

Table S1 Log marginal likelihood estimates for two independent fittings of the HIV dataset using the harmonic mean estimator (HME), posterior-simulation Akaike information content (AICM), path sampling (PS) and stepping-stone sampling (SSS). We show the overall ranking of the models in parentheses for each estimator and fitting. For all estimators, we employ equal amounts of computational work (MCMC iterations), as well as equal numbers of samples from which to estimate the marginal likelihood. The HME shows drastic differences in the overall ranking of the demographic models and, depending on the fitting, may very well select a constant population size as the preferred coalescent prior. The AICM is consistent across both fittings but selects a constant population size above all other coalescent priors. PS and SSS consistently select the Bayesian skyline plot (BSP) coalescent prior as the optimal choice and put the constant population size far behind the other coalescent priors.

Coalescent prior	Fitting 1				Fitting 2			
	HME	AICM	PS	SSS	HME	AICM	PS	SSS
Constant	-36189.3 (5)	72705.9 (1)	-37217.4 (5)	-37217.3 (5)	-36176.1 (1)	72698.2 (1)	-37215.4 (5)	-37215.3 (5)
Expansion	-36183.6 (1)	72713.9 (2)	-37039.3 (2)	-37039.2 (2)	-36184.7 (4)	72713.7 (2)	-37034.2 (2)	-37034.0 (2)
Exponential	-36183.6 (2)	72744.4 (4)	-37041.9 (4)	-37041.8 (4)	-36187.4 (5)	72719.1 (4)	-37046.3 (4)	-37046.2 (4)
Logistic	-36188.4 (4)	72756.1 (5)	-37040.9 (3)	-37040.8 (3)	-36181.9 (3)	72723.0 (5)	-37045.9 (3)	-37045.8 (3)
Bayesian skyline	-36186.0 (3)	72734.1 (3)	-37022.8 (1)	-37022.7 (1)	-36180.2 (2)	72714.7 (3)	-37023.0 (1)	-37022.9 (1)

Table S2 Log marginal likelihood estimates for two independent fittings for the HSV dataset (Firth et al., 2010) using HME, AICM, PS and SSS (with the overall ranking of the models shown in parentheses for each estimator) using a strict clock (SC), an uncorrelated relaxed clock with an exponential distribution (UCED) and an uncorrelated relaxed clock with a lognormal distribution (UCLD). Equal amounts of computational work (MCMC iterations) were run for all estimators, as well as an equal number of posterior samples being used to estimate the marginal likelihood. While the HME shows drastic differences in the overall ranking of the (clock) models, the AICM as well as PS and SSS exhibit consistent behaviour, although disagreeing on the performance of a strict clock when the sampling dates are omitted.

Dates	Clock	Fitting 1				Fitting 2			
		HME	AICM	PS	SSS	HME	AICM	PS	SSS
No	SC	-2541.6 (6)	5112.4 (4)	-2842.6 (6)	-2842.7 (6)	-2538.7 (4)	5112.9 (4)	-2842.4 (6)	-2842.5 (6)
No	UCED	-2533.3 (4)	5123.8 (6)	-2769.8 (5)	-2769.9 (5)	-2536.7 (3)	5124.8 (6)	-2771.8 (5)	-2771.8 (5)
No	UCLD	-2537.7 (5)	5113.2 (5)	-2767.1 (4)	-2767.2 (4)	-2539.5 (5)	5113.1 (5)	-2765.6 (4)	-2765.7 (4)
Yes	SC	-2532.9 (3)	5088.0 (1)	-2756.9 (1)	-2756.9 (1)	-2534.6 (2)	5087.3 (1)	-2756.8 (1)	-2756.8 (1)
Yes	UCED	-2532.7 (2)	5102.2 (3)	-2766.7 (3)	-2766.7 (3)	-2528.9 (1)	5101.9 (3)	-2765.3 (3)	-2765.4 (3)
Yes	UCLD	-2532.2 (1)	5090.6 (2)	-2758.2 (2)	-2758.3 (2)	-2541.0 (6)	5091.3 (2)	-2757.4 (2)	-2757.4 (2)