bp, Step: 50 bp) and B) BootScan (Window: 500bp, Step: 100 bp) show % identity is highest to the putative subtype L reference sequences except in the well-conserved *pol* region (approximate positions 3500-4600).[3]

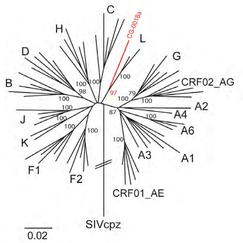


Figure 4b. Results of Neighbor-Joining phylogenetic analysis of positions 3500-4600 show that the sequence of CG-0018a-01 continues to group with the putative subtype L sequences with a bootstrap of 97.[3]

	L	A	B	C	D	F	G	H	J	K
L		0.146	0.142	0.146	0.147	0.142	0.148	0.140	0.145	0.137
A	0.143		0.156	0.159	0.160	0.157	0.145	0.154	0.156	0.151
B	0.139	0.153		0.155	0.123	0.144	0.160	0.153	0.153	0.140
C	0.143	0.155	0.152		0.158	0.153	0.160	0.152	0.155	0.148
D	0.143	0.155	0.122	0.154		0.149	0.160	0.156	0.157	0.145
F	0.139	0.153	0.141	0.149	0.145		0.154	0.150	0.154	0.128
G	0.144	0.141	0.156	0.156	0.155	0.150		0.154	0.149	0.149
H	0.137	0.151	0.150	0.149	0.152	0.147	0.150		0.150	0.142
J	0.141	0.151	0.149	0.151	0.151	0.149	0.144	0.146		0.142
K	0.135	0.148	0.138	0.145	0.141	0.125	0.145	0.139	0.138	

Table 2. Evolutionary distances. The number of amino acid substitutions per site from averaging over all sequence pairs between groups are shown below the diagonal and the number of base substitutions per site from averaging over all sequence pairs between groups are shown above the diagonal. Comparisons to L are highlighted in blue with white text. Analyses were conducted using the Poisson correction model[4] for amino acids and the Maximum Composite Likelihood model for nucleotides[5] in MEGA6[6] with 55 reference strain sequences.

Serological testing To evaluate whether current diagnostic and blood screening tests can detect subtype L, we tested a 1:10,000 dilution of CG-0018a-01 into normal human plasma on the following tests and platforms. The specimen was reactive in all cases.

Platform	Test	Result
Abbott ARCHITECT	HIV Ag/Ab Combo	repeat reactive
ABBOTT PRISM	HIV O+	repeat reactive
ABBOTT PRISM	*HIV Ag/Ab Combo	repeat reactive
Abbott Alinity i	*HIV Ag/Ab Combo	repeat reactive
Abbott Alinity s	HIV Ag/Ab Combo	repeat reactive
Abbott Rapid	*Alere HIV Combo	reactive

Table 3. Serological testing. Sample CG-0018a-01 was tested at a 1:10,000 dilution in normal human plasma. *These tests are not commercially available in the US.

Molecular testing To evaluate whether current diagnostic molecular tests can detect subtype L, we tested the neat specimen CG-0018a-01 or a 1:4 dilution into normal human plasma on the following tests and platforms. The specimen was detected in all cases.

Platform	Test	Result
Abbott m2000	RealTime HIV	Detected
Abbott Alinity m	*HIV-1	Detected
Abbott mPIMA	*DETECT HIV 1/2	Detected

Table 4. Molecular testing. Sample CG-0018a-01 was tested undiluted. *This test is not commercially available in the US

Conclusions

1) The HIV-xGen NGS target enrichment method can yield complete genomes from diverse low viral load specimens
2) A complete HIV genome sequence was generated for CG-0018a-01 and has been deposited into Genbank under accession number MN271384
3) **Subtype L is a new Group M subtype defined by three genome sequences: CG-0018a-01, 83CD003, and 90CD121E12.**
4) All three subtype L strains were identified in Democratic Republic of Congo over the course of 30 years, prevalence is therefore low
5) All Abbott serological and molecular tests were able to detect the new subtype L
6) Continued molecular surveillance is important for ensuring diagnostic and blood screening tests are able to detect all circulating strains

References

1) Robertson DL, Anderson JP, Bradac JA, et al. HIV-1 nomenclature proposal. *Science*. 2000;288(5463):55-56.

2) Yamaguchi J, Olivo A, Laeyendecker O, Forberg K, Ndembi N, Mbanya D, Kaptue L, Quinn TC, Cloherty GA, Rodgers MA, Berg MG. Universal Target Capture of HIV Sequences From NGS Libraries. *Front Microbiol*. 2018; 9:2150.

3) Yamaguchi J, Vallari A, McArthur C, Sthreshley L, Cloherty G, Berg MG, and Rodgers MA. *JAIDS*. 2020 Mar 1; 83(3):319-322.

4) Zuckerkandl E. and Pauling L. (1965). Evolutionary divergence and convergence in proteins. Edited in Evolving Genes and Proteins by V. Bryson and H.J. Vogel, pp. 97-166. Academic Press, New York.

5) Tamura K., Nei M., and Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *PNAS*. 2004; 101:11030-11035.

6) Tamura K., Stecher G., Peterson D., Filipski A., and Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution 2013*; 30: 2725-2729.