Evolution of Resistance Uncovered by Single-Genome Sequencing (SGS) During Continuous or Interrupted Antiretroviral Therapy

J.K. Kayondo[1], N. Ndembi[2], P.A. Cane[3], C. Kityo[5], M. Chirara[6], S. Hué[4], R. Goodall[7], D. Dunn[7], P. Kaleebu[1,2], D. Pillay[3,4] and J.L. Mbisa[3] on behalf of the DART Virology Committee

[1]Uganda Virus Research Institute (UVRI) and [2]Medical Research Council (MRC)/UVRI, Uganda Research Unit on AIDS, Entebbe, Uganda; [3]Health Protection Agency, London, UK ; [4]University College London Medical School, London, UK; [5]JCRC, Kampala, Uganda; [6]University of Zimbabwe, Harare, Zimbabwe; and [7]MRC Clinical Trials Unit, London, UK.

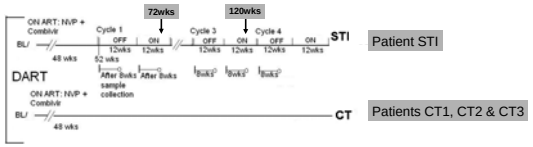
BACKGROUND

Detailed study of HIV-1 evolution under drug selective pressure, through SGS, uncovers the dynamics and *in vivo* fitness determinants of intrapatient genetic variants. We have investigated patients enrolled within the DART study, in which retrospective viral load testing identified individuals with persistent viraemia while on continuous therapy (CT), and also during a structured treatment interruption (STI) protocol. We asked whether differences in the persistence and evolution of the plasma virus population between these individuals could be identified.

METHODS

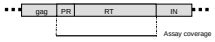
Four patients had persistent viraemia during combivir/nevirapine therapy, of which one was enrolled in a cycling on-off STI protocol; 12 weeks off and 12 weeks on drugs starting from week 52 of initiating therapy. Samples were studied from weeks 72 and 120. Thirty or more sequences from each time point were generated by SGS of complete protease-reverse transcriptase (PR-RT) genes, and phylogenetic analyses were performed.

I. STI and CT Protocol:



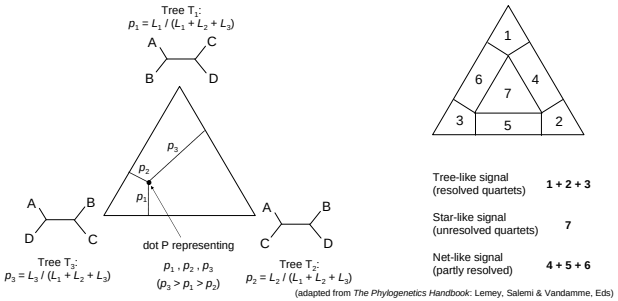
II. Single-genome sequencing assay:

Single-genome sequencing based on limiting dilution of cDNA of full length PR-RT region was performed (Palmer *et al.*, 2005). Drug resistance mutations were defined by the use of the Stanford University HIV Drug Resistance Database.



III. Phylogenetic analyses:

(a) Phylogenetic structure. Evolutionary information contained within a data set of single genome sequences was analyzed using likelihood-mapping method (Strimmer & von Haeseler, 1997). The method computes three posterior probabilities (p_1, p_2, and p_3) for three possible unrooted trees of a quartet of randomly chosen sequences A, B, C, and D (see figure below). These are then reported as a point (P) inside an equilateral triangle representing specific tree topologies with the likelihoods L_1, L_2, and L_3, respectively. The triangle is divided into seven main areas supporting different evolutionary information. The distribution of 10,000 randomly chosen quartets were computed for each data set.



(adapted from *The Phylogenetics Handbook*: Lemey, Salemi & Vandamme, Eds)

(b) Population diversity. The sequence variability in the viral population at each time point was analyzed by computing the pairwise genetic distances. Box-and-whisker plots were used to visualize data distribution.

Table 1. *Genetically linked resistance mutations on single genomes of STI patient (subtype D).*

Wild-type residue	Week 72 (n=35; VL=9,640)	Week 120 (n=34; VL=22,000)
D67	N N N G	
V90		I
L100	Q	P
K103	N N N	N N N
M184	V V V V	V V V
G190	A	
K219	E	
M230	L	
N348	I I I	I I I
% of genomes	83 6 3 3	94 3 3

Table 2. *Genetically linked resistance mutations on single genomes of CT1 patient (subtype A).*

Wild-type residues	Week 72 (n=31; VL=126,000)	Week 120 (n=34; VL=566,000)
M41		L
D67	N	N N N N N N
T69		N
K70		R R R R R
V90	I	
A98	G	
L100	S	
K101	E	
K103	N S N N N	N S S S N N S S
M184	V V V V V V V	V V V V V V V
G190	A A A A A A A	A A A A A A A
T215		F F F F
K219		E E E E
N348	I I I I I	
% of genomes	74 10 3 3 3 3 3	68 9 3 3 3 3 3 3

Table 3. *Genetically linked resistance mutations on single genomes of CT2 patient (subtype A).*

Wild-type residues	Week 72 (n=30; VL=87,900)	Week 120 (n=39; VL=119,000)
M41	L L L L L L L	L L L L L L L
E44	D D D D	D D D D D D D
D67	N N N N	N N N N N N N
V90	I	
A98	G G G G G G G	G G G G G G G
V106		I
V118	I I	I I I I I
V179	I I I I I I I	I I I I I I I
M184	V V V V V V V	V V V V V V V
G190	A A A A A A A	A A A A A A A
L210	W W W W W W W	W W W W W W W
T215	Y Y F Y Y F Y	Y Y Y Y Y Y Y
K219		N N
H221		Y Y Y
% of genomes	63 20 13 7 3 3 3	31 26 13 13 10 3 3 3

Table 4. *Genetically linked resistance mutations on single genomes of CT3 patient (subtype D).*

Wild-type residues	Week 72 (n=46; VL=3,130)	Week 120 (n=32; VL=35,500)
K65		R
T69		S N
K70	R R R	R R R
V90	I I I	I
V108	I I I I I I I	I I I I
V179	I	
Y181	C C C C C C C	C C C C C C C
M184	V V V V V V V	V V V V V V
G190	A R	
K219		Q
H221	Y Y Y Y Y Y Y	Y Y Y Y Y Y
N348	I I I I I I I	I I I I I I
% of genomes	63 22 7 4 2 2 50	31 9 3 3 3 3

- STI patient shows a decay in resistance mutations between time points.
- A higher percentage of genetically linked single genomes belong to one species in the STI patient while the distribution is more diverse in the CT patients.

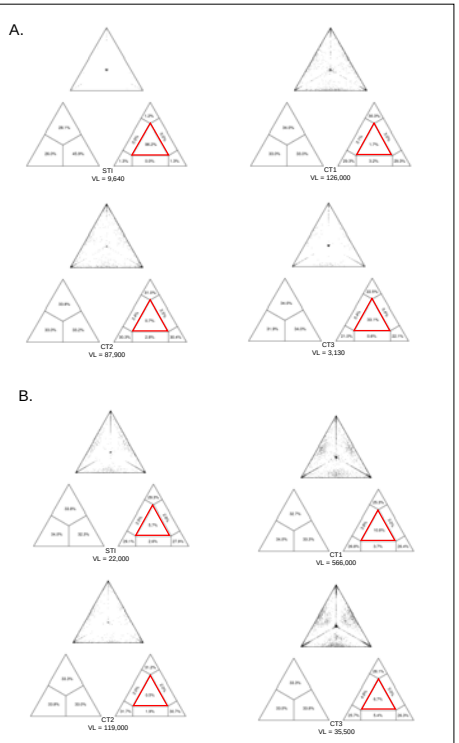


FIG 1. Likelihood mapping analysis of the single genome sequences. *The dots in the top triangle of each analysis represent the likelihood of the three possible trees computed for a quartet of sequences randomly selected from each data set. The bottom two triangles give the percentages of the dots in the different areas. The percentage in the red triangle represents unresolved quartets indicating the extent of sequence homogeneity. Likelihood mapping of 10,000 random quartets from each patient is shown at 72 weeks (A) and 120 weeks (B). The STI sequences at 72 weeks have the highest homogeneity with 96.2% of the dots located in the red triangle.*

FIG 2. Box-and-whisker plots showing the distribution of the pairwise genetic distances for single genome data sets. *The box contains the middle 50% of all values (25th to 75th percentile), also called the interquartile range (IQR) and the horizontal bar within the box represents the median value (50th percentile). The whiskers represent the remainder of the values within 1.5× IQR and any outliers are shown as dots. Pairwise genetic distance data (nucleotide substitutions/site) from each patient is shown at 72 weeks (A) and 120 weeks (B). The IQR of the STI patient data at 72 weeks (0.0005) is significantly less than that of the CT patients (0.0032, 0.0056 and 0.0089 for CT1, CT2 and CT3 respectively).*

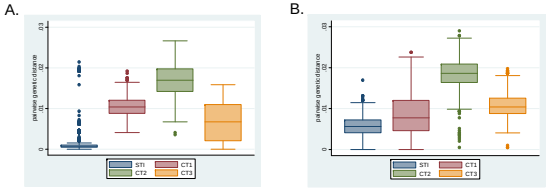
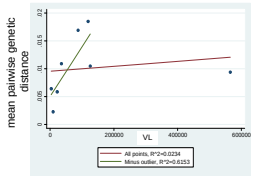


FIG 3. Linear regression analysis of pairwise genetic distance and viral load. *Mean pairwise genetic distances for the patients at each time point were plotted against the viral loads (VL) to determine if population diversity is correlated to the amount of virus. The data shows correlation at lower viral loads (up to 150,000 copies/mL). This suggests that data from STI and CT patients with similar VL should be used for comparison (e.g. STI vs CT3).*



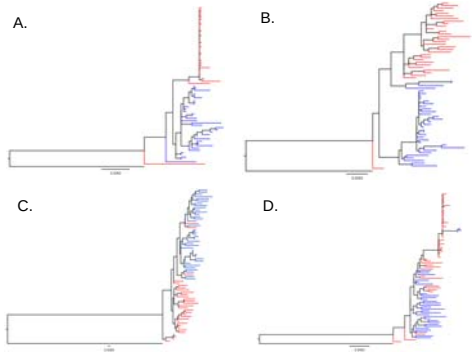


FIG 4. Maximum likelihood phylogenetic analysis of single genome sequences. *Branch lengths were estimated using the GTR model of substitution and are drawn in scale with the bar at the bottom representing 0.006 nucleotide substitutions per site. The colour for each branch tip represents single genomes from each time point; red, 72 weeks; blue, 120 weeks. (A) STI, (B) CT1, (C) CT2 and (D) CT3. The trees are rooted (black branch tip) using subtype-specific sequences; subtype D sequence 94UG114 (A and D); subtype A sequence 92UG037 (B and C). In general, there is clustering of sequences from each time point in all patients, however, the 72 weeks sequences from the STI patient show the shortest branch lengths in keeping with the low diversity of the viral population.*

CONCLUSIONS

- The data shows that there is a delay in viral population diversity in the patient undergoing STI at 72 weeks compared to patients on the CT protocol.
- Viral population diversity is correlated to viral load, however, this relationship does not hold at very high viral loads.
- The diversity in genetically linked drug resistance viral genomes exists at both time points suggesting a continual battle in the viral fitness landscape.
- More work is required to fully appreciate the impact of variable adherence/treatment interruption on within host evolutionary dynamics including sequences from off-treatment periods and from baseline samples.

Acknowledgements: We thank Maria Nannyonjo, Margaret Nabankema, Deogratius Ssemwanga and other Basic Science Lab members at MRC/UVRI; patients, staff and project management of the DART study; Centre for Infections, Health Protection Agency where most of the work was carried out; Wellcome Trust Strategic Award, grant no. 084344 for a Mak-UVRI *Post-doctoral* Fellowship in Infection and Immunity (JKK).