www.sciencetranslationalmedicine.org/cgi/content/full/2/18/18re1/DC1

Supplementary Materials for

The Origins of Sexually Transmitted HIV Among Men Who Have Sex with Men

David M. Butler, Wayne Delport, Sergei L. Kosakovsky Pond, Malcolm K. Lakdawala, Pok Man Cheng, Susan J. Little, Douglas D. Richman, Davey M. Smith*

*To whom correspondence should be addressed. E-mail: davey@ucsd.edu

Published 10 February 2010, *Sci. Transl. Med.* **2**, 18re1 (2010)
DOI: 10.1126/scitranslmed.3000447

This PDF file includes:

Supporting Online Material

Table S1: Genetic differentiation between compartments

Transmission Pair		Intra-individual			Inter-individual		
		SBP-SSP	SBP-SSC	SSP-SSC	SBP-RBP	SSP-RBP	SSC-RBP
A	F_{st}	0.842	0.727	0.707	0.827	0.549	0.692
	sd	0.019	0.030	0.042	0.020	0.022	0.046
	P	***<0.001***	***<0.001***	***<0.001***	***<0.001***	***<0.001***	***<0.001***
B	F_{st}	0.034	0.995	0.993	-0.010	0.045	0.996
	sd	0.046	0.002	0.002	0.037	0.046	0.002
	P	0.203	***<0.001***	***<0.001***	0.501	0.112	***<0.001***
C	F_{st}	0.010	0.992	0.991	-0.012	-0.008	0.990
	sd	0.086	0.004	0.003	0.060	0.117	0.004
	P	0.309	***<0.001***	***<0.001***	0.518	0.526	***<0.001***
D	F_{st}	0.004	0.579	0.563	0.156	0.132	0.460
	sd	0.027	0.057	0.060	0.072	0.077	0.078
	P	0.324	***<0.001***	***<0.001***	0.07	0.084	***<0.001***
E	F_{st}	-0.081	0.698	0.702	0.216	0.384	0.693
	sd	0.104	0.076	0.077	0.129	0.035	0.073
	P	0.441	***<0.001***	***<0.001***	***<0.001***	***<0.001***	***<0.001***
F	F_{st}	0.449	0.821	0.784	0.376	0.088	0.760
	sd	0.121	0.035	0.032	0.123	0.047	0.035
	P	***<0.001***	***<0.001***	***<0.001***	***<0.001***	**0.021**	***<0.001***

Legend: F_{st}: Pair-wise measures of genetic distance: SBP: Source Blood Plasma; SSP: Source Seminal Plasma; SSC: Source Seminal Cells; RBP: Recipient Blood Plasma, sd: bootstrapped estimates of the standard deviation of the mean, *P*: probability that a random assortment of sequences between compartments/individuals has equal or greater F_{st}, and values <0.05 are bolded. Significant *P* values with Bonferroni correction for multiple comparisons (P < 0.008) are in italics.

Table S2: Estimated divergence times of compartments

Pair	Compartments Investigated	T_{MRCA}	Standard error	ESS	P
A	RBP-SBP	1006.48 (161.78, 1463.68)[#]	63.54 (0.34, 1.05)[#]	358.70	
	RBP-SSP	242.14	13.79	562.86	
	RBP-SSC	1437.88	30.42	345.04	0.999
B	RBP-SBP	1155.96	7.05	397.25	
	RBP-SSP	1168.37	7.05	416.37	
	RBP-SSC	1453.66	29.8	262.91	0.972
C	RBP-SBP	1156.79	7.12	392.75	
	RBP-SSP	1168.86	7.09	412.24	
	RBP-SSC	1453.77	29.8	262.78	0.974
D	RBP-SBP	58.86	4.47	280.93	
	RBP-SSP	58.85	4.48	280.88	
	RBP-SSC	61.01	5.18	254.11	0.999
E	RBP-SBP	465.35	16.41	372.81	
	RBP-SSP	127.82	2.046	657.79	
	RBP-SSC	502.96	18.06	360.53	0.999

Legend: T_{MRCA}: Time to Most Recent Common Ancestor was estimated (S1) in days between source and recipient viral populations in the 5 transmission pairs for which time of infection could reliably be estimated; SBP: Source Blood Plasma HIV RNA; SSP: Source Seminal Plasma HIV RNA; SSC: Source Seminal Cell-associated HIV DNA; RBP: Recipient Blood Plasma HIV RNA; ESS: Effective Sample Size; P is the posterior probability that T_{MRCA} (RBP-SSP) <= T_{MRCA} (RBP-SSC). [#]Multiple runs failed to converge on the estimates of time to the most recent common ancestor. The means and standard errors of both modes are thus presented.

Table S3: Number of purifying (dN/dS < 1) and positive (dN/dS > 1) selection sites

Transmission Pair	FEL		iFEL		FEL & iFEL	
	dN/dS < 1	dN/dS > 1	dN/dS < 1	dN/dS > 1	dN/dS < 1	dN/dS > 1
A	6	2	8	2	5	2
B	3	0	3	0	2	0
C	2	0	1	0	1	0
D	10	0	4	0	4	0
E	8	0	7	0	3	0
F	14	4	11	6	11	4

Legend: FEL: Fixed Effects Likelihood inference of selection at individual sites across all lineages; iFEL: Fixed Effects Likelihood inference of selection at individual sites along internal branches; dN: non-synonymous rate, dS: synonymous rate. FEL results indicate sites experiencing significant ($P < 0.05$) positive/purifying selection along all lineages, whereas iFEL results are only at internal branches.

Table S4: Inferred substitutions along transmission branches for pairs A, E and F (panel A) and along pre-transmission branches for all pairs A-K (panel B).

Panel A		**Transmission Pairs (codon substitutions with aa)**			
hxb2 position	aa	A	E	F(1)*	F(2)*
285	I		CTA (L) > TTA(L)		
299	P	CTC(L) > CTT(L)			
301	N	AAC(N) > AAT(N)			
324	--				
329	Q	CAA(Q) > AAA(K)			CAA(Q) > CGA(R)
342	N			AAC(N) > GAC(D)	
345	K			AAA(K) > ACA(T)	
347	I				ATA(I) > GTA(V)

Panel B		**Transmission Pairs (codon substitutions with aa)**					
hxb2 position	aa	A	B	C	D	E	F
252	R						AGG(R) > AAG(K)
259	L				TTG(L) > CTG(L)		
260	L						CTG(L) > CTA(L)
261	L						TTA(L) > TTG(L)
269	E	GAG(E) > GGG(G)				GGG(G) > GAG(E)	
270	V						GTA(V) > GTC(V)
271	V						ATA(I) > ATG(M)
272	I					GTT(V) > ATT(I)	
275	V		GCC(A) > GTC(V)				GAG(E) > GAA(E)
276	N					AAC(N) > AAT(N)	
277	F						TTC(F) > CTC(L)
278	T	ACG(T) > TCG(S)	TCG(S) > ACG(T)	TCG(S) > ACG(T)		ACA(T) > ACG(T)	

279	D	GAC(D) > AAC(N)	AAC(N) > GAC(D)	AAC(N) > GAC(D)			
283	T				ACA(T) > ACC(T)		ACC(T) > AAC(N)
285	I					CTA(L) > ATA(I)	
287	Q						CAG(Q) > CAT(H)
288	L				CTA(L) > CTG(L)		
290	T					AAC(N) > AAA(K)	GAG(E) > CAG(Q)
293	E	GAA(E) > GTA(V)	GTA(V) > GAA(E)	GTA(V) > GAA(E)	CCC(P) > AAA(K)		AAA(K) > ACA(T)
295	N				CAT(H) > AAT(N)		
296	C				TGC(C) > TGT(C)		
297	T				ATA(I) > ACA(T)		
299	P	CCC(P) > CTC(L)	CTC(L) > CCC(P)	CTC(L) > CCC(P)			CCC(P) > CCT(P)
300	N					GGC(G) > AAC(N)	AAC(N) > AAT(N)
305	K					AAG(K) > AAA(K)	
306	R	AGT(S) > GGT(G)	GGT(G) > AGT(S)	GGT(G) > AGT(S)			
308	R		CAT(H) > CGT(R)		CCT(P) > CAT(H)		
318	V						TTT(F) > TAT(Y)
322	K	GAA(E) > GCC(A)			GAC(D) > GAA(E)	GAT(D) > GAA(E)	
323	I			GTA(V) > ATA(I)			
324	-		GTA(V) > ATA(I)				
329	Q		AAA(K) > CAA(Q)	AAA(K) > CAA(Q)			
331	H						CAT(H) > CAC(H)
332	C						TGT(C) > TGC(C)
334	I	ATT(I) > CTT(L)	CTT(L) > ATT(I)	CTT(L) > ATT(I)			
335	S				ACT(T) > AGT(S)		
336	R					AGT(S) > AGA(R)	AGA(R) > GGG(G)
337	A		ACA(T) > GCA(A)			ACA(T) > GCA(A)	GCA(A) > GAA(E)
338	K		GAA(E) > AAA(K)		AAT(N) > AAA(K)		
340	N				ACT(T) > AAT(N)		
341	-				ACT(T) > AAC(N)		
342	N				AGC(S) > ACT(T)		
345	K	AAA(K) > GAA(E)	GAA(E) > AAA(K)		CAA(Q) > AAA(K)	GAA(E) > AAA(K)	
346	Q				CTG(L) > CAG(Q)		CAG(Q) > AAG(K)
348	A					GTT(V) > GCT(A)	
349	S		AAA(K) > AGA(R)		GCA(A) > AGA(R)		AGA(R) > ACA(T)

351	L					TTG(L) > TTA(L)	
352	R				GGA(G) > AGA(R)		AAA(K) > CAA(Q)
353	E						GAA(E) > AAA(K)
354	Q						CAA(Q) > GAA(E)
355	F	TTT(F) > TAC(Y)	TAC(Y) > TTT(F)		TAT(Y) > TTT(F)		TTT(F) > TTC(F)
358	N				AAC(N) > AAT(N)		
359	K				GCA(A) > AAA(K)		
362	I	ATC(I) > GTC(V)					ATC(I) > GTC(V)
364	K					AAT(N) > AAG(K)	

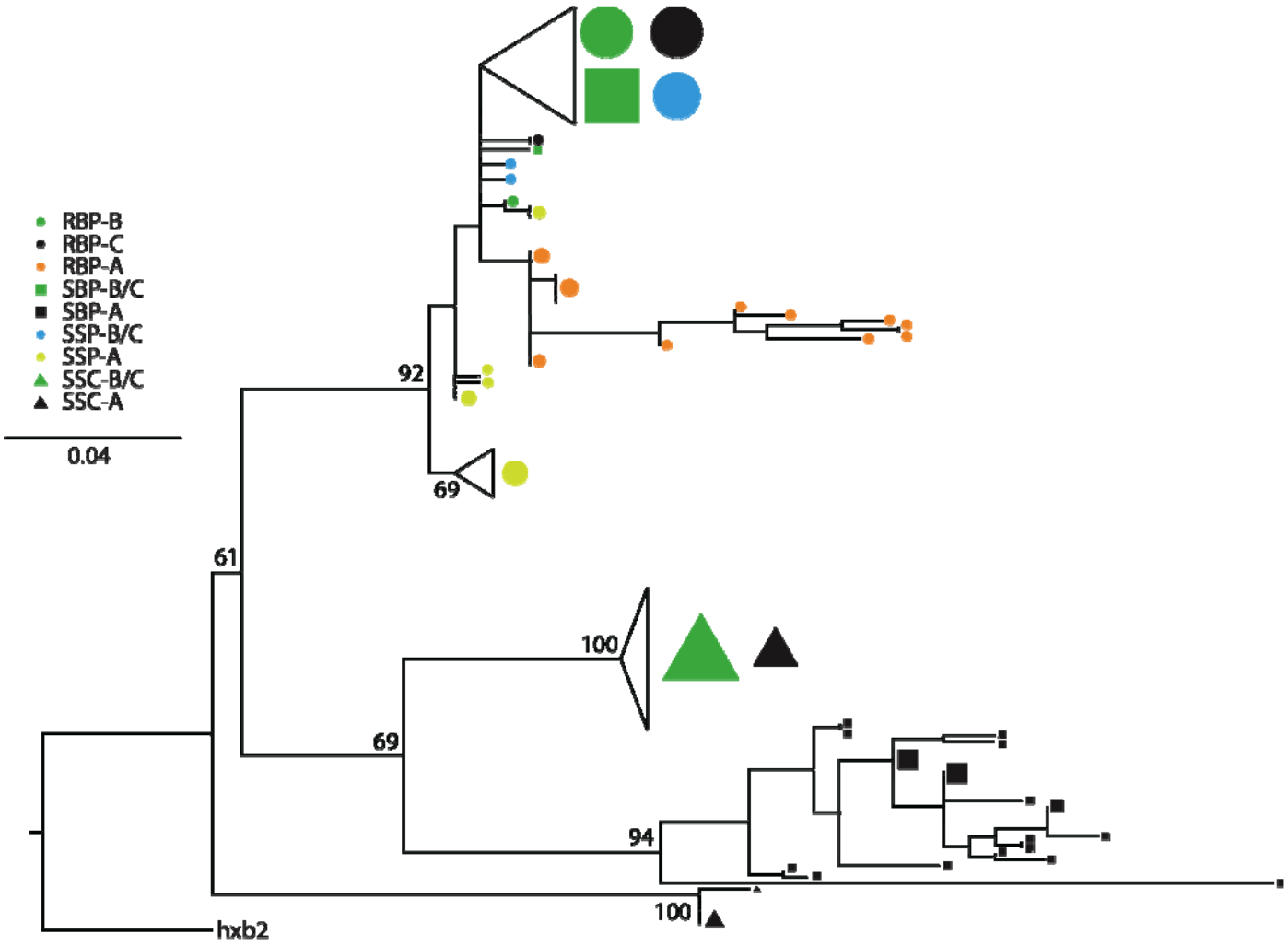
Legend: aa: one letter amino acid code; -: indel; >: substitution direction. *Transmission pair F had two potential transmission branches leading to each of two clades containing both source and recipient virus (see Figure 1).

Table S5: Marginal likelihoods, estimated as the harmonic mean of the sampled likelihoods in Bayesian MCMC analysis (S1).

Substitution model	Demographic model	Marginal likelihood
GTR+Γ	Constant coalescent	-1145.29 ± 0.16
GTR+Γ	Exponential coalescent	-1148.15 ± 0.18
GTR+Γ	Bayesian Skyline	-1145.27 ± 0.18
SRD06	Constant coalescent	-1146.27 ± 0.16
SRD06	Exponential coalescent	-1146.64 ± 0.16
SRD06	Bayesian Skyline	-1144.46 ± 0.18

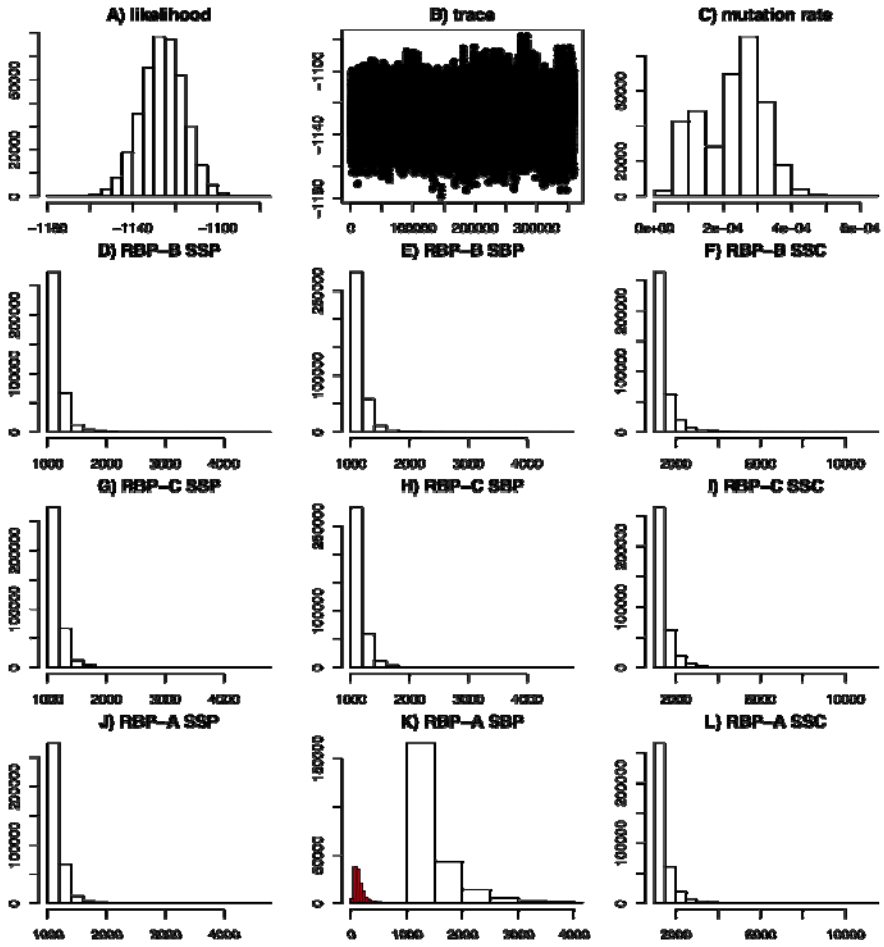
Figure Legends

Figure S1: Maximum likelihood phylogenetic tree of viral sequences from a source who infected multiple partners at two independent time points.



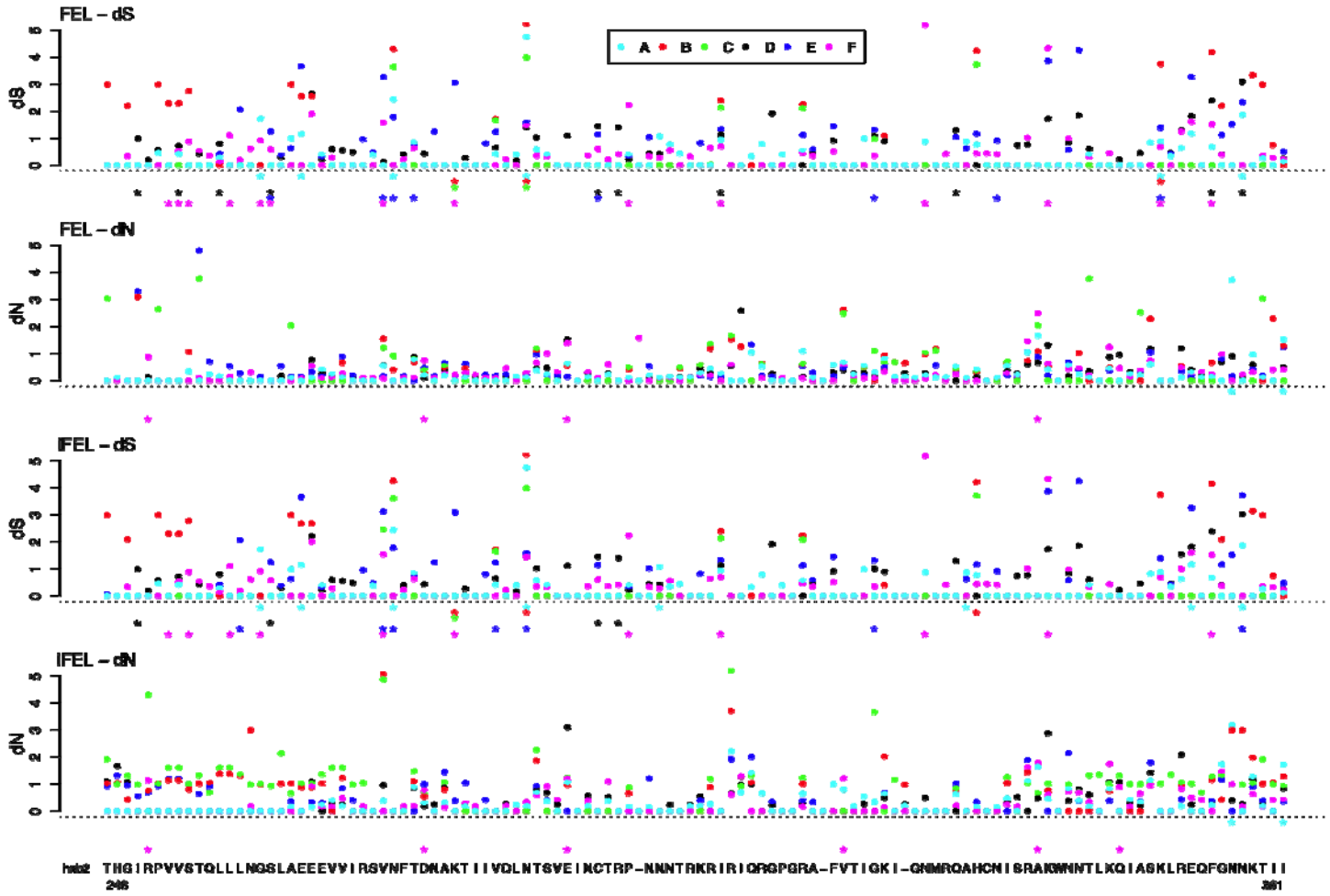
Legend: RBP: recipient HIV RNA in blood plasma, SBP: source HIV RNA in blood plasma, SSP: source HIV RNA in seminal plasma, SSC: source seminal cell-associated virus. B/C: first time point corresponding to transmission pairs B & C; A: second time point corresponding to transmission pair A.

Figure S2: Posterior distributions of parameters sampled during Bayesian MCMC analyses (S1).



Legend: As in Fig S1

Figure S3: Site-specific non-synonymous (dN) and synonymous (dS) rates of substitution.



Legend: FEL: Fixed Effects Likelihood; and iFEL: Internal Fixed Effects Likelihood. Estimated rates for each of the transmission pairs (A-F) are shown, with asterisks (*) below plots indicating sites with significant ($P < 0.05$) purifying (*dS* plots) or diversifying (*dN* plots) selection.

References

References

S1. A. J. Drummond, A. Rambaut, BEAST: Bayesian evolutionary analysis by sampling trees *BMC Evol Biol* **7**, 214 (2007).