

**Table S1.** Results of the model selection of the different BEAST analyses used to estimate divergence times.

			<i>Model 2</i>							
			<i>Mut-normal</i>				<i>Mut-lognormal</i>			
			<i>yule_strict</i>	<i>yule_rlxln</i>	<i>bd_strict</i>	<i>bd_rlxln</i>	<i>yule_strict</i>	<i>yule_rlxln</i>	<i>bd_strict</i>	<i>bd_rlxln</i>
<i>Model 1</i>	<i>Mut-normal</i>	<i>yule_strict</i>	-	62.2	*	54.3	9.2	59.6	#	57.2
		<i>yule_rlxln</i>	#	-	#	#	#	#	#	#
		<i>bd_strict</i>	#	58.6	-	50.7	-	55.9	#	53.5
		<i>bd_rlxln</i>	#	7.9	#	-	#	5.2	#	*
	<i>Mut-lognormal</i>	<i>yule_strict</i>	#	53.1	#	45.2	-	50.4	#	48
		<i>yule_rlxln</i>	#	*	#	#	#	-	#	#
		<i>bd_strict</i>	<b>6.1</b>	<b>68.3</b>	<b>9.8</b>	<b>60.4</b>	<b>15.3</b>	<b>65.7</b>	-	<b>63.3</b>
		<i>bd_rlxln</i>	#	*	#	#	#	*	#	-

Bayes Factors (BF) were estimated comparing eight BEAST models that combined a mutation prior with either a normal (*Mut-normal*) or lognormal (*Mut-lognormal*) distribution, either a Yule (*yule*) or a Birth-Death (*bd*) model, and either a *strict* or a LOGN relaxed (*rlxln*) clock. Models were compared in a pairwise fashion, by first estimating their Marginal Likelihoods (mL) and corresponding Standard Deviation (SD) and then calculating the Bayes Factors as  $BF = mL_1 - mL_2$ , where model 1 and model 2 are those given in the respective row and column. We only report BF values that satisfied the conditions  $BF > 0$  and  $BF - (SD_1 + SD_2) > 0$ , where  $SD_1$  and  $SD_2$  are the SD values estimated for model 1 and 2. Highlighted on bold are the values for the model (*Mut-lognormal + bd + strict*) favored over all other seven models. # =  $BF < 0$ ; \* =  $BF - (SD_1 + SD_2) < 0$ .