

branch		A to C	A to T	C to G	C to A	A to G	C to T	CG to CA	GC*	GC content
(4,1)	used	0.00080	0.00080	0.00100	0.00100	0.00250	0.00400	0.05000	0.34139	0.42838(47)
	estimated	0.000801(47)	0.000798(45)	0.000996(56)	0.000998(61)	0.00251(11)	0.00400(15)	0.0499(12)	0.3419(97)	-
(4,2)	used	0.00080	0.00080	0.00100	0.00100	0.00250	0.00400	0.05000	0.34139	0.42838(47)
	estimated	0.000800(48)	0.000799(46)	0.000996(61)	0.000994(62)	0.00251(12)	0.00400(15)	0.0499(13)	0.3421(98)	-
(0,4)	used	0.00320	0.00320	0.00400	0.00400	0.01000	0.01600	0.20000	0.34139	0.43190(48)
	estimated	0.0033(19)	0.0035(27)	0.0040(17)	0.0042(25)	0.0158(81)	0.0119(93)	0.2006(41)	0.45(12)	0.43174(52)
(0,3)	used	0.00400	0.00400	0.00500	0.00500	0.01250	0.02000	0.25000	0.34139	0.42838(48)
	estimated	0.0039(20)	0.0038(27)	0.0050(17)	0.0049(24)	0.0159(77)	0.0132(99)	0.2495(47)	0.44(10)	-
root node	used	-	-	-	-	-	-	-	-	0.45001(51)
	estimated	-	-	-	-	-	-	-	-	0.4426(51)

Supplementary Table S2: Nucleotide and substitution frequencies for synthetic sequence data on the phylogeny in Figure S2. For each branch (i, j) we report the used and estimated substitution frequencies, the stationary GC content, GC*, and the GC content at the descendant node j . For leaf nodes $j = 1, \dots, 3$ the used and estimated values coincide and only one is given. We generated synthetic alignments of 1 Mbp length and report the mean values and standard deviations from 500 independently repeated experiments. The standard deviations in the last digits are given in brackets, e.g. $0.000801(47) = 0.000801 \pm 0.000047$. The substitution frequencies on the two edges connected to the root, (0,3) and (0,4), as well as the GC content at the root node (gray shaded areas) cannot reliably be reconstructed, as mentioned in the text.