



**Supplemental Figure 1. Phylogram of all genital tract and plasma HIV *env* sequences from the eight study subjects generated using DIVEIN.** Each subject's sequences are color-coded, and cluster together. The phylogram shows no cross-contamination of viral sequences across subjects or extraneous sequences. Phylograms were rooted using representative reference sequences, indicated with the letter B or A, for the corresponding subtype from GenBank (Clade B: B.US.83.RF, B.US.90.WEAU160, B.FR.83.HXB2, B.US.86.JRFL; Clade A: A1 A1.KE.93.Q23-17, A1.SE.94.SE7253, A1.UG.85.U455, A1.UG.92.92UG037, 02AG.CM.99.pBD61, 02AG.NG.x.IBNG.L, A1.AU.03.PS1044D). The scale bar (horizontal line) indicates the horizontal branch length corresponding to 3 substitutions per 100 sites.

**Table S1. Number of HIV *env* sequences, monotypic viruses, and potential n-linked glycosylation sites from each subject by study visit and specimen type.**

Patient ID	Visit	Genital Tract Secretions				Plasma			
		Total # sequences <sup>1</sup>	Total # monotypic sequences <sup>2</sup>	% monotypic sequences	# pNLGs (median <sup>3</sup> )	Total # sequences <sup>1</sup>	Total # monotypic sequences <sup>2</sup>	% monotypic sequences	# pNLGs (median <sup>3</sup> )
1	1	8	5	63%	16	11	0	0%	16
	2	<b>16</b>	<b>10</b>	<b>63%</b>	14*	<b>14</b>	<b>0</b>	<b>0%</b>	16*
	3	10	0	0%	16*	22	7	32%	15*
2	1	<b>17</b>	<b>11</b>	<b>65%</b>	15*	<b>12</b>	<b>7</b>	<b>58%</b>	14*
	2	12	4	33%	15	10	0	0%	15
3	1	18	0	0%	15*	22	7	32%	14*
	2	<b>25</b>	<b>6</b>	<b>24%</b>	15	<b>19</b>	<b>0</b>	<b>0%</b>	15
4	1	10	0	0%	10*	17	2	12%	9*
	2	15	8	53%	9	20	3	15%	9
	3	23	12	52%	9	16	3	19%	9
5	1	10	0	0%	11	9	0	0%	11
	2	10	5	50%	11	11	3	27%	11
	3	9	3	33%	11	16	8	50%	11
6	1	12	0	0%	13*	10	0	0%	12*
	2	11	3	27%	13	14	3	21%	13
7	1	14	3	21%	14	17	9	53%	14
	2	20	8	40%	14*	14	0	0%	13*
8	1	9	2	22%	13	9	2	22%	13
	2	14	6	43%	12	13	4	31%	12

*Notes:*

<sup>1</sup> The number of sequences generated per tissue at each study visit from each subject.

<sup>2</sup> The number of monotypic viruses defined as those with 0% pair wise distances generated using PhyML. The frequencies between tissues were evaluated using a Wilcoxon rank sum test. Overall the number of monotypic virus between the genital tract and plasma were similar ( $p=0.22$ ). Visits without compartmentalization had a median of three monotypic sequences each in the genital tract and plasma, while visits with compartmentalization had a median of 10 monotypic sequences in the genital tract and 0 in the plasma ( $p=0.049$ ).

<sup>3</sup> The median number of pNLGs per sequence by subject and tissue site at each study visit. Comparisons of the number of pNLG between the plasma and genital tract within a visit are shown. Those pairs (i.e. V2 plasma vs. V2 genital tract) with significantly different frequencies in the number of pNLG between tissues ( $p<0.05$ , using a Wilcoxon rank sum test) are indicated by an \*.

**Bold** indicate the study visits with significant compartmentalization by all four tests of compartmentalization.