EARLY HIV-1 GENETIC DIVERSITY INCLUDES CTL AND DRUG RESISTANCE MUTATIONS

346

Adam A. Capoferri[1], Valerie F. Boltz[1], Wei Shao[2], Clarissa Halpern[1], Rasmi Thomas[3], Morgane Rolland[3,4], Nittaya Phanuphak[5], Lydie Trautmann[3,4], Sandhya Vasan[3,4], Julie Ake[3], Carlo Sacdalan[6,7], Somchai Sriplienchan[6], Jason W. Rausch[1], John W. Mellors[8], John M. Coffin[9], Mary F. Kearney[1], on behalf of the RV254/SEARCH 010 Study Team

[1] HIV Dynamics and Replication Program, National Cancer Institute, Frederick, MD, USA [2] Leidos Biomedical Research, Inc., Frederick National Laboratories for Cancer Research, Frederick, MD, USA [3] US Military HIV Research Program, Center for Infectious Diseases Research, Walter Reed Army Institute of Research, Sliver Spring MD, USA [4] Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, MD, USA [5] Institute of HIV Research and Innovation (IHRI), Bangkok, Thailand [6] SEARCH Research Foundation, Bangkok, Thailand [7] Research Affairs, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand [8] Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA [9] Department of Molecular Biology and Microbiology, Tufts University, Boston, MA, USA

BACKGROUND/AIMS

To better define the number of transmitted founder variants and sequence divergence from founders in the first few weeks following HIV-1 transmission, we performed ultra-sensitive single genome sequencing (uSGS) of *pol* and *env* HIV-1 RNA in plasma samples from individuals with acute HIV-1 infection.

METHODS

- HIV-1 RNA was extracted from plasma of donors with acute CRF01AE infection (Fiebig II-IV) enrolled in the RV254/SEARCH010 Cohort (NCT00796146) (**Table 1**).
- Donors were all Single transmitted founders (TFs).
- Ultrasensitive SGS method with primer IDs and paired-end Illumina sequencing was applied to identify >10,000 independent *pol* (RT, aa 184-261) and *env* (gp120/V3, aa 259-331) sequences per sample.
- Comprehensive genetic analyses were performed.

Table 1. Donor demographics of 10 RV254/SEARCH 010 trial plasma samples at sampling.

Participant ID (PID)	Sex	Age (years)	Fiebig Stage	Viral load (copies/mL)	CD4 count (cells/uL)	MHC-I HLA alleles		
						A1/A2	B1/B2	C1/C2
3928	M	46	III	5,517,440	525	11/11	15/15	08/08
3832	M	36	II	36,694,000	269	02/33	44/46	01/07
5436	M	32	III	30,811,000	621	11/33	44/52	07/07
3698	M	30	III	4,939,160	352	02/11	18/35	04/07
6609	M	29	III	4,112,500	206	24/33	07/51	07/14
8123	F	45	III	25,579,700	132	02/11	13/35	03/04
3513	M	28	III	13,557,900	359	11/11	18/40	07/15
9114	M	22	III	2,656,900	350	02/24	46/52	01/12
3193	M	24	III	2,412,840	289	33/33	44/58	03/07
8522	M	29	III	10,000,000	265	24/33	51/58	03/14

RESULTS

Table 2. Total numbers of single genomes resulting from uSGS from *pol* and *env*.

PID	Fiebig Stage	*pol* subgenomic region		*env* subgenomic region	
		# total genomes sequenced (# unique genomes)	% of total genomes for each T/F	# total genomes sequenced (# unique genomes)	% of total genomes for each T/F
3928	III	19,251 (472)	86	17,242 (351)	94
3832	II	12,518 (219)	95	12,993 (215)	96
5436	III	7,766 (206)	91	10,329 (209)	85
3698	III	8,148 (289)	69, 18[a]	10,120 (238)	95
6609	III	10,386 (269)	92	13,476 (228)	91
8123	III	11,339 (238)	94	13,867 (241)	93
3513	III	12,913 (230)	96	12,459 (209)	92
9114	III	27,884 (438)	87	10,957 (192)	96
3193	III	3,704 (96)	94	6,914 (144)	96
8522	III	22,675 (351)	94	12,602 (138)	95

[a] 1nt different from other viral population, possible Founder Effect

Acknowledgements:
We would like to thank the study participants who committed so much of their time for this study. The participants were from the RV254/SEARCH 010, which is supported by cooperative agreements (WW81XWH-18-2-0040) between the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., and the United States Army Medical Research and Development Command (USAMRDC) and by an intramural grant from the Thai Red Cross AIDS Research Centre and, in part, by the Division of AIDS, the National Institute of Allergy and Infectious Diseases, National Institutes of Health (AAI21058-001-01000). This work was supported by intramural NCI funding to the HIV Dynamics and Replication Program and by the Office of AIDS Research. Antiretroviral therapy for RV254/SEARCH 010 participants was supported by the Thai Government Pharmaceutical Organization, Gilead Sciences, Merck and ViiV Healthcare.

Disclaimer:
The views expressed are those of the authors and should not be construed to represent the positions of the U.S. Army, the Department of Defense, the National Institutes of Health, the Department of Health and Human Services, or the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc. The investigators have adhered to the policies for protection of human subjects as prescribed in AR-70-25.

Table 3. Genetic diversity after transmission of single founders for *pol* and *env* regions.

PID	Fiebig Stage	*pol subgenomic region*				*env subgenomic region*			
		%APD[a]	dN/dS (rel. T/F)[b]	dN/dS (rel.Con CRF01AE)[c]	Ti/Tv[d]	%APD[a]	dN/dS (rel. T/F)[b]	dN/dS (rel.Con CRF01AE)[c]	Ti/Tv[d]
3928	III	0.13	0.30	3.66E-03	4.1	0.06	1.00	0.62	2.6
3832	II	0.05	0.25	0.20	7.2	0.03	1.00	0.56	3.2
5436	III	0.08	0.67	0.04	8.5	0.05	0.67	0.84	4.1
3698	III	0.25	0.06	0.04	24.1	0.05	1.25	0.044	3.1
6609	III	0.07	0.57	0.14	5.1	0.09	0.17	0.43	5.7
8123	III	0.06	0.38	2.48E-03	4.7	0.06	0.33	0.33	7.2
3513	III	0.04	0.38	0.12	5.0	0.07	1.00	2.81	4.4
9114	III	0.12	0.33	0.24	2.5	0.04	1.00	0.59	4.8
3193	III	0.05	0.38	0.12	7.6	0.04	1.25	0.60	3.5
8522	III	0.06	0.67	0.13	1.6	0.04	0.75	0.29	2.5
Median		**0.07**	**0.38**	**0.12**	**5.1**	**0.05**	**1.00**	**0.57**	**3.8**
IQR		**0.05-0.11**	**0.31-0.52**	**0.04-0.14**	**4.2-7.5**	**0.04-0.06**	**0.69-1.00**	**0.36-0.61**	**3.1-4.7**
Average (SD)		**0.09 (0.06)**	**0.40 (0.19)**	**0.10 (0.08)**	**7.0 (6.4)**	**0.05 (0.02)**	**0.84 (0.36)**	**0.71 (0.77)**	**4.1 (1.5)**

[a] The percent average pairwise distance
[b] The ratio of nonsynonymous to synonymous mutations relative to the determined transmitted/founder virus across the entire subgenomic region
[c] The ratio of nonsynonymous to synonymous mutations relative to the consensus CRF01AE virus across the entire subgenomic region
[d] The transition to transversion ratio for the subgenomic region

Figure 1. Nucleotide substitution rates of single transmitted/founders.

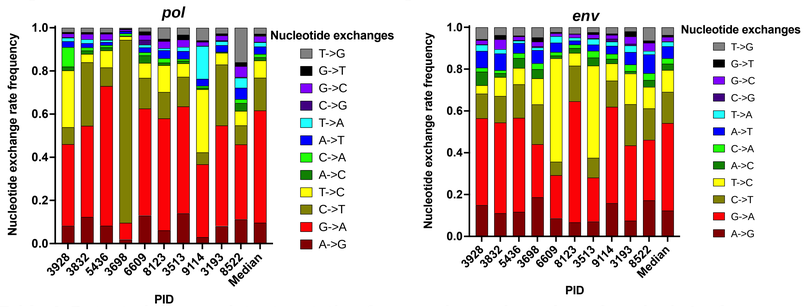

Table 4. Detected drug resistant mutation frequencies and number of copies/mL of plasma in inferred single transmitted/founder virus participants.

PID	% Frequency observed (drug resistant mutation log_{10} copies/mL)										Level of plasma viremia (log_{10} copies/mL)
	M184I	M184V	L210W	K219E	K219Q	Y188C	Y188H	G190A	G190E	M230L	
3928	4.03 (5.35)	-	0.21 (4.07)	0.21 (4.07)	-	0.21 (4.07)	0.21 (4.07)	-	1.69 (4.97)	0.21 (4.07)	6.74
3832	0.46 (5.22)	-	-	-	-	-	-	-	1.83 (5.83)	-	7.56
5436	0.97 (5.48)	-	-	-	-	-	-	0.49 (5.17)	3.40 (6.02)	-	7.49
3698	0.35 (4.23)	-	-	-	-	-	0.69 (4.53)	-	0.69 (4.53)	-	6.69
6609	1.12 (4.66)	-	-	-	0.37 (4.18)	-	-	-	1.49 (4.79)	-	6.61
8123	1.26 (5.51)	0.42 (5.03)	-	-	-	0.42 (5.03)	0.42 (5.03)	-	0.84 (5.33)	0.42 (5.03)	7.41
3513	1.74 (5.37)	-	-	0.43 (4.77)	-	0.43 (4.77)	-	0.43 (4.77)	1.74 (5.37)	-	7.13
9114	3.42 (4.96)	-	-	-	-	-	0.23 (3.78)	0.23 (3.78)	3.65 (4.99)	-	6.42
3193	-	1.04 (4.40)	-	-	-	-	1.04 (4.40)	-	1.04 (4.40)	-	6.38
8522	1.14 (5.06)	-	-	-	-	0.28 (4.45)	-	-	2.85 (5.45)	0.58 (4.76)	7.00

- Average of 11,733 *pol* and 12,096 *env* single genomes were obtained from each participant with 93% of the genomes being identical (**Table 2**).
- Genetic diversity by average pairwise distance (APD) had a median of 0.07% in *pol* and 0.05% in *env* (p=0.10). The average dN/dS was 0.38 in *pol* and 1.0 in *env* (p=0.02). Preference toward transitions was observed with a median Ti/Tv of 5.1 in *pol* and 3.8 in *env* (p=0.21) (**Table 3**).
  - Minimal inferred single step rate was found to be significantly less than HIV-1 RT, with the *pol* median rate of $5.1x10^{-6}$ mut·site^{-1}·day^{-1} (6.9-fold, p<0.0001) and *env* median rate of $3.7x10^{-6}$ mut·site^{-1}·day^{-1} (9.5-fold, p<0.0001).
- In *pol*, 52% of transitions were G>A and in *env*, 42% were G>A, with the split divided over the other 3 mutations **(Figure 1).**
- Drug resistance mutations were detected in all participants (0.2-4.0% of genomes), but were not linked to one another (**Table 4**).
- Reversions to conAE were observed at 57% of sites that varied from the TF, but these mutations only comprised 0.3% of all amino acid changes in *pol* and 2.8% in *env* (p<0.0003) (**Table 5**).
- The remainder were mutations away from conAE, with 42% in *pol* and 51% in *env* falling within CTL epitopes that matched the participants' HLA (**Table 6**).

Table 5. Percent of positions undergoing reversions to the consensus CRF01AE indicating lack of driving diversity.

PID	# of positions in T/F that were different than Con CRF01AE[a]	# of positions that had a reversion back to Con CRF01AE (%)[b]	Total # of different amino acid changes in dataset[c]	% of amino acid mutations that were reversions to Con CRF01AE
		subgenomic *pol*		
3928	0	0 (—)	358	—
3832	3	1 (33)	128	0.78
5436	1	1 (100)	157	0.64
3698	1	0 (0)	142	<0.704
6609	1	0 (0)	180	<0.006
8123	0	0 (—)	156	—
3513	2	1 (50)	152	0.66
9114	4	2 (50)	390	0.01
3193	2	1 (50)	71	0.01
8522	4	3 (75)	273	0.01
Median	**2**	**1 (50)**	**157**	**0.33**
IQR	**1-3**	**0-1 (8-69)**	**139-294**	**0.007-0.69**
		subgenomic *env*		
3928	14	7 (50)	280	2.50
3832	8	4 (50)	151	2.65
5436	12	10 (83)	160	6.25
3698	10	9 (90)	175	5.14
6609	6	4 (67)	136	2.94
8123	7	4 (57)	174	2.30
3513	15	6 (40)	139	4.32
9114	6	2 (33)	152	1.32
3193	6	4 (67)	103	3.88
8522	8	4 (50)	152	2.63
Median	**8**	**4 (54)**	**152**	**2.80**
IQR	**6-13**	**4-8 (48-71)**	**138-174**	**2.45-4.52**
p-value[d]	**0.0003**	**0.001**	**0.19**	**<0.0003**

[a] The total number of positions in the inferred single transmitted/founder virus that were different to the consensus CRF01AE sequence were counted
[b] The number of mutations in the inferred single transmitted/founder virus that resulted in reversions were counted
[c] The sum of all amino acid changes that were different from the inferred single transmitted/founder virus
[d] Paired t-test between *pol* and *env*

Table 6. Percent of mapped CTL epitopes.

PID	Total # of different amino acid changes in dataset	Number of amino acid changes in CTL epitopes from Con CRF01AE	% of all amino acid changes in the dataset fall within CTL epitope
		subgenomic *pol*	
3928	358	65	18.2
3832	128	83	64.8
5436	157	63	40.1
3698	142	95	66.9
6609	180	49	27.2
8123	156	77	49.4
3513	152	66	43.4
9114	390	76	19.5
3193	71	47	66.2
8522	273	54	19.8
Median	**157**	**66**	**41.8**
IQR	**139-294**	**53-79**	**19.7-65.2**
		subgenomic *env*	
3928	280	59	21.1
3832	151	82	54.3
5436	160	54	33.8
3698	175	105	60.0
6609	136	81	59.6
8123	174	102	58.6
3513	139	67	48.2
9114	152	91	59.9
3193	103	42	40.8
8522	152	66	43.4
Median	**152**	**74**	**51.3**
IQR	**138-174**	**58-94**	**39.1-59.7**
p-value[a]	**0.19**	**0.12**	**0.35**

[a] Paired t-test between *pol* and *env*

CONCLUSION

Genetic diversity in Fiebig II-III was ~6-fold lower than expected from the HIV-1 mutation rate over the first 21.5 days of infection, implying purifying selection during acute infection. While many transmitted amino acid mutations reverted to consensus CRF01AE; these mutations were a small contribution to the overall genetic diversity. The main drivers of early HIV-1 diversity were synonymous mutations and amino acid changes in CTL epitopes, suggesting the possibility of early CTL escape.