Supplementary Appendix

Contents

Supplementary Methods

Single genome amplification

Single genome amplification (SGA) was performed as previously described(1). cDNA was synthesized using the SuperScript III reverse transcriptase (Invitrogen) using the primer 1.R3.B3R 5'- ACTACTTGAAGCACTCAAGGCAAGCTTTATTG-3' (nt 9642-9611 in HXB2). To amplify the 3' half viral genome, the first round PCR was performed using the primers 07For7 5'-CAAATTAYAAAAATTCAAAATTTTCGGGTTTATTACAG-3' (nt 4875-4912) and 2.R3.B6R 5'-TGAAGCACTCAAGGCAAGCTTTATTGAGGC-3' (nt 9636–9607), and the second round PCR was performed with the primers VIF1 5'-GGGTTTATTACAGGGACAGCAGAG-3' (nt 4900-4923) and Low2c 5'-TGAGGCTTAAGCAGTGGGTTCC-3' (nt 9591-9612). Two microliters of the first round PCR products were used for the second round PCR amplification. The PCR thermocycling conditions were as follows: one cycle at 94^{o}C for 2 min; 35 cycles of a denaturing step at 94^{o}C for 15 sec, an annealing step at 60^{o}C for 30 sec, an extension step at 68^{o}C for 4 min; and one cycle of an additional extension at 68^{o}C for 10 min. All PCR amplifications were carried out using the Platinum™ Taq DNA Polymerase High Fidelity (Invitrogen). The PCR amplicons were directly sequenced by the cycle sequencing and dye terminator methods. Individual sequences were assembled and edited using Sequencher (Gene Codes). The sequences were aligned using the Gene Cutter in the Los Alamos HIV Sequence Database, followed by manual adjustment to obtain optimal alignment.

Determination of the number of T/F virus

The Poisson-Fitter tool in the Los Alamos HIV Sequence data base was used to determine the multiplicity of infection(2). Single T/F infection was determined by analyzing whether sequences from an infected individual exhibited Poisson distribution of mutations and star-like phylogeny after hypermutations with APOBEC3G/F signatures were excluded. Multiple T/F infection was determined when sequences from an individual deviated from both Poisson distribution of

mutations and star-like phylogeny when positions with APOBEC mutations were excluded. Transmission of multiple T/Fs was further confirmed by visual inspection of the highlighter plot, which showed distinct viral lineages with an inter-lineage distance higher than the maximum achievable distance given their Fiebig stage.

Virus isolation

To isolate viruses from HIV-1 infected individuals, PBMCs from a healthy donor were stimulated for 3 days in RPMI1640 containing 10% fetal bovine serum (FBS), interleukin 2 (IL-2) (50 U/ml; PeproTech), soluble anti-CD3 (0.5 μg/ml; eBioscience) and soluble anti-CD28 (0.5 μg/ml; eBioscience). After stimulation, the cells were washed twice with RPMI1640 to remove the antibodies. A total of 10^7 stimulated PBMCs from a healthy donor were then mixed with 10^7 of cryopreserved PBMCs from an HIV-1 infected individual. The cells were cultured in a T25 flask in RPMI1640 containing 10% FBS and 50 U/ml IL-2 for up to 4 weeks. Every 3 days, half volume of the culture supernatant was replaced with fresh medium. Every 7 days, half of the culture (including the cells) was removed, and a total of 5×10^6 of stimulated PBMCs from a healthy donor were added. The p24 concentration in the culture supernatant was measured every week. The viruses were harvested between two and four weeks when the p24 concentration in the culture supernatant achieved at least 5 ng/ml.

Pseudovirus preparation and titration

To generate HIV-1 envelope (*env*) clones, the *rev-vpu-env* cassette was PCR amplified from the corresponding SGA products or chemically synthesized (GenScrip) and cloned into the expressing vector pcDNA3.3-TOPO (Invitrogen). The desired mutations were introduced by site-directed mutagenesis. All envelope clones were confirmed by sequencing. The pseudovirus stocks were prepared as previously described(3). In brief, 2 μg of *env* clone was co-transfected with 4 μg of the pNL4.3-ΔEnv-vpr+-luc+ into 293T cells (ATCC CRL3216) in a T25 flask using the

FuGENE6 transfection reagent (Promega). The cells were cultured at 37°C for 6 hours before the medium was completely replaced with fresh medium. The culture supernatants containing the pseudoviruses were harvested at 72 hours post transfection, aliquoted and stored at −80°C until use. The 293T cell line from ATCC was not validated further.

Next generation sequencing

The library for sequencing on Illumina MiSeq was prepared using a nested PCR approach as described previously(4). The first round PCR was carried out using the forward primer 44F 5'-ACAGTRCARTGYACACATGG-3' (nt 6954-6973) and the reverse primer 35R 5'-CACTTCTCCAATTGTCCITCA-3' (nt 7648-7668). The first round PCR conditions were as follow: one cycle at 94°C for 2 min; 30 cycles of a denaturing step at 94°C for 15 sec, an annealing step at 55°C for 30 sec, an extension step at 68°C for 1 min; and one cycle of an additional extension at 68°C for 10 min. A total of 5 µL of first round PCR products were used for second round PCR amplification. The second round PCR was carried out using a panel of primers containing the Illumina index and adaptors as previously described(4), using the following PCR conditions: one cycle at 94°C for 2 min; 5 cycles of a denaturing step at 94°C for 15 sec, an annealing step at 55°C for 30 sec, an extension step at 68°C for 1 min; and another 30 cycles of a denaturing step at 94°C for 15 sec, an annealing step at 60°C for 30 sec, an extension step at 68°C for 1 min; and one cycle of an additional extension at 68°C for 10 min. The Platinum™ Taq DNA Polymerase High Fidelity (Invitrogen) was used for all PCR amplification. The second round PCR products were gel purified, and the final library was generated by pooling the second round PCR product from each sample with an equal concentration. To increase the diversity of the library, 40% of the PhiX was spiked in (Illumina). The library was sequenced on the Illumina MiSeq using the MiSeq Reagent Kit v3 (Illumina). Deep sequencing was performed in duplicate for each sample.

Next generation sequencing data analysis

The raw fastq reads in files “read 1” and “read 2” were merged using the FLASH software(5). The merged fastq files were then filtered based on data quality using the following parameters: no more than 10 base calls with Q score lower than 30 in each read. After filtering, identical reads in each sample were collapsed to generate the unique sequences. Singletons (sequences which only appeared once in a sample) were excluded from the downstream analysis. To minimize the possibility that a unique sequence was generated due to sequencing error, deep sequencing was performed in duplicate for each sample. Only the deep sequencing reads that appeared in both experiments were used for analysis. The deep sequencing data was processed and analyzed using publicly available software and tools on the Galaxy platform(6).

CD4 subset analysis and sorting

PBMCs from HIV-1 infected individuals were stained by the following antibodies: CD3-Brilliant Violet 605 (clone OKT3, BioLegend), CD4-PerCP-Cy5.5 (clone OKT4, Biolegend), CCR7-PE-CF594 (clone 2-L1-A, BD Biosciences), CD27-PE (clone M-T271, BioLegend), CD45RO-APC (clone UCHL1, BioLegend). The cells were then stained by the Live-dead aqua prior to flow analysis to exclude the dead cells (Invitrogen). The stained cells were sorted on a BD FACSAria II cell sorter (BD Biosciences). Four CD4 subsets were defined as follows: naïve ($CD45RO^-$, $CCR7^+$, and $CD27^+$), central memory ($CD45RO^+$, $CCR7^+$, and $CD27^+$), transitional memory ($CD45RO^+$, $CCR7^-$, and $CD27^+$) and effector memory ($CD45RO^+$, $CCR7^-$, and $CD27^-$). The purity of each sorted subset was higher than 95%. All commercial antibodies were validated by the vendors.

Determination of CCR5 and CXCR4 expression

To determine the level of CCR5 and CXCR4 expression on each CD4 subset, PBMCs were stained with the following antibodies: CD3-Brilliant Violet 605 (clone OKT3, BioLegend), CD4-

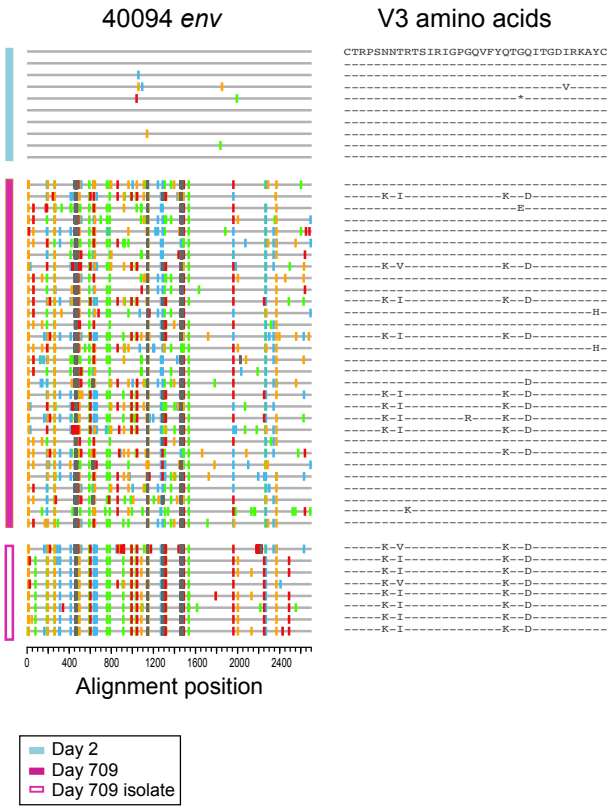
PerCP-Cy5.5 (clone OKT4, BioLegend), CCR7-PE-CF594 (clone 2-L1-A, BD Biosciences), CD27-FITC (clone M-T271, BioLegend), CD45RO-APC (clone UCHL1, BioLegend), CCR5-PE (clone J418F1, BioLegend) or CXCR4-PE (clone Q18A64, BioLegend). The CCR5 and CXCR4 staining antibodies were titrated to determine the optimal concentration. To evaluate the non-specific staining, fluorescent stanning was cold-inhibited by a 100-fold excess of the unlabeled CCR5 or CXCR4 antibody (the same clone as the labeled antibody) mixed with the respective labeled antibody. A fluorescence minus one (FMO) staining was also determined for CCR5/CXCR4 staining. The highest concentration of the labeled antibody with which the cold inhibition showed virtually overlapping staining with the FMO was used to quantify the levels of CCR5 and CXCR4 expression on each CD4 subset. All flow data were collected using the DIVA 7.0 software on the FACSAria II (BD Biosciences) cell sorter/analyzer and analyzed by the FlowJo software (FlowJo LLC, Ashland, OR). All commercial antibodies were validated by the vendors.

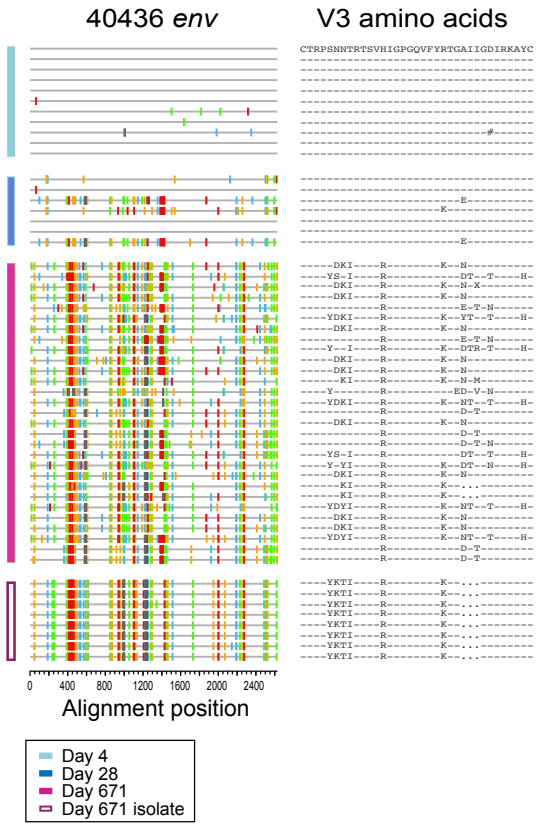
Quantification of cell associated HIV-1 RNA

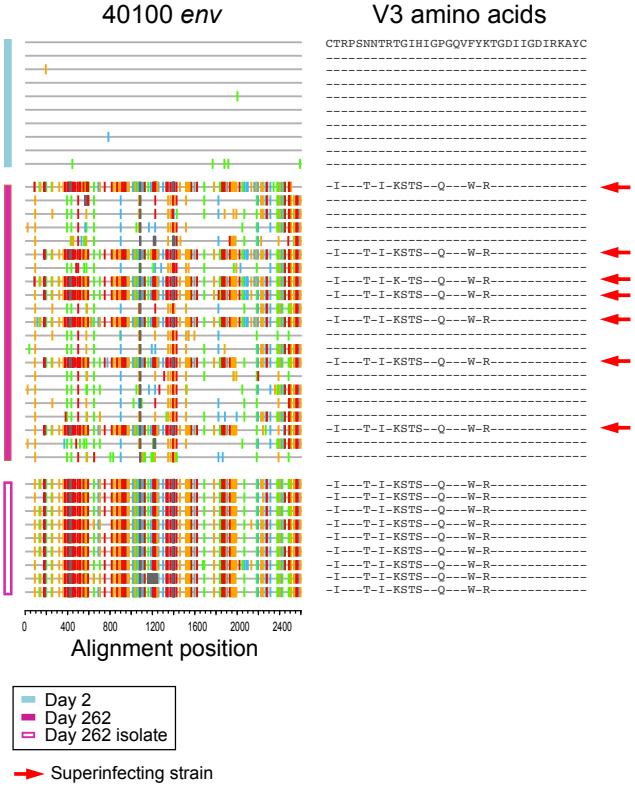
Cell associated HIV-1 RNA was quantified for each sorted CD4 subset by amplifying part of the *pol* gene. RNA was extracted from the sorted cells using the RNeasy Mini kit (Qiagen). A total of 8.5 µL extracted RNA was subjected to one-step RT-PCR using the Superscript III one-step RT-PCR system (Invitrogen). The one-step RT-PCR was performed using the forward primer Pol F1 5'-TACAGTGCAGGGGAAAGAATA-3' (nt 4809-4829) and the reverse primer Pol R1 5'-CTTCTTGGCACTACTTTTATGTCAC-3' (nt 4993-5017). The PCR conditions were as follow: a reverse transcription step at 50^{o}C for 1h; A denaturing step at 94^{o}C for 2 min; 16 cycles of a denaturing step at 94^{o}C for 15 sec, an annealing step at 55^{o}C for 30 sec, an extension step at 68^{o}C for 1 min, and one cycle of an additional extension at 68^{o}C for 5 min. The first round PCR products were diluted 10-fold and a total of 6.4 µL of diluted PCR products were used for the real-time PCR using the forward primer Pol F1, the reverse primer Pol R2 5'-CTGCCCCTTCACCTTTCC-3' (nt 4957-4974), and the probe Pol Famzen: 5'-/56-

FAM/TTTCGGGTT/ZEN/TATTACAGGGACAGCAG/3IABkFQ/-3' (nt 4896-4921). The real-time PCR was performed on the QuantStudio 3 Real-Time PCR Systems (Thermo Fisher Scientific) using the following conditions: A denaturing step at 94^{o}C for 4 min, 45 cycles of a denaturing step at 94^{o}C for 3 sec, an annealing and extension step at 60^{o}C for 20 sec. The copy number of the input RNA was determined by using the RNA standard generated by *in vitro* transcription. In brief, the amplicon region (we used the CRF01_AE consensus sequence in this study) was cloned into the pUC57 vector downstream of the T7 promoter. The DNA fragment containing the amplicon was PCR amplified and the RNA was generated by *in vitro* transcription using the MEGAscript T7 Transcription Kit (Invitrogen).

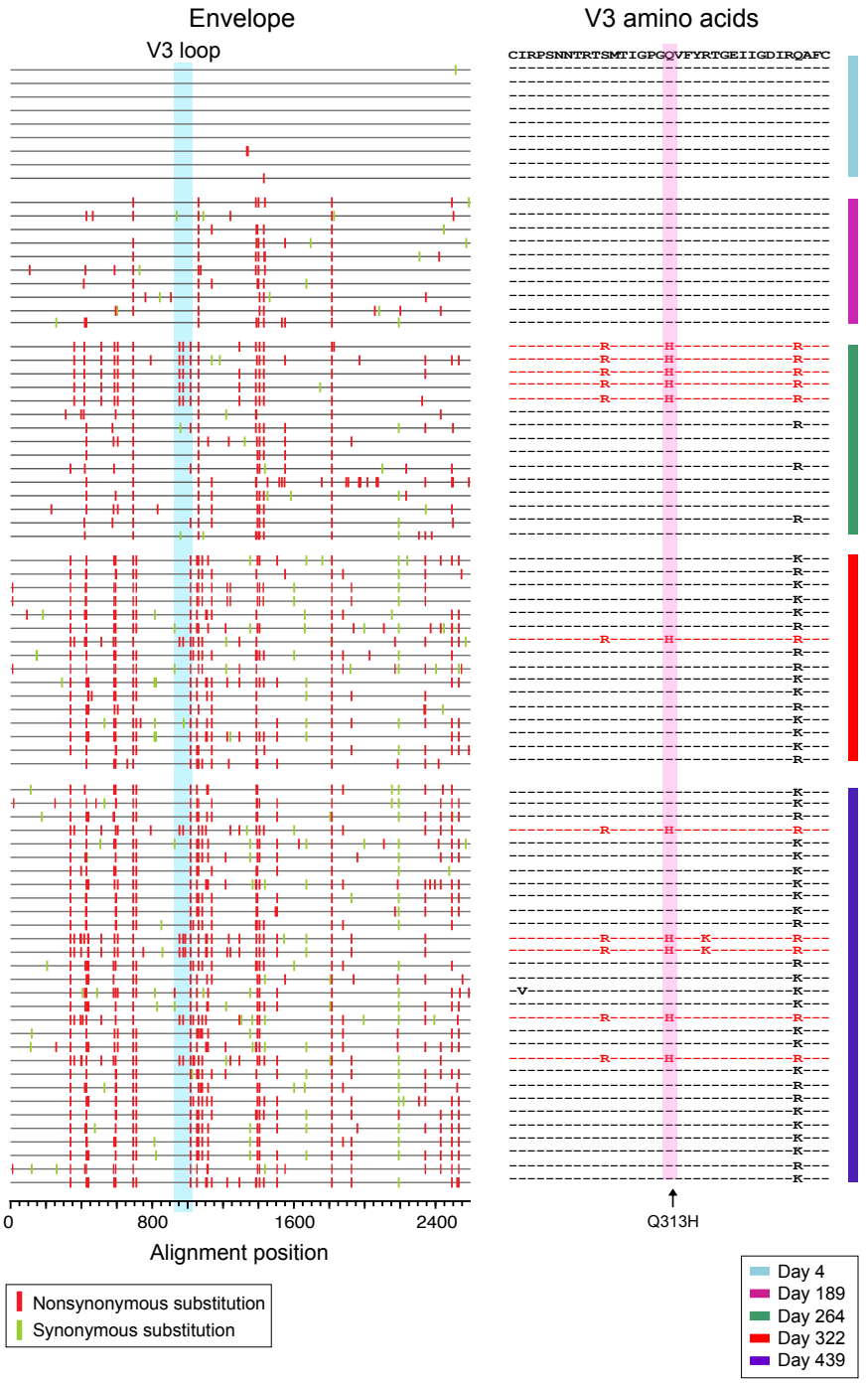
Supplementary Figures



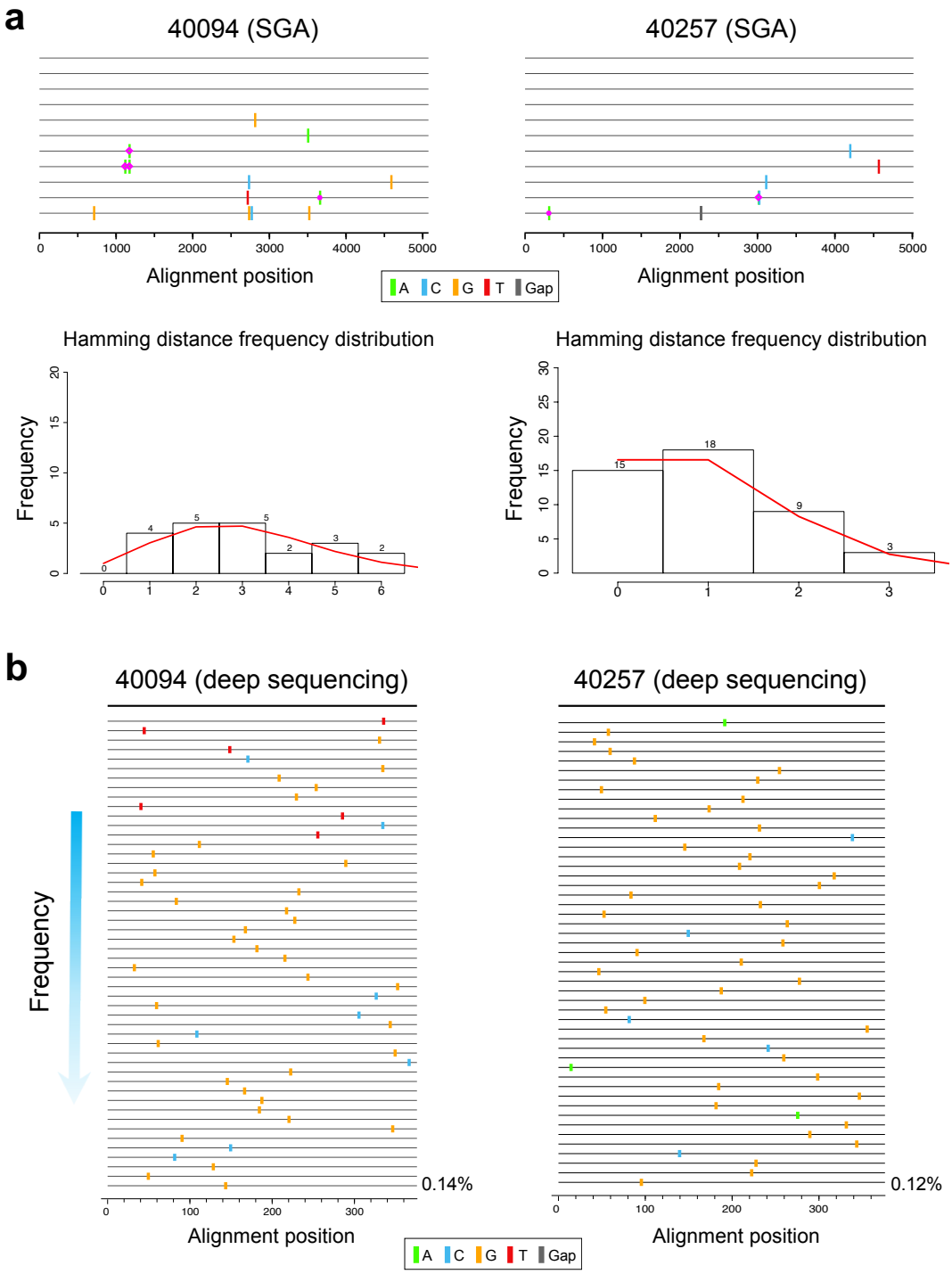




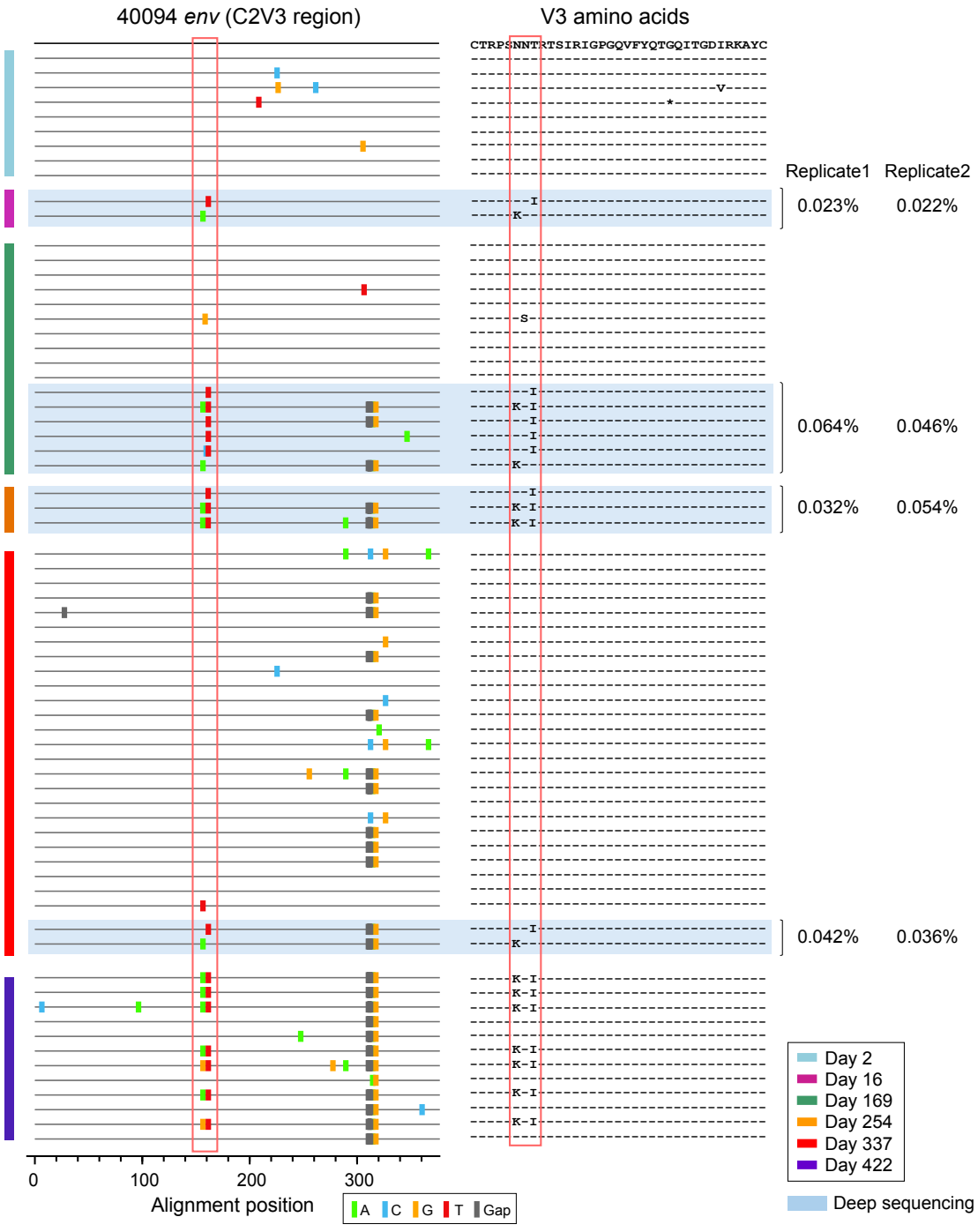
Supplementary Figure 1: The SGA sequences of the X4-using isolates from participants 40094, 40436 and 40100 are aligned to the SGA sequences of the plasma viruses. Sequences from different time points are color coded. Participant 40100 was initially infected by a single, R5 tropic T/F virus and was super-infected by an X4 strain at day 262. The sequences of the superinfecting strain are indicated by red arrows. The sequences of the X4-using isolate matched the super-infecting strain. Viral isolation was not performed for participant 40257 due to PBMC availability. For 40257, the existence of X4 variants was identified using envelope (*env*) clones (Supplementary Table 3).



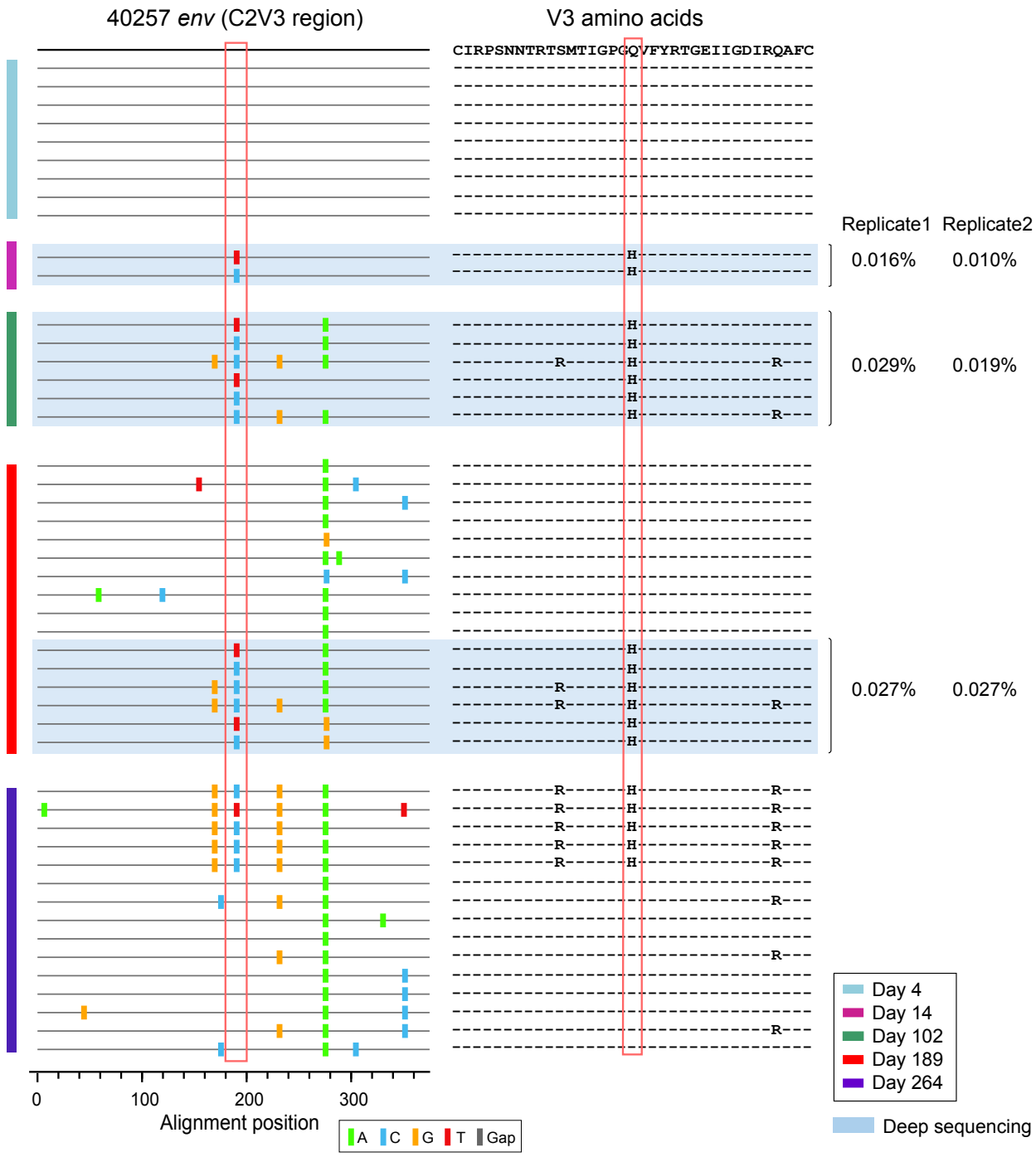
Supplementary Figure 2: The longitudinal evolution of the 40257 T/F virus is illustrated by the highlighter plot. Sequences from different time points are color coded. The sequence on the top represents the T/F virus. The red and green tics indicate non-synonymous and synonymous substitutions compared to the T/F virus, respectively. The V3 loop is highlighted in blue. In the V3 amino acids alignment (right), the position of the driver mutation (313) is shaded in red, and the driver mutation (Q313H) is indicated by black arrow. The X4 variants carrying the driver mutation are shown in red, while the R5 variants are shown in black.



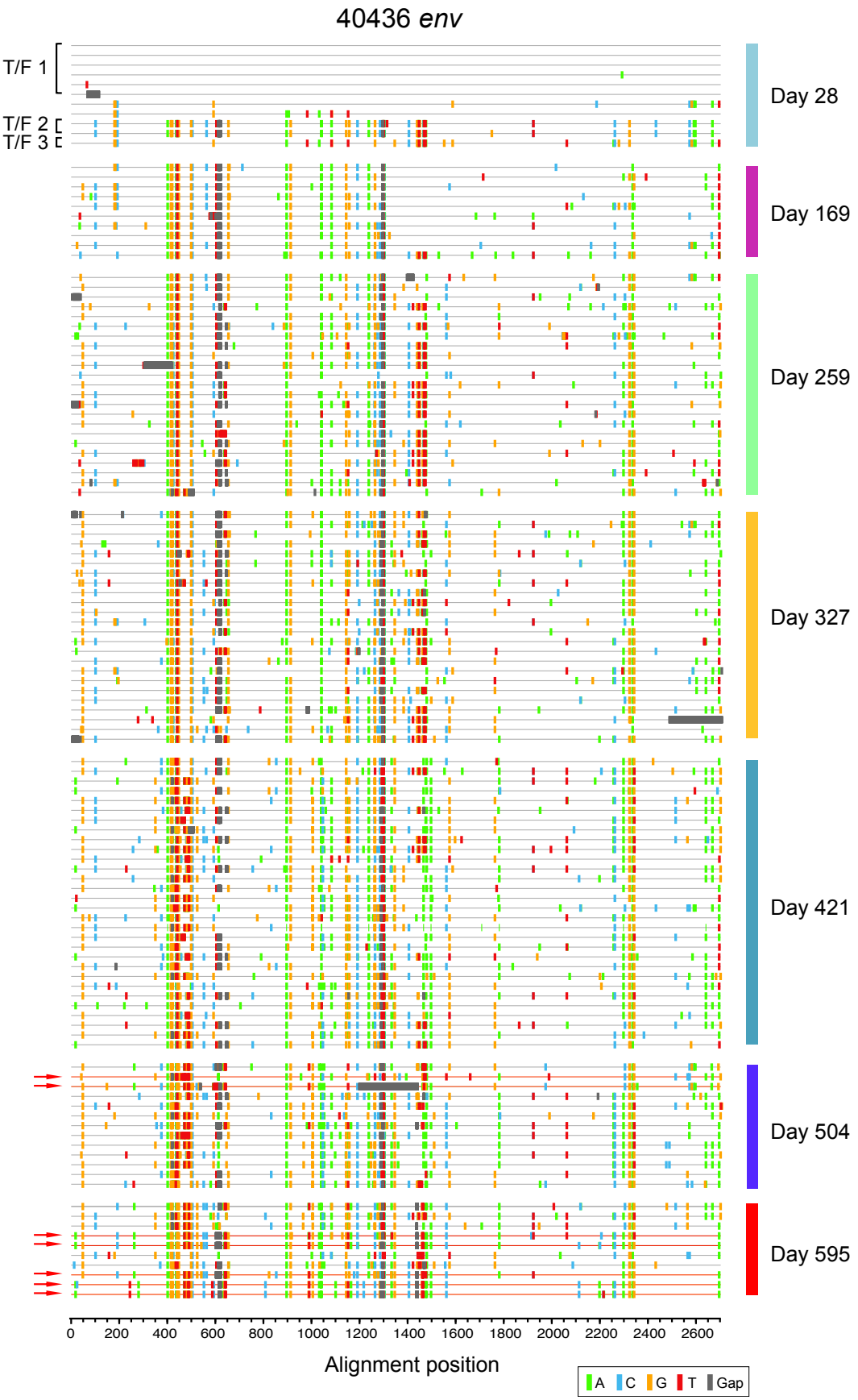
Supplementary Figure 3: (**a**) SGA sequences obtained from acute infection were used to determine the number of T/F virus by using the Poisson-Fitter tool in the Los Alamos HIV Sequence Database. For both participants, the Hamming Distance (HD) frequency distributions follow Poisson distribution, and the sequences follow a star-phylogeny. (**b**) Deep sequencing of the C2V3 region detected a single viral lineage during acute infection. The sequences are ordered according to their frequency. A total of 50 unique sequences are shown. The T/F sequence is used as the master sequence (the black line on the top). The frequency of the last sequence in the highlighter plot is shown.



Supplementary Figure 4: Evolution of the viral variants carrying the driver mutations in participant 40094. The deep sequencing reads carrying the driver mutations (shaded in blue) are aligned to plasma viral sequences obtained by SGA. Sequences from different time points are color coded. The mutations responsible for coreceptor switch are indicated in red boxes. Deep sequencing was carried out in duplicate. The frequency of the deep sequencing reads is shown on the right.

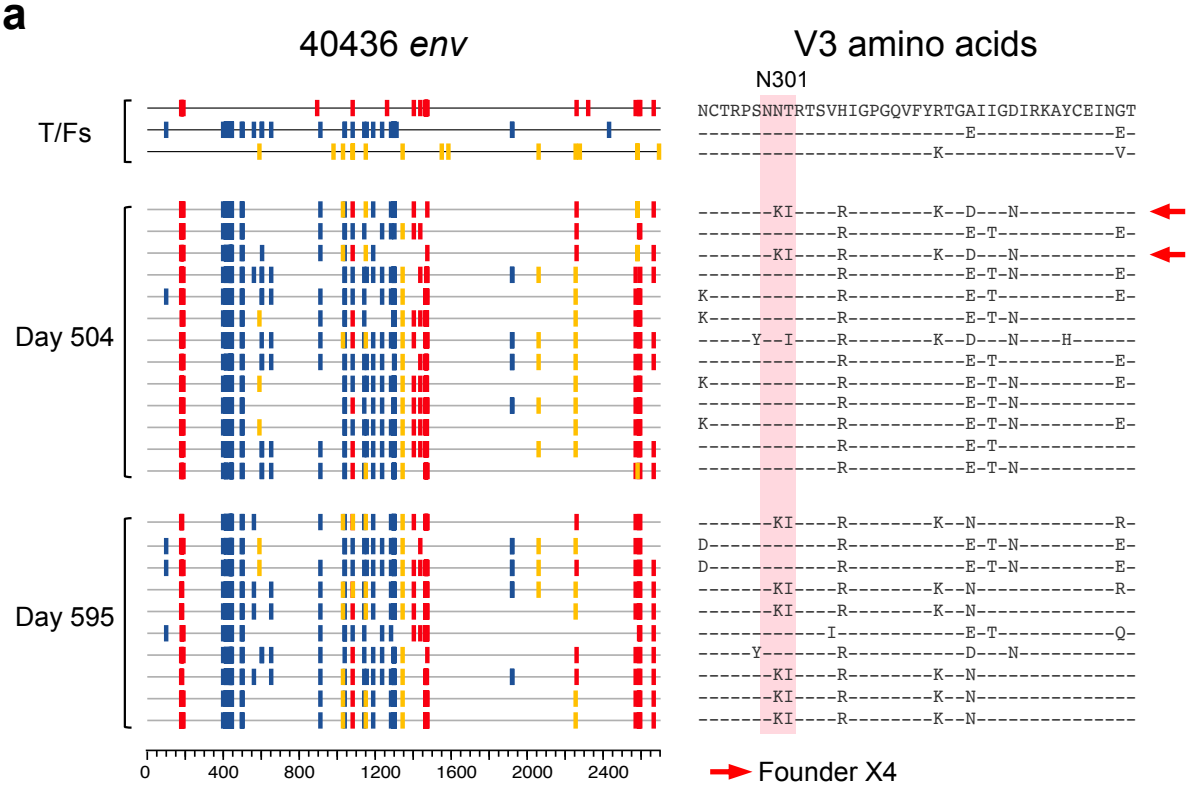


Supplementary Figure 5: Evolution of the viral variants carrying the driver mutations in participant 40257. The deep sequencing reads carrying the driver mutations (shaded in blue) are aligned with plasma viral sequences obtained SGA. Sequences from different time points are color coded. The mutations responsible for coreceptor switch are indicated in red boxes. Deep sequencing was carried out in duplicate. The frequency of the deep sequencing reads is shown on the right.

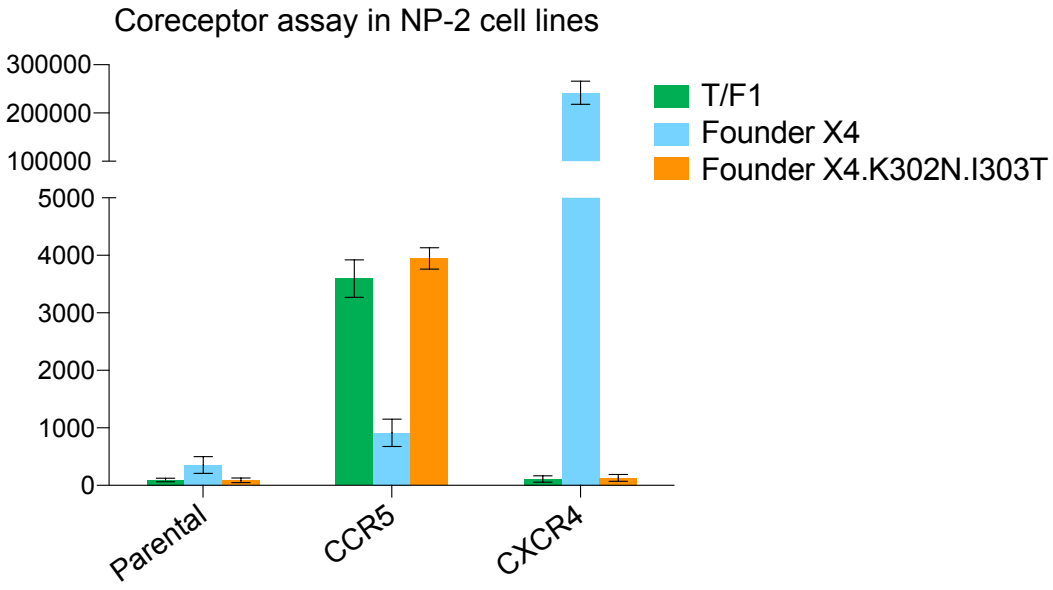

40436 *env*
T/F 1
T/F 2
T/F 3
Day 28
Day 169
Day 259
Day 327
Day 421
Day 504
Day 595
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
2600
Alignment position
A
C
G
T
Gap

Supplementary Figure 6: Longitudinal evolution of the envelope sequences in participant 40436. Participant 40436 was infected by three different T/F lineages. SGA sequences from different time points are color coded. The X4-using variants, which emerged at day 504, are shown in red and are indicated by red arrows.

a

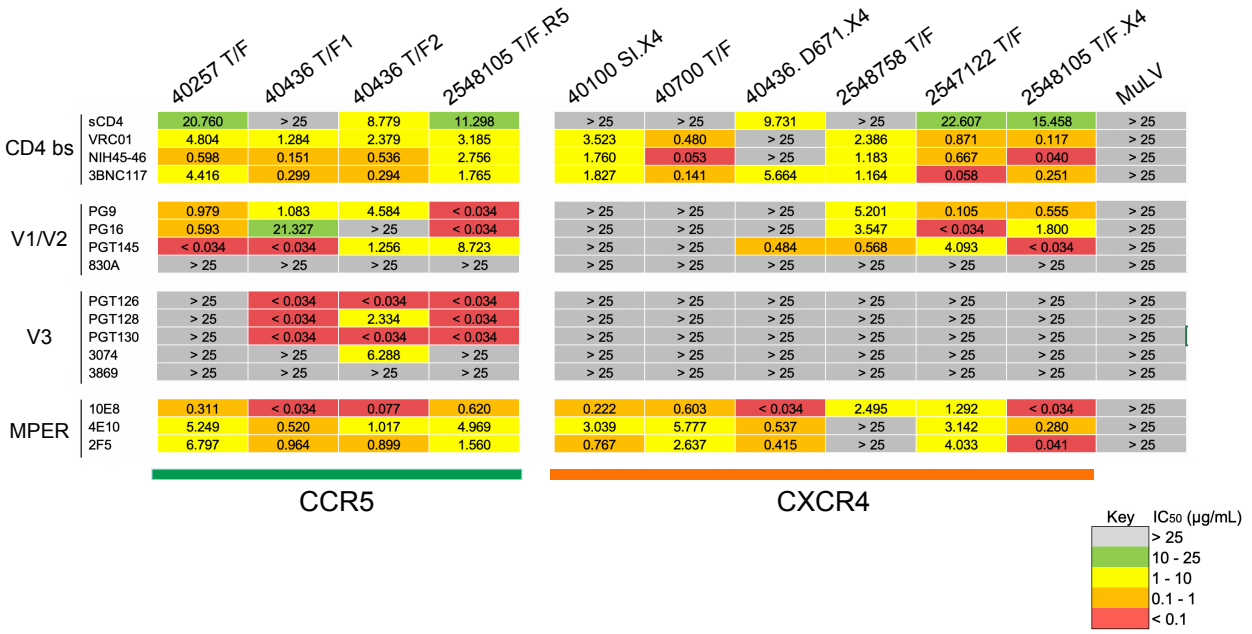


b

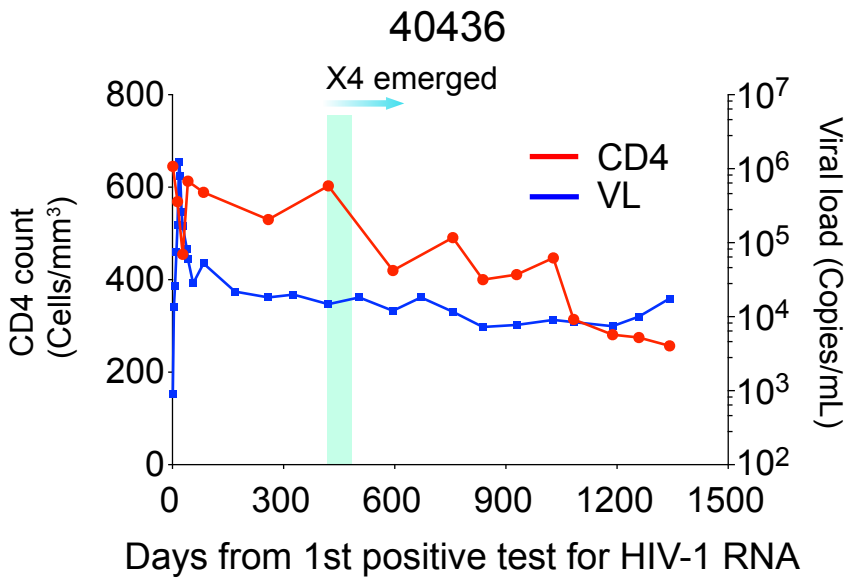


PGT128 IC50 (µg/mL)	
Founder X4	>25
Founder X4.K302N.I303T	0.240

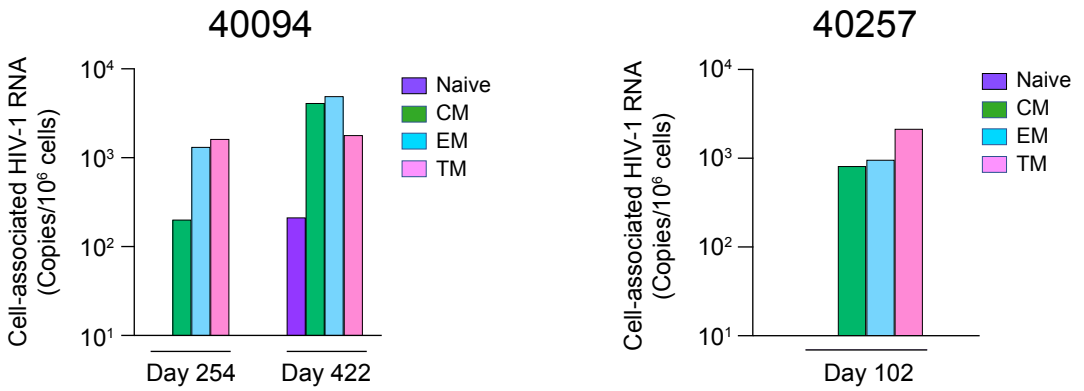
Supplementary Figure 7: (a) Recombination analysis of the envelope sequences from day 504 and day 595 in participant 40436 (left). The T/F viruses are used as reference sequences. Unique genetic signatures in each T/F virus are shown in red, blue, and yellow, respectively. In the V3 amino acid alignment (right), the N301 glycan site where the driver mutations emerged is shaded in red. The two earliest X4 variants carrying mutations at the N301 glycan site are indicated by red arrows. (**b**) Coreceptor usage and neutralization sensitivity to PGT128 of the founder X4 virus and the K302N.I303T mutant which had the N301 glycan restored. All coreceptor assays were performed in triplicate. The error bar indicates the SD.



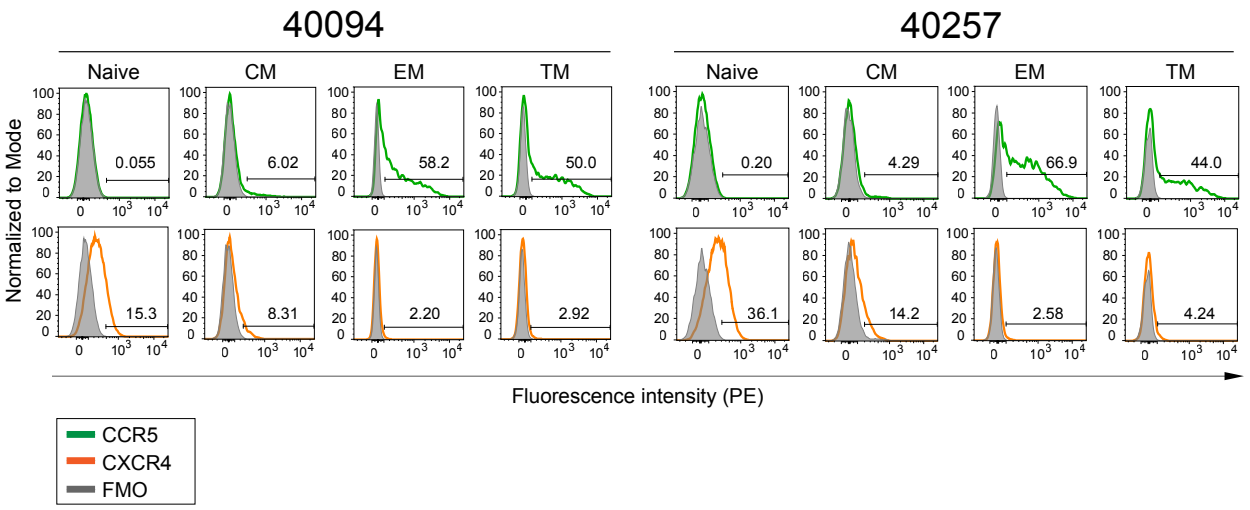
Supplementary Figure 8: Neutralization sensitivity of R5 and X4 HIV-1 to broadly neutralizing antibodies (bNAbs). The CCR5 group includes four R5 tropic T/F viruses identified in the RV217 and RV254 cohorts. The CXCR4 group includes six phenotypically confirmed X4-using viruses, including the superinfecting strain in 40100, the X4-tropic T/F virus in 40700, an X4 variant isolated from 40436, as well as three X4-using T/F viruses identified in participants 2548758, 2547122 and 2548105. Among them, the 40700 T/F virus uses CXCR4 exclusively, while the other five viruses have low-level R5 using ability. The IC_{50} (µg/mL) of each bNAb is shown. The MuLV was used as the negative control.



Supplementary Figure 9: CD4 and viral load dynamics in participant 40436. The time frame when the earliest/founder X4 viruses became datable in plasma by SGA is highlighted in green.

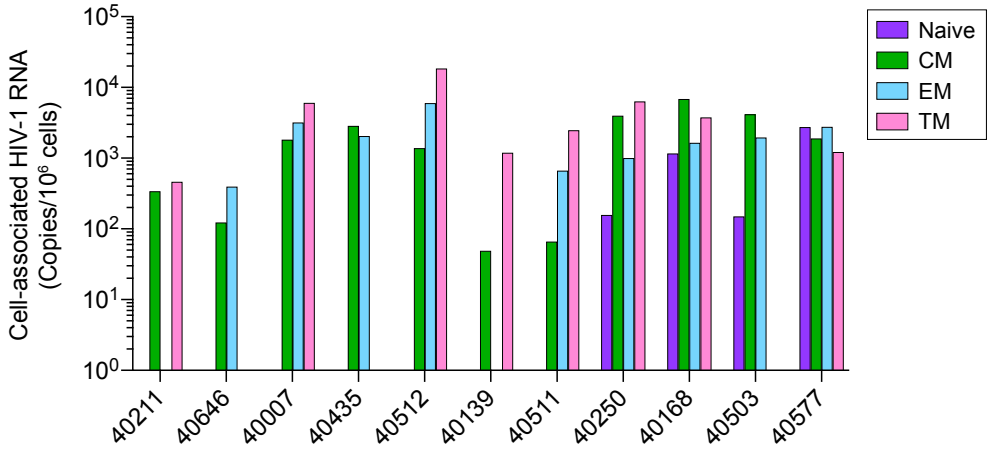


Supplementary Figure 10: Quantification of cell associated HIV-1 RNA in participants 40094 and 40257. Cell associated HIV-1 RNA in different CD4 subsets were quantified in participant 40094 at day 254 (before coreceptor switch) and day 422 (the earliest time point of coreceptor switch). The cell associated RNA was undatable in the naïve subset at day 254. Cell associated HIV-1 RNA in different CD4 subset were determined in participant 40257 at day 102 (before coreceptor switch). The cell associated RNA was undetectable in the naïve subset.

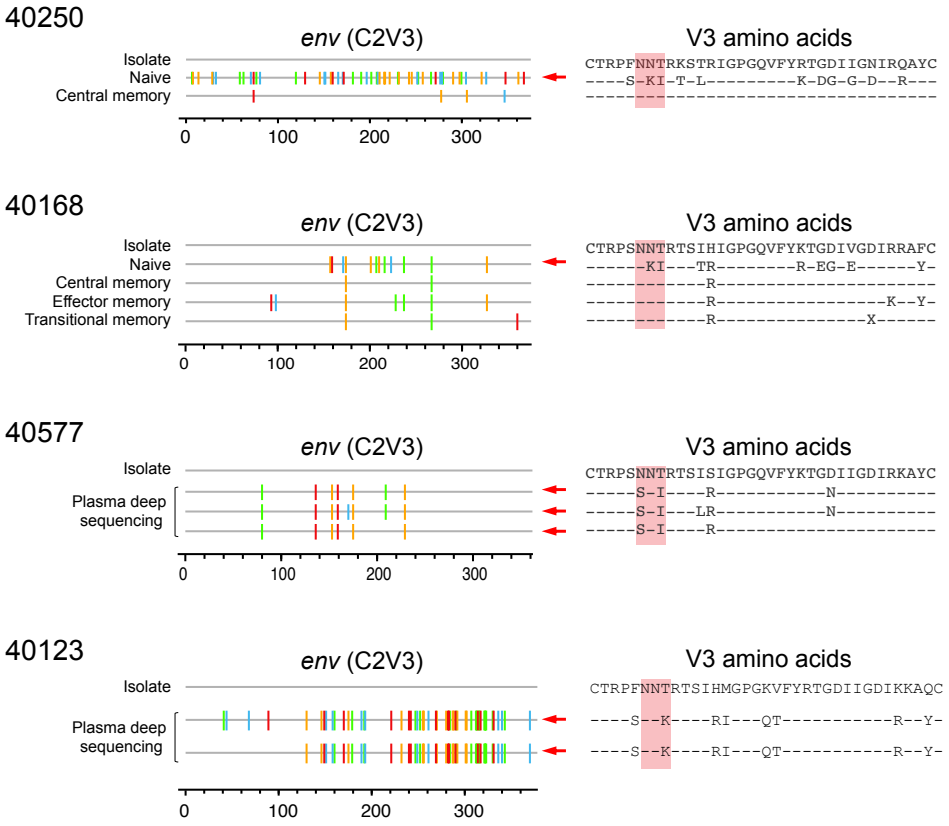


Supplementary Figure 11: Quantification of CCR5 and CXCR4 expression on each CD4 subset by flow cytometry. Positive cells are shown by percentage in each figure.

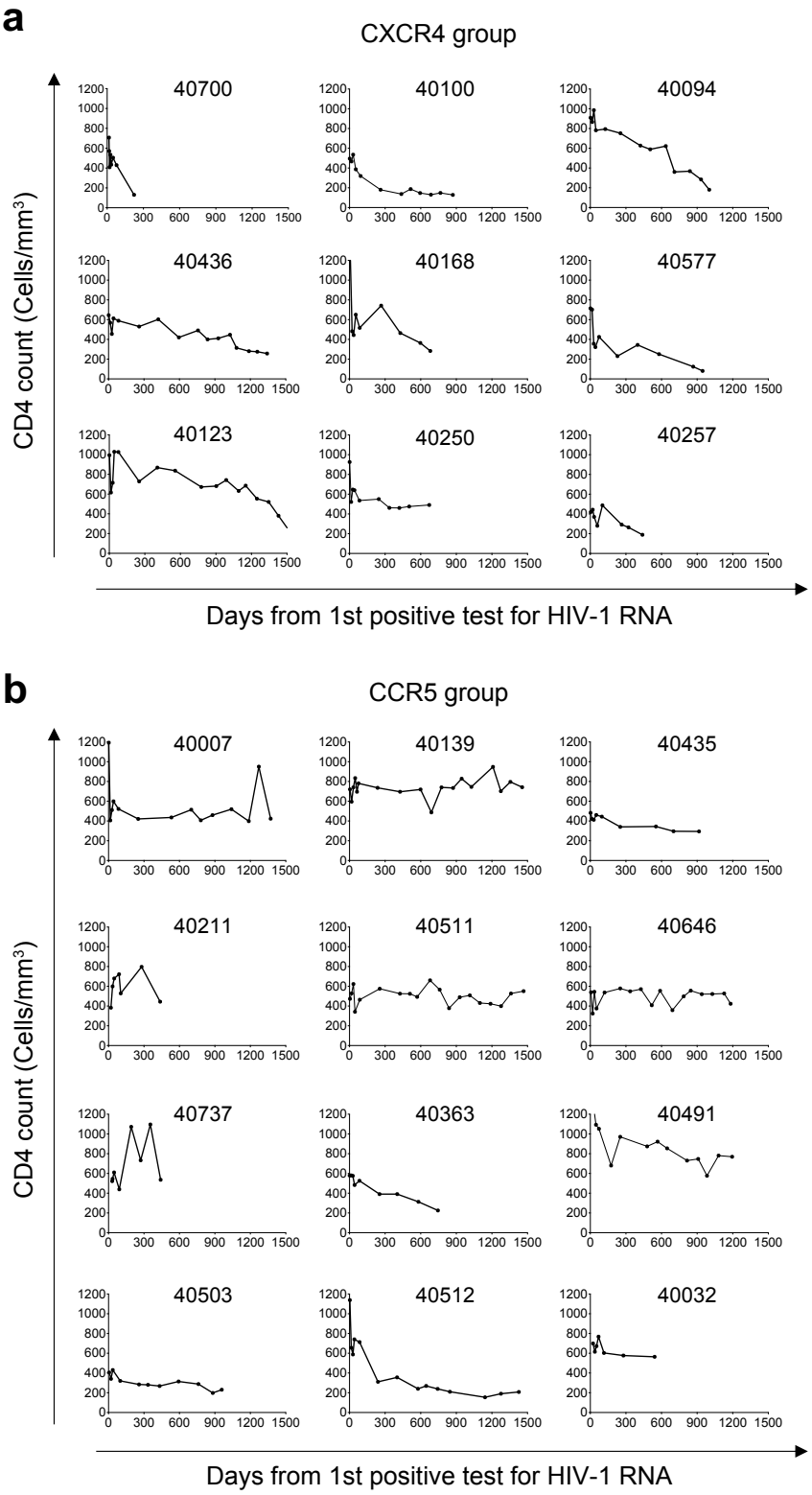
a



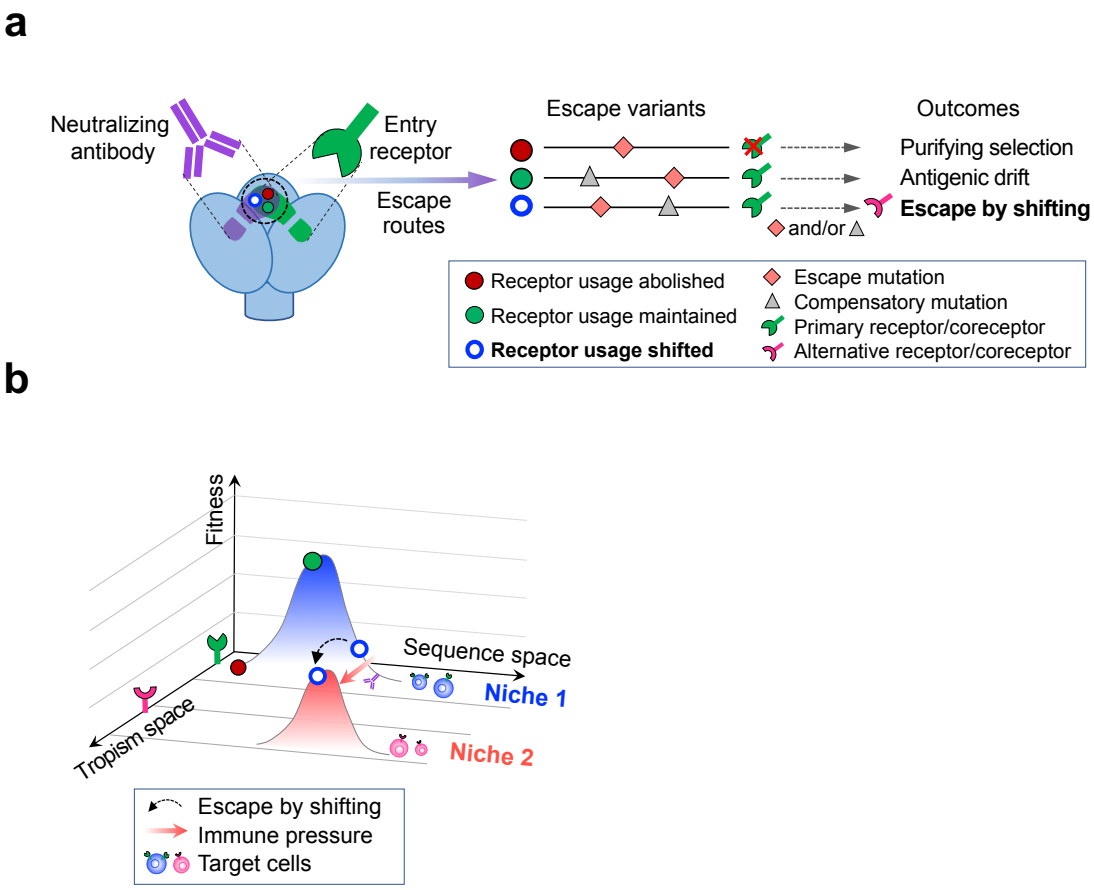
b



Supplementary Figure 12: (**a**) Cell associated HIV-1 RNA in different CD4 subsets of 11 participants for whom the primary viral isolates showed pure R5 phenotype. (**b**) X4 viruses were detected in naïve CD4+ T cells for participants 40250 and 40168. A low frequency of X4 variants was detected in plasma in participants 40577 and 40123 by deep sequencing. The phenotypically confirmed X4 variants are indicated by red arrows.



Supplementary Figure 13: (**a**) CD4 dynamics of participants harboring X4 viruses. Participant 40700 was infected by an X4 T/F HIV-1 without CCR5 using ability. (**b**) CD4 dynamics of participants without the evidence of harboring X4 variants. The longitudinal data from the earliest available time point to the last available time point before ART initiation is shown.



Supplementary Figure 14: Conceptual framework of the "escape by shifting" concept. (**a**) When the footprint of a neutralizing antibody overlaps with virus receptor binding site, receptor usage alteration functions as an evolutionary route of immune evasion. (**b**) Escape variants explore receptor tropism space, sequence space and fitness landscape under host immune pressure. Alteration in receptor usage allows an escape variant to find its fitness peak in a novel niche.

Supplementary Tables

Supplementary Table 1. Characteristics of study participants of the RV217 and RV254 cohorts

Participant ID	Cohort	Country	Subtype	Fiebig Stage	Gender
40007	RV217	Thailand	CRF01_AE	I-II	Male
40032	RV217	Thailand	CRF01_AE	I-II	Female
40094	RV217	Thailand	CRF01_AE	I-II	Male
40100	RV217	Thailand	CRF01_AE	I-II	Transgender
40123	RV217	Thailand	CRF01_AE	I-II	Male
40139	RV217	Thailand	CRF01_AE	V-VI	Transgender
40168	RV217	Thailand	CRF01_AE/B	I-II	Transgender
40211	RV217	Thailand	CRF01_AE	I-II	Male
40250	RV217	Thailand	CRF01_AE	I-II	Male
40257	RV217	Thailand	CRF01_AE	I-II	Transgender
40363	RV217	Thailand	CRF01_AE	I-II	Transgender
40435	RV217	Thailand	CRF01_AE	IV	Male
40436	RV217	Thailand	CRF01_AE	I-II	Male
40491	RV217	Thailand	CRF01_AE	I-II	Transgender
40503	RV217	Thailand	CRF01_AE	I-II	Male
40511	RV217	Thailand	CRF01_AE	I-II	Transgender
40512	RV217	Thailand	CRF01_AE	III	Male
40577	RV217	Thailand	CRF01_AE	I-II	Male
40646	RV217	Thailand	CRF01_AE	I-II	Male
40700	RV217	Thailand	CRF01_AE	IV	Male
40737	RV217	Thailand	CRF01_AE	I-II	Male
2547122	RV254	Thailand	CRF01_AE	III	Male
2548758	RV254	Thailand	CRF01_AE/B	IV	Male
2548105	RV254	Thailand	CRF01_AE	V	Male

No males or females were excluded during the design of the study. Gender was self-reported by the study participants.

Supplementary Table 2. Information of monoclonal neutralizing antibodies

Reagent	Binding domain	Specificity	Source	Product Number
sCD4	gp120	CD4bs	HIV Reagent Program	ARP-4615
VRC01	gp120	CD4bs	HIV Reagent Program	ARP-12033
NIH45-46	gp120	CD4bs	HIV Reagent Program	ARP-12174
3BNC117	gp120	CD4bs	HIV Reagent Program	ARP-12474
PG9	gp120	V1V2	Polymun Scientific	AB015
PG16	gp120	V1V2	Polymun Scientific	AB016
PGT145	gp120	V1V2	HIV Reagent Program	ARP-12703
830A	gp120	V1V2	Susan Zolla-Pazner (Icahn School of Medicine at Mount Sinai)	
PGT126	gp120	V3	HIV Reagent Program	ARP-12344
PGT128	gp120	V3	HIV Reagent Program	ARP-13352
PGT130	gp120	V3	Devin Sok (International AIDS Vaccine Initiative)	
3074	gp120	V3	HIV Reagent Program	ARP-12040
3869	gp120	V3	HIV Reagent Program	ARP-12039
10E8	gp41	MPER	HIV Reagent Program	ARP-12294
4E10	gp41	MPER	Polymun Scientific	AB004
2F5	gp41	MPER	Polymun Scientific	AB001

Supplementary Table 3. Determination of coreceptor usage of 20 participants in the RV217 cohort

Participants ID	Days since first positive test for HIV-1 RNA	Coreceptor usage
40094	709	X4-using
40436	671	X4-using
40257	439	X4-using
40100	262	X4-using
40007	700	R5
40032	615	R5
40123	684	R5
40139	690	R5
40211	360	R5
40250	672	R5
40363	676	R5
40435	701	R5
40491	647	R5
40503	761	R5
40511	679	R5
40512	743	R5
40577	579	R5
40646	692	R5
40737	441	R5
40168	681	R5

Coreceptor usage was determined using primary viral isolates except for participant 40257, for whom the existence of X4 variants was identified by using the envelope (*env*) clones. The primary viral isolates of 40094, 40436, 40100, as well as the *env* clones from 40257 used CXCR4 with high efficiency while had low level CCR5 using ability (Participant 40700 who was initially infected by a CXCR4 tropic T/F virus was not listed on this table). The second column shows the time (days from the first positive test for HIV-1 RNA) when the coreceptor assays were performed.

Supplementary Table 4. Infectivity of the 40257 T/F pseudovirus and its variants

Variant	$TCID_{50}$ / p24
40257 T/F	8.4
S306R	1.0
Q313H	3.4
Q327R	0.7
S306T/Q313H/Q327R	0.1
S306T/Q313H/Q327R.V1V2	6.8

The infectivity of the pseudovirus was shown as the ratio of $TCID_{50}$ ($TCID_{50}$/mL) and p24 (ng/mL). The S306T/Q313H/Q327R.V1V2 variant contains the V1/V2 regions of the founder X4 virus.

Supplementary Table 5. GenBank accession numbers of previously submitted sequences and newly generated sequences

Previously submitted sequences	GenBank accession number
40094 visit 1 (day 2)	MN792035, MN792037, MN792039, MN792041, MN792043, MN792045, MN792047, MN792048, MN792051, MN792053, OM825813
40094 visit 9 (day 30)	MN792055-MN792065
40094 visit 14 (day 169)	MN792066-MN792075
40094 visit 20 (day 709)	OM825816-OM825845
40100 visit 1 (day 2)	KY580582-KY580591
40100 visit 15 (day 262)	OM825866-OM825886
40257 visit 1 (day 4)	MN792312, MN792314, MN792316, MN792317, MN792319, MN792321, MN792323, MN792325, MN792327, MN792329
40257 visit 14 (day 189)	MN792360
40257 visit 17 (day 439)	OM826197-OM826226
40436 visit 2 (day 4)	KU230426, KY580684-KY580693
40436 visit 9 (day 28)	KY580694-KY580700, MN792466-MN792469
40436 visit 14 (day 169)	KY580701-KY580710
40436 visit 17 (day 421)	OM826337-OM826366
40436 visit 20 (day 671)	OM826368-OM826396
Newly generated sequences	**GenBank accession number**
40094 visit 16 (day 337)	OQ286397-OQ286421
40094 visit 17 (day 422)	OQ286422-OQ286433
40094 visit 18 (day 505)	OQ286434-OQ286446
40094 visit 19 (day 636)	OQ286447-OQ286457
40257 visit 15 (day 264)	OQ286475-OQ286489
40257 visit 16 (day 322)	OQ286490-OQ286505
40436 visit 15 (day 259)	OR287360-OR287382
40436 visit 16 (day 327)	OR287383-OR287406
40436 visit 18 (day 504)	OR287407-OR287419
40436 visit 19 (day 595)	OR287420-OR287429
40094 primary isolate (visit 20, day 709)	OQ286458-OQ286465
40436 primary isolate (visit 20, day 671)	OQ286506-OQ286514
40100 primary isolate (visit 15, day 262)	OQ286466-OQ286474

References:

1. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A. 2008;105(21):7552-7.
2. Giorgi EE, Funkhouser B, Athreya G, Perelson AS, Korber BT, Bhattacharya T. Estimating time since infection in early homogeneous HIV-1 samples using a poisson model. BMC Bioinformatics. 2010;11:532.
3. Marichannegowda MH, Song H. Immune escape mutations selected by neutralizing antibodies in natural HIV-1 infection can alter coreceptor usage repertoire of the transmitted/founder virus. Virology. 2022;568:72-6.
4. Song H, Ou W, Feng Y, Zhang J, Li F, Hu J, et al. Disparate impact on CD4 T cell count by two distinct HIV-1 phylogenetic clusters from the same clade. Proc Natl Acad Sci U S A. 2019;116(1):239-44.
5. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011;27(21):2957-63.
6. Galaxy C. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. Nucleic Acids Res. 2022;50(W1):W345-51.