References

- [1] HIV Sequence Database (2015) Los Alamos National Laboratory. http://www.hiv.lanl.gov/. Accessed June 24, 2015.
- [2] Simmonds P, et al. (1991) Discontinuous sequence change of human immunodeficiency virus (HIV) type 1 env sequences in plasma viral and lymphocyte-associated proviral populations in vivo: implications for models of HIV pathogenesis. *Journal of Virology* 65(11):6266–6276.
- [3] Shankarappa R, et al. (1999) Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J. Virol.* 73(12):10489–10502.
- [4] Edwards CTT, et al. (2006) Population genetic estimation of the loss of genetic diversity during horizontal transmission of HIV-1. *BMC Evolutionary Biology* 6(28).
- [5] Fischer M, et al. (2004) Attenuated and nonproductive viral transcription in the lymphatic tissue of HIV-1-infected patients receiving potent antiretroviral therapy. *J. Infect. Dis.* 189(2):273–285.
- [6] Novitsky V, et al. (2009) Timing constraints of in vivo gag mutations during primary HIV-1 subtype C infection. *PLoS ONE* 4(11):e7727.
- [7] Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6):716–723.

Table S1: Summary of the patient data collected from the LANL HIV sequence database [1] in the data sets from public sources.

1 1 1 2 2 2 2 2 2 2	Reference	П	S	Sequences		Ti	Time points		Time span		Linear model	Σ	MAE	M	MAD	p-Latency
2658 69 6 5 5 5 5 64 -1.2 (-2.0, -0.46) 0.42 0.081 825 49 6 82 6 8.2 60 0.11 (-0.53, 0.74) 0.73 0.089 7259 28 4 4 2.4 9.8 -0.97 (-2.2, 0.30) 0.69 0.28 7265 20 3 4 2.4 2.4 9.8 -0.97 (-2.2, 0.30) 0.69 0.89 1333 38 4 4 4 1.4 1.7 -0.49 (-1.1, 0.13) 0.78 0.20 1334 36 4 4 4 2.0 1.7 -0.49 (-1.1, 0.13) 0.31 0.20 1334 42 4 4 2.0 2 0.73 (-1.0.04) 0.8 0.73 821 42 4 4 2.0 2 0.73 (-1.0.024) 0.18 0.73 822 23 24 4 2.0 2 0.20 (-0.46, 0.064) <			Plasma	PBMC	Total	Plasma		Total	Plasma	AAIC	Root	Years	Scaled	Years	Scaled	
825 49 49 6 8.2 60 0.11 (-0.53, 0.74) 0.73 0.89 7259 28 4 4 2.4 9.8 -0.97 (-2.2, 0.30) 0.69 0.28 7265 20 3 4 2.4 2.4 9.8 -0.97 (-2.2, 0.30) 0.69 0.28 1333 38 4 4 1.4 1.7 0.49 (-1.1, 0.13) 0.73 0.25 0.23 1334 36 2 4 1.4 1.7 0.49 (-1.1, 0.13) 0.31 0.25 1335 42 4 1.4 1.7 0.4 1.1 0.40 (-1.1, 0.13) 0.31 0.23 821 5 2 2 2 2 2 0.41 (-1.1, 0.13) 0.31 0.25 821 6 17 4 2 2 0.21 (-2.4, 0.071) 0.41 0.72 822 19 18 1 1 1 1 1 1 1	[2]	2658	69		69	5		5	5.2	64	-1.2 (-2.0, -0.46)	0.42	0.081	0.64	0.12	0.23
7259 28 4 4 2.4 9.8 -0.97 (-2.2, 0.30) 0.69 0.28 7265 20 3 4 2.4 5.4 -0.73 (-1.4, -0.024) 0.18 0.25 1333 38 4 4 1.4 1.7 -0.49 (-1.1, 0.13) 0.31 0.25 13334 36 5 5 2.0 11 -0.49 (-1.1, 0.13) 0.31 0.20 0.20 0.20 0.41 0.20 0.31 0.47 0.23 0.23 0.20 0.20 0.41 0.47 0.23 0.20 0.41 0.42 0.42 0.44 0.47 0.42 0.43 0.44 0.43 0.44 0.44 0.44 0.40 0.41 0.41 0.41 0.41 0.41 0.41 0.41 0.42 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44	[3]	825	49		49	9		9	8.2	09	0.11 (-0.53, 0.74)	0.73	0.089	0.89	0.11	1.0
7265 20 3 3 6.75 84 -0.73 (-1.4, -0.024) 0.18 0.25 1333 38 4 4 1.4 1.7 -0.49 (-1.1, 0.13) 0.31 0.22 1334 36 5 5 2.0 11 -0.49 (-1.1, 0.13) 0.31 0.22 1334 42 4 4 2.0 11 0.47 0.47 0.49 0.17 0.47 0.47 0.49 0.17 0.49 0.21 0.49 0.20 0.21 0.21 0.24 0.23 0.20 0.49 0.17 0.21 0.24 0.21 0.21 0.24 0.21 0.23 0.20 0.23 0.23 0.21 0.21 0.22 0.20 0.11 0.20 </td <td>[4]</td> <td>7259</td> <td>28</td> <td></td> <td>28</td> <td>4</td> <td></td> <td>4</td> <td>2.4</td> <td>8.6</td> <td>-0.97 (-2.2, 0.30)</td> <td>69.0</td> <td>0.28</td> <td>0.61</td> <td>0.25</td> <td>0.79</td>	[4]	7259	28		28	4		4	2.4	8.6	-0.97 (-2.2, 0.30)	69.0	0.28	0.61	0.25	0.79
1333 38 4 4 1.4 1.4 1.7 -0.49 (-1.1, 0.13) 0.31 0.22 1334 36 36 5 2.0 11 -1.2 (-2.4, 0.071) 0.47 0.23 1335 42 4 4 2.0 28 -0.31 (-0.70, 0.079) 0.34 0.17 821 6 178 24 7 17 17 6.5 180 -0.20 (-0.46, 0.064) 0.34 0.17 822 29 119 3 10 10 5.8 70 -1.6 (-2.2, 0.95) 0.50 0.080 824 5 102 14 14 18 13 86 120 -1.6 (-2.2, 0.95) 0.50 0.080 13889 77 65 142 13 14 18 13 13 1.5 (-3.0, 1.5, -0.22) 0.69 0.09 10769 108 16 18 13 5 6 0.09 -1.4, -0.57 0.14 0.14		7265	20		20	ю		ဇာ	0.75	8.4	-0.73 (-1.4, -0.024)	0.18	0.25	0.33	0.44	1.0
1334 36 5 5 20 11 -1.2 (-2.4, 0.071) 0.47 0.23 1336 42 4 20 28 -0.31 (-0.70, 0.079) 0.34 0.17 821 69 178 247 7 17 6.5 180 -0.20 (-0.46, 0.064) 0.46 0.70 824 52 102 119 3 10 10 5.8 70 -1.6 (-2.2, 0.95) 0.50 0.080 1384 7 65 164 7 9 13 8.6 120 -0.87 (-1.5, -0.22) 0.50 0.080 13889 77 65 164 10 4 11 5.4 5.8 -17 (-30, -3.3) 5.6 1.0 10769 108 16 10 4 11 5.4 5.8 -17 (-30, -3.3) 5.6 1.0 34391 12 3 3 3 6 1.3 2 -0.027 (-0.35, 0.30) 0.14 0.11 <		13333	38		38	4		4	1.4	17	-0.49 (-1.1, 0.13)	0.31	0.22	0.30	0.22	1.0
1336 42 4 4 2.0 28 -0.31 (-0.70, 0.079) 0.34 0.17 821 69 178 247 7 17 17 6.5 180 -0.20 (-0.46, 0.064) 0.46 0.000 822 29 119 3 10 10 5.8 70 -1.6 (-2.2, 0.95) 0.50 0.086 824 52 102 154 7 9 13 8.6 120 -0.87 (-1.5, -0.22) 0.69 0.086 13889 77 65 142 13 14 18 13 130 -3.5 (-4.7, -2.2) 1.5 0.11 10769 108 5 6 0.91 21 -0.99 (-1.4, -0.57) 0.12 0.14 34411 14 19 3 3 5 6 0.91 20 -0.027 (-0.35, 0.30) 0.14 0.11		13334	36		36	5		5	2.0	11	-1.2 (-2.4, 0.071)	0.47	0.23	0.44	0.22	0.24
821 69 178 247 7 17 6.5 180 -0.20 (-0.46, 0.064) 0.46 0.70 822 29 90 119 3 10 10 5.8 70 -1.6 (-2.2, 0.95) 0.50 0.086 824 52 102 154 7 9 13 8.6 120 -0.87 (-1.5, -0.22) 0.50 0.086 13889 77 65 142 13 14 18 13 130 -3.5 (-4.7, -2.2) 1.5 0.11 10769 108 5 6 0.91 21 -0.99 (-1.4, -0.57) 0.12 0.14 34411 14 19 3 3 6 1.3 2 -0.027 (-0.35, 0.30) 0.14 0.11		13336			42	4		4	2.0	28	-0.31 (-0.70, 0.079)	0.34	0.17	0.34	0.17	1.0
8222990119310105.870-1.6 (-2.2, 0.95)0.500.08682452102154791386120-0.87 (-1.5, -0.22)0.690.08013889776514213141813130-3.5 (-4.7, -2.2)1.50.111076910856164104115.45.8-17 (-30, -3.3)5.61.0343911235473560.9121-0.99 (-1.4, -0.57)0.120.1434411141933361.322-0.027 (-0.35, 0.30)0.140.11	[3]	821	69	178	247	7	17	17	6.5	180	-0.20 (-0.46, 0.064)	0.46	0.070	0.78	0.12	< 10 ⁻⁵
8245210215479138.6120-0.87 (-1.5, -0.22)0.690.08013889776514213141813130-3.5 (-4.7, -2.2)1.50.111076910856164104115.45.8-17 (-30, -3.3)5.61.0343911235473560.9121-0.99 (-1.4, -0.57)0.120.1434411141933361.322-0.027 (-0.35, 0.30)0.140.11		822	29	06	119	8	10	10	5.8	70	-1.6 (-2.2, 0.95)	0.50	0.086	0.89	0.15	0.0080
13889 77 65 142 13 14 18 13 130 -3.5 (-4.7, -2.2) 1.5 0.11 10769 108 56 164 10 4 11 5.4 5.8 -17 (-30, -3.3) 5.6 1.0 34391 12 35 47 3 5 6 0.91 21 -0.99 (-1.4, -0.57) 0.12 0.14 34411 14 19 33 3 6 1.3 22 -0.027 (-0.35, 0.30) 0.14 0.11		824	52	102	154	7	6	13	8.6	120	-0.87 (-1.5, -0.22)	69.0	0.080	1.3	0.15	0.0020
10769 108 56 164 10 4 11 5.4 5.8 -17 (-30, -3.3) 5.6 1.0 34391 12 35 47 3 5 6 0.91 21 -0.99 (-1.4, -0.57) 0.12 0.14 34411 14 19 33 3 6 1.3 22 -0.027 (-0.35, 0.30) 0.14 0.11		13889	77	65	142	13	14	18	13	130	-3.5 (-4.7, -2.2)	1.5	0.11	1.8	0.13	$< 10^{-5}$
34391 12 35 47 3 5 6 0.91 21 -0.99 (-1.4, -0.57) 0.12 0.14 34411 14 19 33 3 6 1.3 22 -0.027 (-0.35, 0.30) 0.14 0.11	[5]	10769	108	99	164	10	4	11	5.4	2.8	-17 (-30, -3.3)	9.6	1.0	9.7	1.4	0.02
14 19 33 3 6 1.3 22 -0.027 (-0.35, 0.30) 0.14 0.11	[9]	34391	12	35	47	3	5	9	0.91	21	-0.99 (-1.4, -0.57)	0.12	0.14	0.24	0.27	0.017
		34411	14	19	33	3	3	9	1.3	22	-0.027 (-0.35, 0.30)	0.14	0.11	0.25	0.20	0.019

Patient ID corresponds to the anonymized patient identifiers in the LANL database. Time span is in years. AAIC is the Akaike Information Criterion (AIC) [7] of the null model minus the AIC of the linear model. Root is the estimate of the root time in years by the linear model with respect to the time of the first sample. MAE is Mean Absolute Error between collection date and estimated date of the training data. MAD is Mean Absolute Difference is between collection date and estimated date) of the censored data. Scaled MAE/D is the mean absolute error/difference divided by the time span of the training data. The model failed to be calibrated for bold rows.