Cell Host & Microbe, Volume *26*

Supplemental Information

HIV Rebound Is Predominantly Fueled by Genetically Identical Viral Expansions from Diverse Reservoirs

Marie-Angélique De Scheerder, Bram Vrancken, Simon Dellicour, Timothy Schlub, Eunok Lee, Wei Shao, Sofie Rutsaert, Chris Verhofstede, Tessa Kerre, Thomas Malfait, Dimitri Hemelsoet, Marc Coppens, Annemieke Dhondt, Danny De Looze, Frank Vermassen, Philippe Lemey, Sarah Palmer, and Linos Vandekerckhove

Supplementary data:

Figure S1: Maximum likelihood phylogenetic trees. Related to Figure 3 and 5. The colored strip represents sampling origin for each sequence as indicated by the legend. The trees are drawn to scale and the grey circles represent the branch length from the root expressed as the number of substitutions per site. The scale values are provided in the inset. Panel A shows the trees as presented in figure 3 for the other participants highlighting the sequences from T1 cell subsets from blood (TCM, TEM, TTM and TN), LN (TCM, TEM) and GALT (CD45+ cells). Plasma viruses from time points T2, T3 and T4 are grouped as plasma after STOP cART. Panel B represents the trees as presented in figure 5 for the other participants highlighting the sequences obtained from plasma RNA at the different time points of sequencing T0, T1, T2, T3, T4. Panel C represents the two participants in which we sequenced the envelope region from CD4+ T cells from Bone Marrow. Panel D represents CSF sequences at T1 and T4 together with all plasma sequences and all proviral sequences from T1 combined for 5 participants in which we obtained CSF sequences.

Figure S2: Haplotype networks. Related to Figure 3. One haplotype network has been generated per participant (A). In these graphs, each haplotype (i.e. unique sequence) is represented by a circle, the size of which is proportional to its overall sampling frequency in a given participant. The genetic relatedness between haplotypes is represented by lines. Each line segment in the networks represents a single mutational change, or multiple changes when the line is annotated by a number indicating the number of mutations. Haplotype colors indicate the sampling origin of each sequence by anatomic compartment (blood, LN and GALT) and plasma from all available time points. An additional haplotype (B) was designed for STAR 10 based on cell subset origin to enlighten on the clonal distribution over subsets in blood, the participation of TN to viral rebound and the stochastic kinetics of rebound (T1 proviral subsets, identical to T4 plasma sequences). Relevant recombinant strains (RC) with rebound virus from STAR 7, 10 and 11 were depicted as haplotype (C). The color code for (B) and (C) indicate the sampling origin of each sequence by cell subset and plasma from all available time points. Circles were used to highlight specific details of the haplotype, as described in the manuscript: A) STAR 3: the left circle shows an intermingling between the different compartments and plasma virus T1 with early rebound virus at T2 and T3, whereas the right circle shows a clear outgrowth of plasma virus after rebound (T3-T4). STAR 7: the circle shows a viral expansion in plasma at T1, having identical sequences at T2 and T3. STAR 12: the circle shows a viral expansion at T0 that was also found in proviral sequences at T1 and at later plasma time points. B) STAR 10: we observe a clonal distribution over subsets in blood, the participation of TN to viral rebound and the stochastic kinetics of rebound within the circle (T1 proviral subsets, identical to T4 plasma sequences). C) STAR 7: right circle: as in A) we observe in the recombinant strain a viral expansion in plasma at T1, having identical sequences at T2 and T3.

Figure S3: Nucleotide diversity computed within different groups of sequences. Related to Figure 5. Evidence for kinetic variability and stochastic reactivation of rebound virus after treatment interruption. For each patient, two distinct hierarchical partitions were considered (i.e. anatomical compartments and cell subsets), both indicating various levels of nucleotide diversity (y axis) in the different compartments, cell subsets and plasma (x axis).

Figure S4: Sample size in terms of independent rebound lineages offers strong support against HIV rebound seeded by a dominant reservoir. Related to Figure 3 and 5. Panel A Left*:* The independent rebound lineages amounts to an estimated total of 216 seeding events. Based on the most probable source assignment, we can summarize those lineages as being seeded from the following cellular populations. Right: if we assume that the most frequently observed source for viral rebound (CD45 GALT) in our data would also be the dominant reservoir population at the population level (a conservative assumption in our argument), we can quantify how strongly our data can reject specific expectations of the frequency of dominance (*p*). For this purpose we compute the likelihood curve for this frequency using a binomial function. The strength of our data is reflected in the narrow binomial asymptotic 95% confidence interval. Based on the upper boundary for this confidence interval, we can reject a population frequency > 36% for a dominant rebound reservoir among the cellular populations we investigate. The ability to formally reject moderate to high frequencies of dominance at the population level provides very strong evidence against the hypothesis of a single dominant rebound population. Panel B Left: Number of independent rebound lineages per participant. Right: Radar plot representing, for each participant, the estimated number of times that a rebound virus lineage originates from the respective cell subsets (depicted as the numbers from 0 to 35). Color legend represents the different participants. Panel C: Plot showing the cumulative number of *env* sequences and the proportion of intact *env* sequences in plasma over time (T0-T4).

Figure S5: Gating strategy. Related to the STAR methods.

1) Example of gating strategy used in flow cytometric sorting experiments using a BD Facs Yazz with red and blue laser**. a)** Lymphocyte populations were first selected using forward and side scatter characteristics. Following this, the exclusion of doublets was performed. Live and then $CD3^+$ lymphocytes were selected, and further gated for CD4 expression by gating CD8 negative cells. TN(CD45RA+) cells were sorted after selecting CD3+CD8- lymphocytes**. b)** Similarly, CD45RO+ cells were selected and expression of CD27 was used to sort TCM (CD27+) and TEM (CD27-) cells from blood and lymph node. **c)** Same sorting strategy was used to sort TTM introducing CCR7 as an additional marker and selecting CD45RO+CD27+CCR7- cells. **d)** CD45 immune cells were sorted from GUT samples, after excluding doublets and dead cells. Purity checks were performed for all the above conditions.

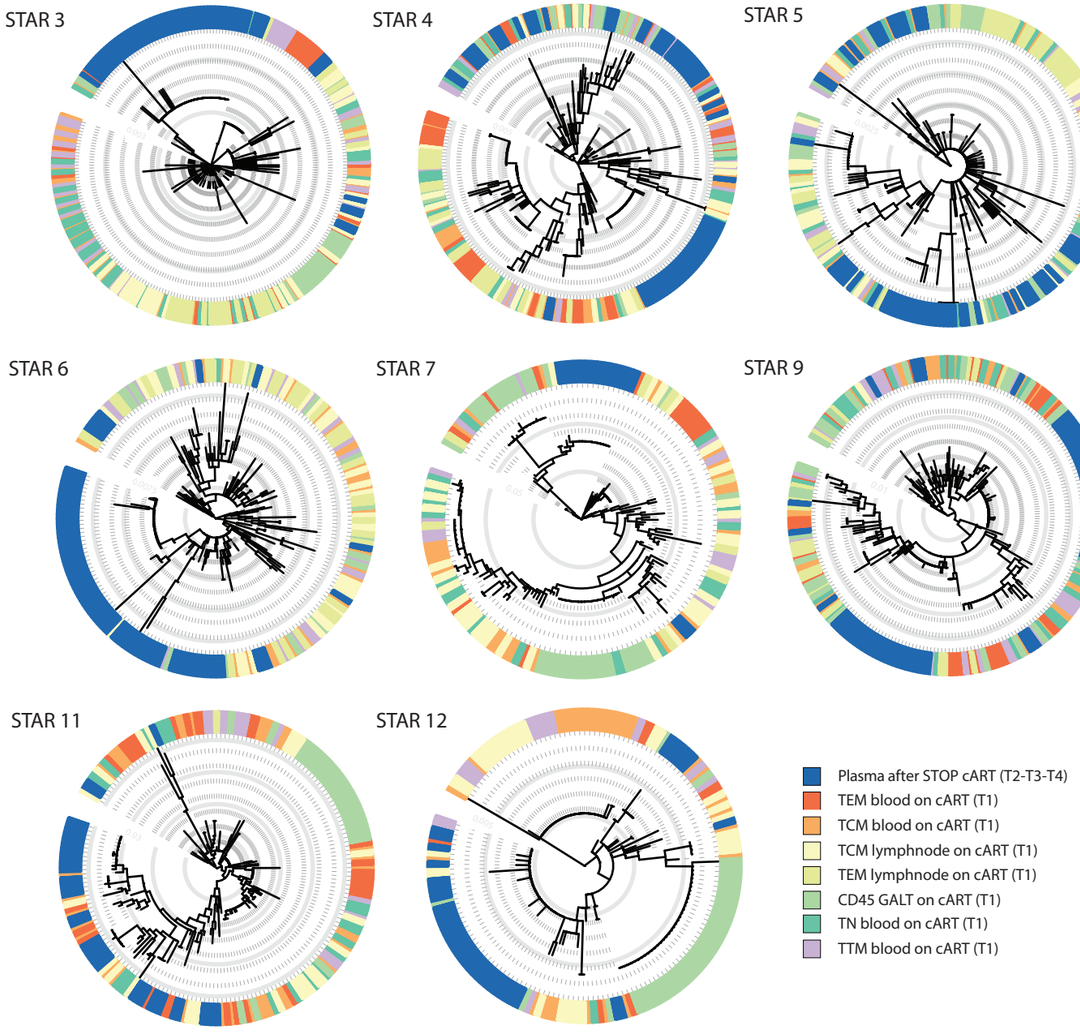
2) Flowcytometry analysis of the T naïve CD4+ cell population where we show that few cells do not express CD27 (<10%), and from the CD27+ population only a fraction were CD95 positive (<1%), therefore concluding that our target cells were indeed mostly TN cells.

Table S1: Relevant demographics of the HIV-infected study participants. Related to the STAR methods.

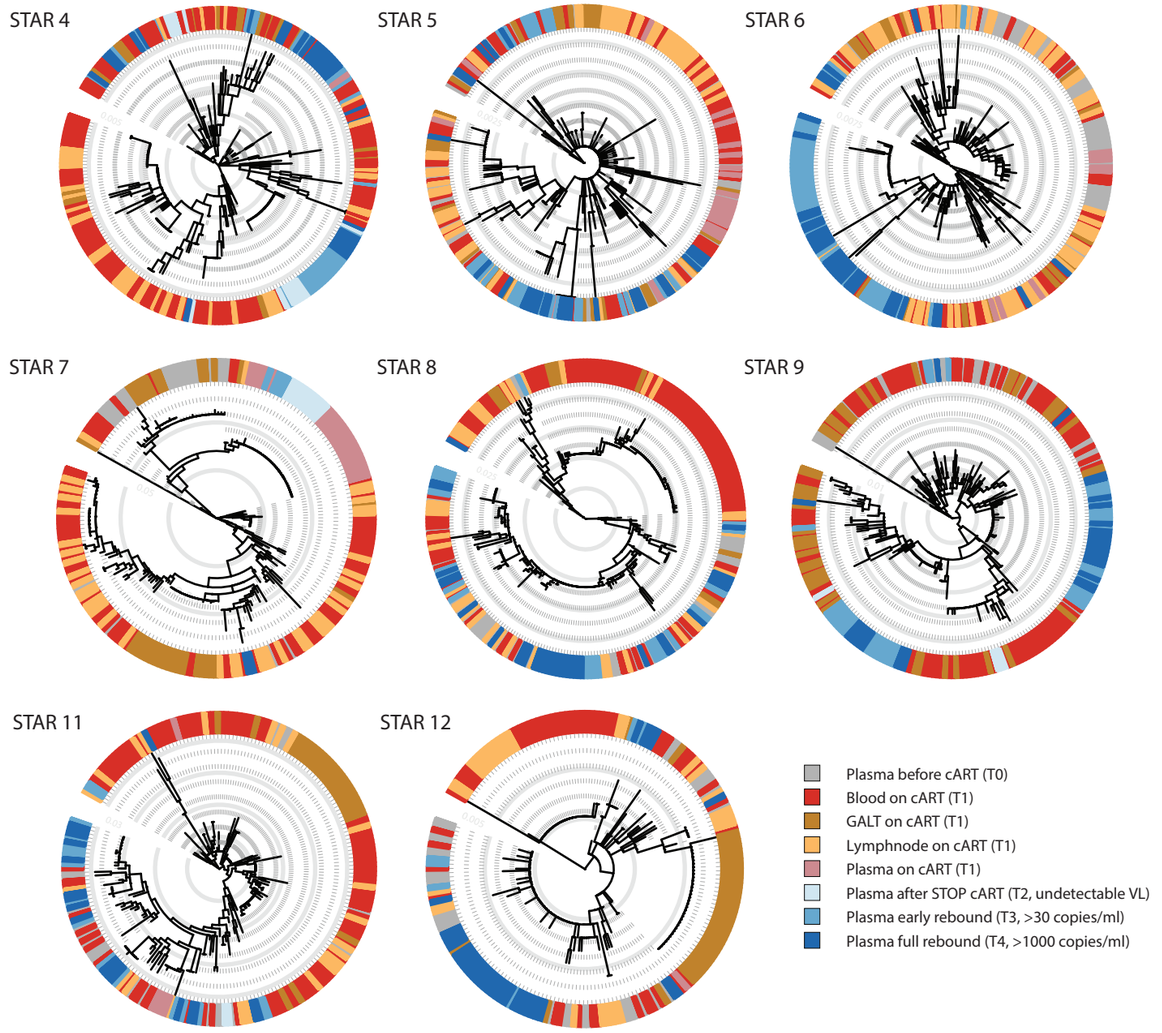
Table S2: Phylogenetic association estimates. Related to Figure 3. The table summarizes the mean ratio of the observed over randomized phylogenetic association values as well as their Bayesian credible interval (BCI). The left side of the table list the values for anatomical compartments as trait, while the right side lists the values for sampling time points as traits. The estimates are provided for the full data sets (T1 complete) as well as for the data set for which identical sequences by trait were reduced to a single representative sequence (T1 aggregated). Values close to 0 and 1 for this ratio represent absolute clustering by trait and randomized clustering respectively. A BCI excluding 0 can be considered as significant evidence against absolute clustering by trait whereas a BCI excluding 1 can be considered as significant evidence against random clustering. These extreme scenarios can be rejected for all participants except for the randomized clustering by anatomic compartment in one STAR6 cladeA.

Table S3: Overview of the sequence data subsets, their size and use for the phylogenetic analyses. Related to the STAR methods.

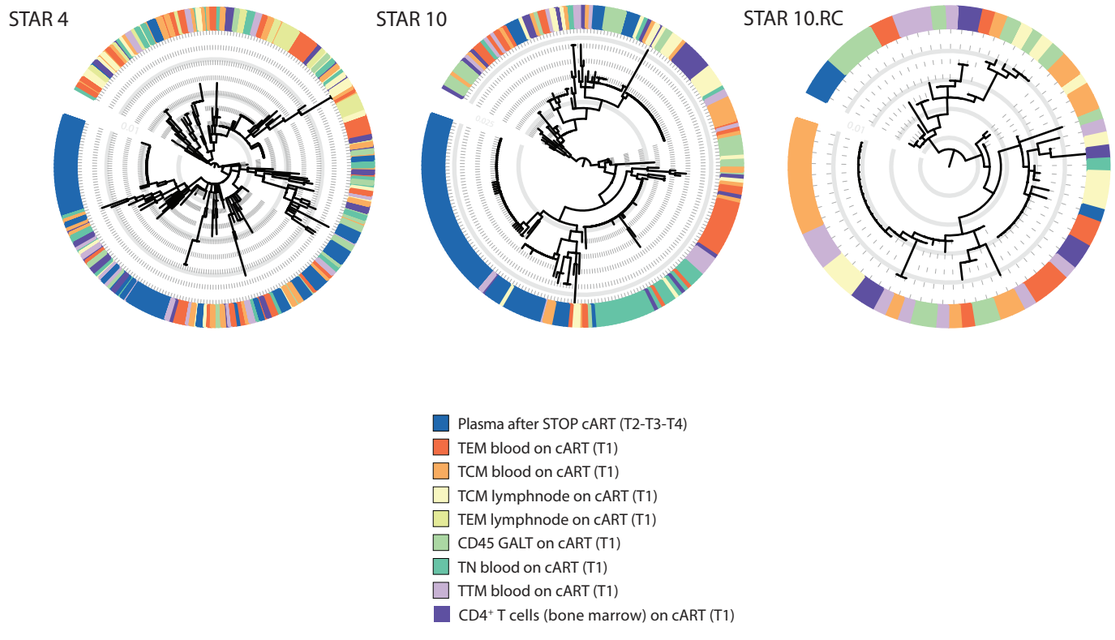
Panel A.



Panel B.



Panel C.



Panel D.

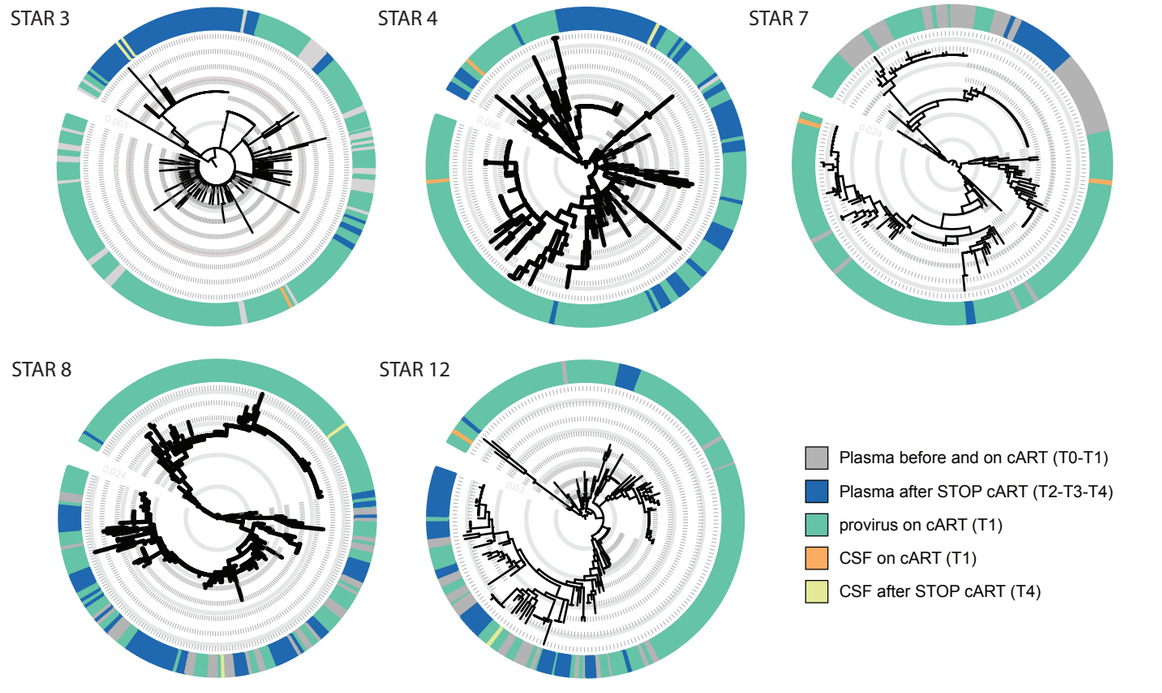


Figure S1: Maximum likelihood phylogenetic trees

Figure S2: Haplotype networks

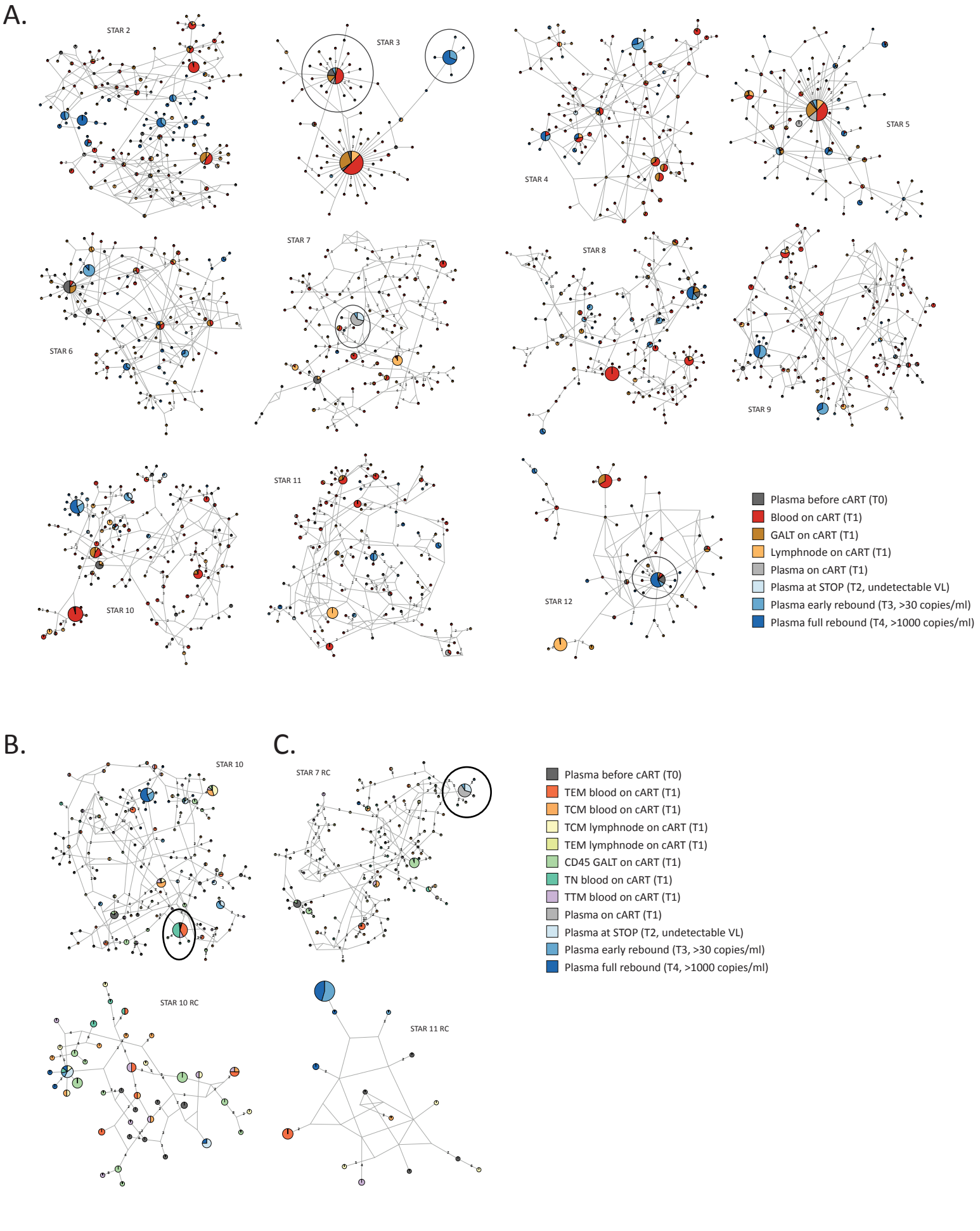


Figure S3: Nucleotide diversity

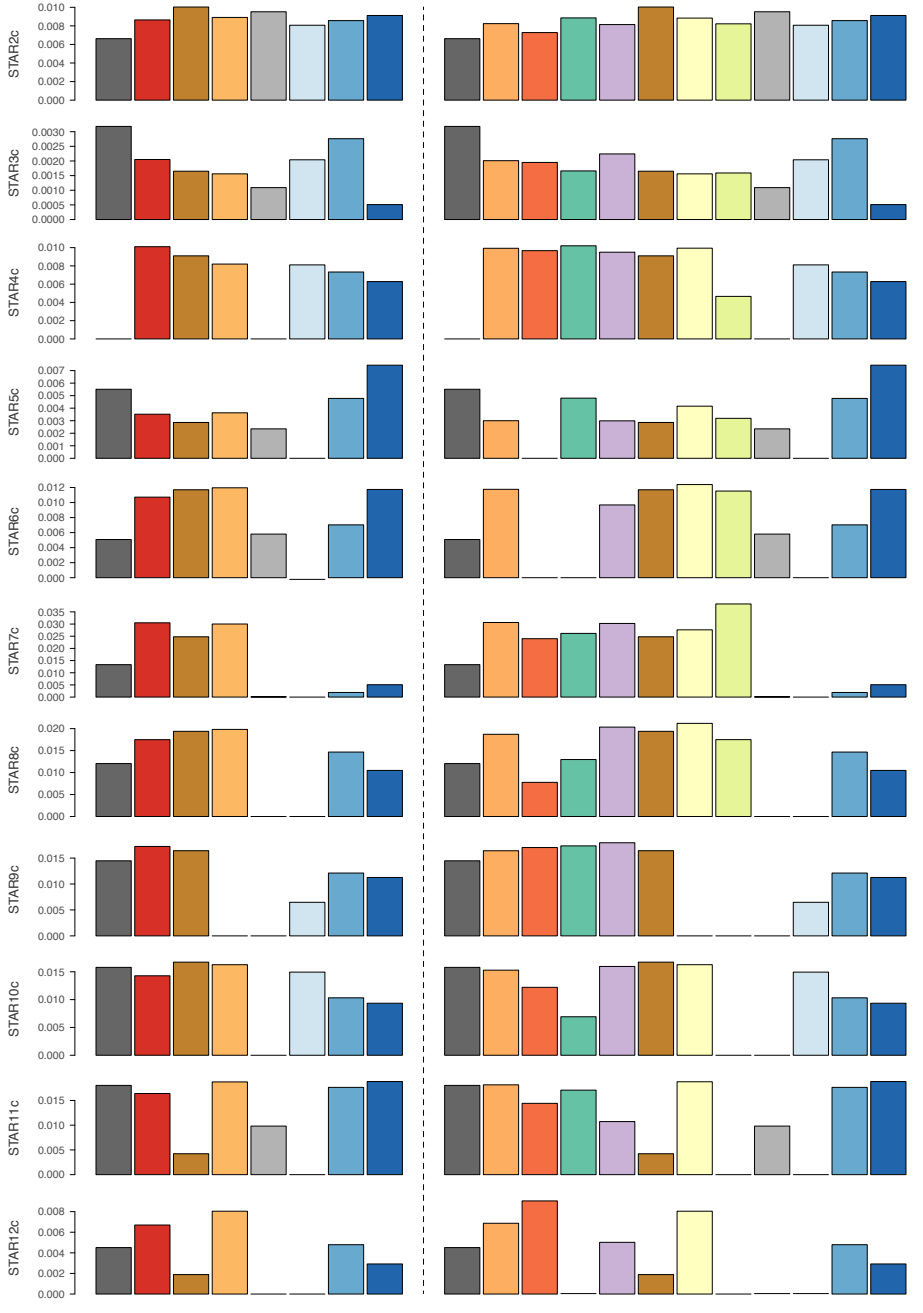
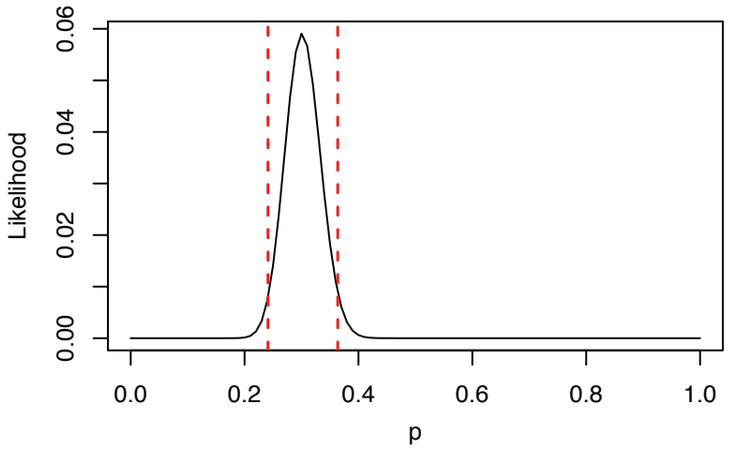


Figure S4

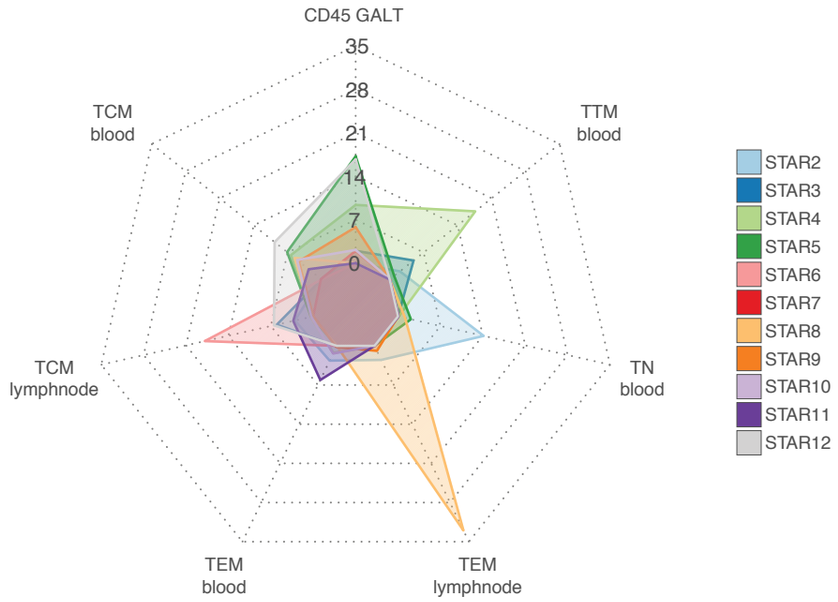
Panel A.

cell population	# rebound lineages
CD45 GALT	65
TCM blood	32
TCM lymph node	33
TEM blood	14
TEM lymph node	33
TN blood	16
TTM blood	24



Panel B.

participant	# rebound lineages
STAR2	24
STAR3	13
STAR4	35
STAR5	27
STAR6	17
STAR7	2
STAR8	39
STAR9	11
STAR10	8
STAR11	10
STAR12	31



Panel C.

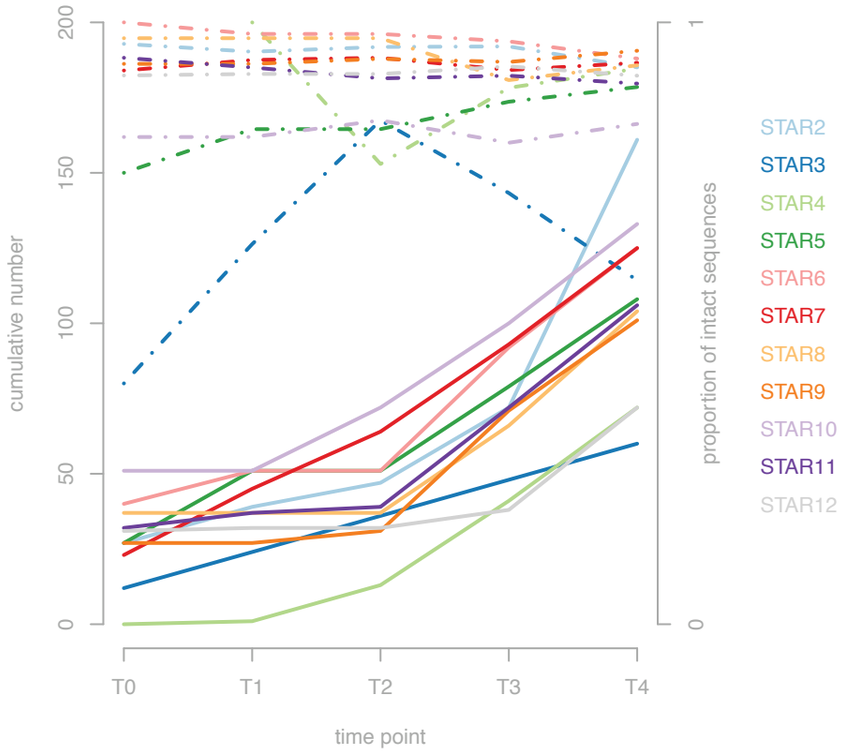
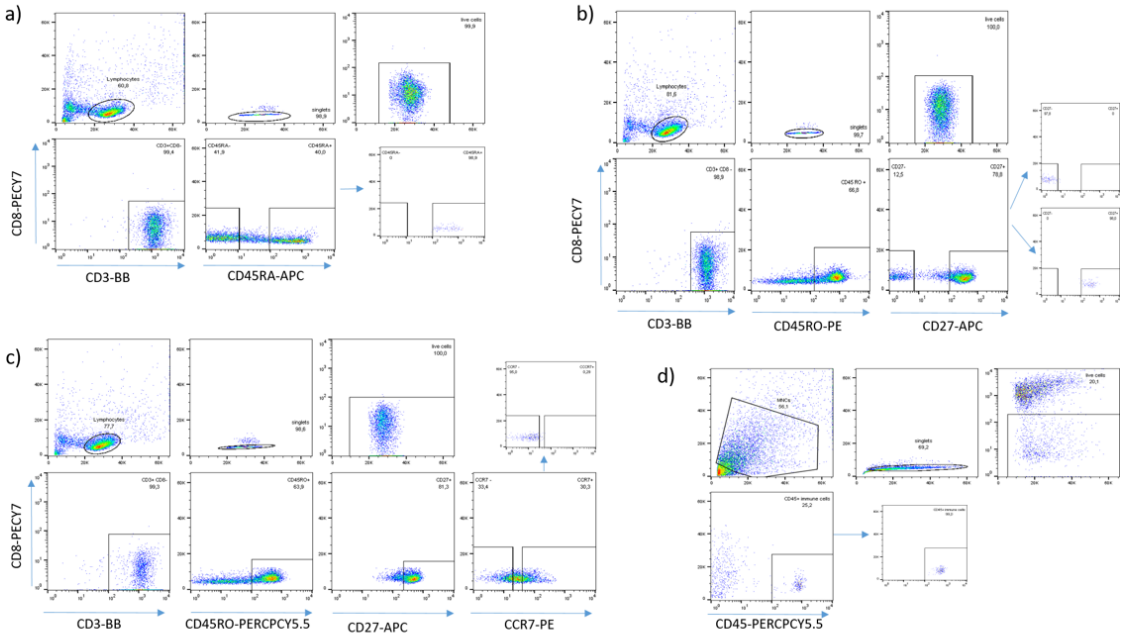


Figure S5: Gating strategy

1.



2.

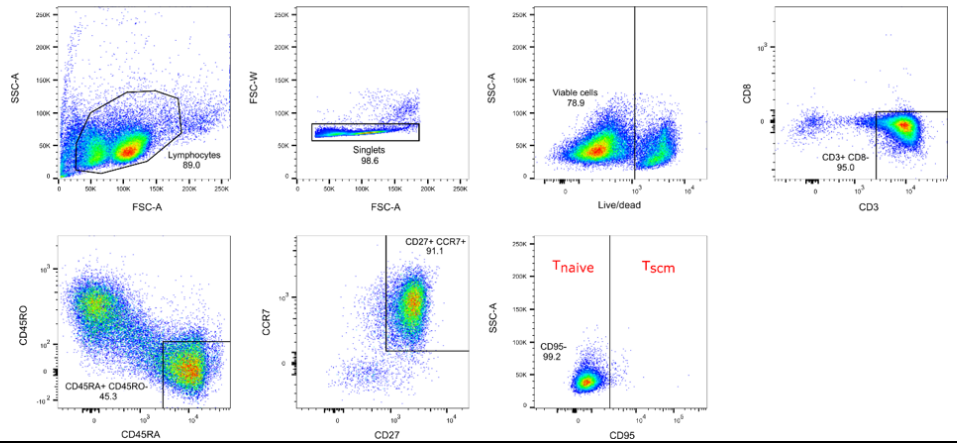


Table S1: Relevant demographics of the HIV-infected study participants

Subject-ID	Gender	Age (y)	CD4 nadir (cells/µl)	VL zenith (log10c/ml)	Time since primo-infection (y)	Time on cART (y)	Time to viral rebound (d)	VL at T0 (c/ml)	VL at T1 (c/ml)	VL at T2 (c/ml)	VL at T3 (c/ml)	VL at T4 (c/ml)	HIV subtype	Viral tropism
STAR 2	M	38	438	5.08	NA	4	36	68219	<20	<20	462	2050	B	R5
STAR 3	M	52	331	5.53	3	3	28	431421	<20	27	38	5730	B	R5
STAR 4	M	36	142	4.31	NA	11	20	NA	<20	<20	790	36500	B	R5
STAR 5	M	39	308	4.77	3	3	15	58482	<20	NA	83	1880	B	X4
STAR 6	M	41	378	4.82	4	3	19	51482	<20	77	103	1610	B	R5
STAR 7	M	40	421	4.8	11	3	21	63499	<20	<20	124	4310	B	R5
STAR 8	M	40	400	4.64	11	8	21	5230	<20	<20	107	1098	B	R5
STAR 9	M	32	405	5.01	9	6	21	30115	<20	<20	101	1967	B	R5
STAR 10	M	54	327	5.49	29	11	21	82110	<20	<20	683	4450	B	R5
STAR 11	M	37	432	3.62	9	2	28	4160	<20	<20	202	1790	B	R5
STAR 12	F	56	348	3.23	7	6	25	1741	<20	<20	74	364	B	R5

Table S2: Phylogenetic association estimates

	anatomic compartment			sampling time point		
complete dataset	mean	95% BCI lower	95% BCI upper	mean	95% BCI lower	95% BCI upper
STAR2	0.57	0.53	0.61	0.75	0.70	0.79
STAR3	0.78	0.70	0.86	0.76	0.67	0.83
STAR4	0.67	0.61	0.72	0.81	0.76	0.86
STAR5	0.84	0.79	0.90	0.87	0.81	0.94
STAR6.cladeA	0.92	0.65	1.12	0.68	0.43	0.86
STAR6.cladeB	0.56	0.52	0.60	0.59	0.55	0.63
STAR7	0.45	0.41	0.49	0.76	0.71	0.81
STAR8	0.56	0.50	0.61	0.81	0.76	0.86
STAR9	0.41	0.36	0.46	0.77	0.73	0.81
STAR10	0.40	0.34	0.45	0.72	0.65	0.78
STAR11	0.47	0.43	0.52	0.77	0.72	0.81
STAR12	0.49	0.41	0.56	0.68	0.60	0.76
T1 dedupl dataset	mean	95% BCI lower	95% BCI upper	mean	95% BCI lower	95% BCI upper
STAR2	0.63	0.59	0.66	0.76	0.72	0.80
STAR3	0.80	0.72	0.89	0.79	0.71	0.87
STAR4	0.66	0.60	0.70	0.78	0.73	0.82
STAR5	0.86	0.79	0.92	0.87	0.80	0.93
STAR6.cladeA	0.96	0.71	1.18	0.68	0.47	0.89
STAR6.cladeB	0.61	0.57	0.65	0.61	0.57	0.65
STAR7	0.53	0.49	0.56	0.79	0.74	0.83
STAR8	0.61	0.55	0.67	0.83	0.78	0.88
STAR9	0.45	0.40	0.49	0.80	0.75	0.83
STAR10	0.41	0.34	0.46	0.73	0.66	0.80
STAR11	0.58	0.54	0.62	0.78	0.73	0.83
STAR12	0.67	0.59	0.76	0.76	0.67	0.84

Table S3: Overview of the sequence data subsets, their size and use for the phylogenetic analyses.

						maximum likelihood tree	Bayesian tree estimation		Inference of trait evolution	
Patient	**dataset**	**subset**	**#taxa w/ recombinants**	**#taxa w/o recombinants**	**#recombinants removed**	**PhyML v3.0**	**MrBayes v3.2.6**	**BEAST v1.8.4**	**BaTS**	**Discrete trait analysis on single, fixed maximum liklihood tree and on a collection of plausible trees**
STAR2	conservative	all	535	533	2	no	yes	no	yes	no
	conservative	all, deduplicated by anatomic compartment at T1	n.a.	488	n.a.	no	yes	no	yes	no
	conservative	TORNA, T1 DNA, T2-T3-T4 RNA	n.a.	358	n.a.	yes	no	yes	no	yes
STAR3	conservative	all	411	411	0	no	yes	no	yes	no
	conservative	all, deduplicated by anatomic compartment at T1	n.a.	291	n.a.	no	yes	no	yes	no
	conservative	TORNA, T1 DNA, T2-T3-T4 RNA	n.a.	302	n.a.	yes	no	yes	no	yes
STAR4	conservative	all	366	366	0	no	yes	no	yes	no
	conservative	all, deduplicated by anatomic compartment at T1	n.a.	327	n.a.	no	yes	no	yes	no
	conservative	TORNA, T1 DNA, T2-T3-T4 RNA	n.a.	250	n.a.	yes	no	yes	no	yes
STAR5	conservative	all	362	362	0	no	yes	no	yes	no
	conservative	all, deduplicated by anatomic compartment at T1	n.a.	288	n.a.	no	yes	no	yes	no
	conservative	TORNA, T1 DNA, T2-T3-T4 RNA	n.a.	274	n.a.	yes	no	yes	no	yes
STAR6 cladeA	conservative	all	43	43	0	no	yes	no	yes	no
	conservative	all, deduplicated by anatomic compartment at T1	n.a.	39	n.a.	no	yes	no	yes	no
	conservative	TORNA, T1 DNA, T2-T3-T4 RNA	n.a.	21	n.a.	yes	no	no	no	no
STAR6 clade B	conservative	all	336	322	14	no	yes	no	yes	no
	conservative	all, deduplicated by anatomic compartment at T1	n.a.	312	n.a.	no	yes	no	yes	no
	conservative	TORNA, T1 DNA, T2-T3-T4 RNA	n.a.	249	n.a.	yes	no	yes	no	yes
STAR7	conservative	all	382	275	107	no	yes	no	yes	no
	conservative	all, deduplicated by anatomic compartment at T1	n.a.	238	n.a.	no	yes	no	yes	no
	conservative	TORNA, T1 DNA, T2-T3-T4 RNA	n.a.	201	n.a.	yes	no	yes	no	yes
	conervative_91RC	all	n.a.	91	n.a.	no	no	no	no	no
	conervative_91RC	TORNA, T1 DNA, T2-T3-T4 RNA	n.a.	69	n.a.	yes	no	yes	no	yes
STAR8	conservative	all	411	385	26	no	yes	no	yes	no
	conservative	all, deduplicated by anatomic compartment at T1	n.a.	322	n.a.	no	yes	no	yes	no
	conservative	TORNA, T1 DNA, T2-T3-T4 RNA	n.a.	274	n.a.	yes	no	yes	no	yes
	conservative_5RC	all	n.a.	5	n.a.	no	no	no	no	no
	conservative_21RC	all	n.a.	21	n.a.	no	no	no	no	no
STAR9	conservative	all	360	360	0	no	yes	no	yes	no
	conservative	all, deduplicated by anatomic compartment at T1	n.a.	333	n.a.	no	yes	no	yes	no
	conservative	TORNA, T1 DNA, T2-T3-T4 RNA	n.a.	251	n.a.	yes	no	yes	no	yes
STAR10	conservative	all	466	360	106	no	yes	no	yes	no
	conservative	all, deduplicated by anatomic compartment at T1	n.a.	280	n.a.	no	yes	no	yes	no
	conservative	TORNA, T1 DNA, T2-T3-T4 RNA	n.a.	264	n.a.	yes	no	yes	no	yes
	conservative_106RC	all	n.a.	106	n.a.	no	no	no	no	no
	conservative_106RC	TORNA, T1 DNA, T2-T3-T4 RNA	n.a.	75	n.a.	yes	no	yes	no	yes
STAR11	conservative	all	417	345	72	no	yes	no	yes	no
	conservative	all, deduplicated by anatomic compartment at T1	n.a.	295	n.a.	no	yes	no	yes	no
	conservative	TORNA, T1 DNA, T2-T3-T4 RNA	n.a.	212	n.a.	yes	no	yes	no	yes
	conservative_19RC	all	n.a.	19	n.a.	no	no	no	no	no
	conservative_53RC	all	n.a.	53	n.a.	no	no	no	no	no
	conservative_53RC	all, deduplicated by anatomic compartment at T1	n.a.	42	n.a.	yes	no	yes	no	yes
STAR12	conservative	all	240	240	0	no	yes	no	yes	no
	conservative	all, deduplicated by anatomic compartment at T1	n.a.	164	n.a.	no	yes	no	yes	no
	conservative	TORNA, T1 DNA, T2-T3-T4 RNA	n.a.	195	n.a.	yes	no	yes	no	yes