Supplementary file SM2: Ordered *versus* unordered characters

The best models within each partitioning scheme for CEA with ordered characters are listed in SM2 Table 1, with their respective values of MgL, and KRS. Besides unpartitioned and homoplasy partitioned data, a single anatomical scheme was analyzed (see "Methods" for details) as well as schemes C-ParFin-1 and C-ParFin-2 (discussed in section "CEA"). There was Bayesian preference for linked branch lengths in all partitioned analyses, except when employing the anatomical (C-Ana-ORD) scheme and for λ=10 except in the case of C-Unp(Ord)-7 (λ=20). ACRV was either not modeled or accommodated via shared or per-partition Γ. All alternative schemes were very strongly rejected in favor of homoplasy (KRS > 59), which employed the same parameters as in all other analyses (i.e. linked branch lengths, equal rates, λ=10; SM2 Table 1). Like in the case of CEA analysis with unordered characters, all PF models were equivalent (KRS < 0.60 for all pairwise comparisons, calculations not shown). Ordering characters greatly improved marginal likelihoods across the board. KRS computed from pairwise comparisons within schemes exceeded 56 in the case of unpartitioned data, 48 in the case of anatomical partitioning, 44 for homoplasy and 62 for PF (calculations not shown).

The distributions of rate multipliers by partition for ordered CEA data (SM2 Figure 1) were remarkably similar to those obtained from the analysis of unordered data (SM2 Figure 1), and so were the joint posterior of tree lengths and multipliers, but the corresponding marginal density was smoother and more concentrated around the mean. These distributions indeed differ significantly ($D=0.14$, $p << 0.001$) but increase in median length was modest (around 3%).

The anatomical scheme employed in the analysis of ordered CEA resulted in more pronounced variation in tree length posteriors across partitions than the scheme applied to the unordered data (SM2 Figure 1a). Rate distributions were also centered around 1 but their masses were somewhat truncated below the median (SM2 Figure 1a). When integrated across partitions, the tree length posterior was clearly bimodal, while the marginal density of rates seems to be a mixture of at least two distributions with very different variances. Most of the joint posterior is concentrated on its lower bivariate mode (SM2 Figure 2b, above), around significantly shorter tree lengths ($D=0.89$, $p << 0.001$) than obtained under homoplasy partitioning (SM2 Figure 2a).

The top PF model employed linked branch lengths and shared Γ. Multiplier posteriors were clearly segregated across partitions while rates had strongly skewed distributions with long tails (SM2 Figure 1c). In log scale, marginal multiplier distribution had most of its mass concentrated in the left mode (Supplementary Fig. S4e). Tree length distribution was continuous and its median was about 18% shorter than those obtained under homoplasy (D=0.40, p << 0.001, Supplementary Fig. S3i) but ~10% longer than those obtained for the top unordered model applied to a PF partitioned matrix (D=0.45, p << 0.001, Supplementary Fig. S3c).

In the anatomical partitioning, when CEA characters were treated as ordered, the top anatomical model had the same parameters as the ones applied to the unordered characters, but it was obtained under C-Ana2 (3 partitions) instead of C-Ana1 (2 partitions - SM2 Table 1). The former scheme seemingly captured some of the APRV due to the obvious differences in tree lengths, but did so inefficiently, as denoted by the multimodality of within-partition rate variation (SM2 Figure 1b). The top model (C-Ana2(Ord)-22) also recovered an odd bimodality in tree lengths which was not observed in other analyses (SM2 Figure 1b). Nevertheless, model simplification due to state change restriction probably led to very strong Bayesian preference for ordered characters.

SM2 Table 1. Best models by criterion for CEA with coding variable when characters were ordered. Best model overall in bold face.

Model	Partitions	APRV	ACRV	λ	MgL	KRS[b]	KRS[c]
C-Unp(Ord)-6	1	N/A	Shared Γ	10	-1596.42	113.56	0.00
C-Unp(Ord)-7	1	N/A	Shared Γ	20	-1597.11	114.94	1.38
C-Unp(Ord)-2[a]	1	N/A	Equal rates	10	-1602.24	125.20	11.64
C-Ana(Ord)-22	4	Unlinked branch lengths	Shared Γ	10	-1588.07	96.86	0.00
C-Hom(Ord)-2	**8**	**Linked branch lengths**	**Equal rates**	**10**	**-1539.64**	**0.00**	**0.00**
C-ParFin1(Ord)-10	2	Linked branch lengths	Shared Γ	10	-1569.11	58.94	0.00
C-ParFin1(Ord)-6	2	Linked branch lengths	Per-partition Γ	10	-1569.14	59.00	0.06
C-ParFin1(Ord)-2	2	Linked branch lengths	Equal rates	10	-1569.31	59.34	0.40
C-ParFin2(Ord)-10	4	Linked branch lengths	Shared Γ	10	-1569.38	59.48	0.54
C-ParFin2(Ord)-6	4	Linked branch lengths	Per-partition Γ	10	-1569.42	59.56	0.62
C-ParFin2(Ord)-2	4	Linked branch lengths	Equal rates	10	-1569.43	59.58	0.64

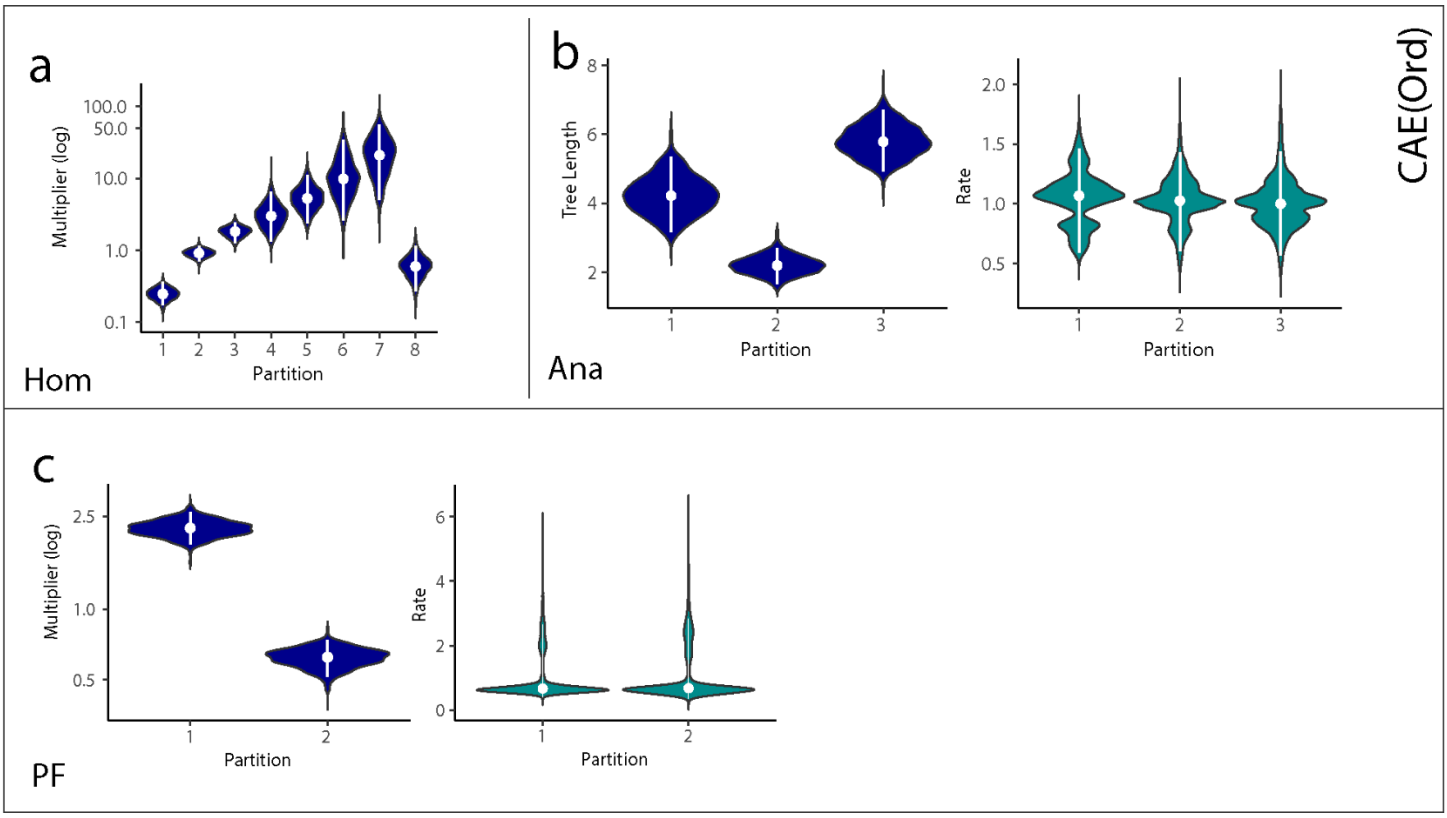
Abbreviations: Part., Partitions; APRV, Among partition rate variation; ACRV, Among character rate variation; λ, inverse scale parameter of the exponential branch length prior; MgL, Marginal likelihood; KRS, Kass & Raftery's statistic.

[a] Default model.

[b] Kass & Raftery's statistic computed with respect to the best model overall.

[c] Kass & Raftery's statistic computed with respect to the best model within each criterion.

SM2 Figure 1. Violin plots representing per-partition Bayesian posterior distributions of tree lengths (model employed unlinked branch lengths to accommodate among partition rate variation APRV), rate multipliers (linked branch lengths) and rates of character evolution (ACRV was approximated by a shared or per-partition Γ distribution). Multipliers and tree lengths are represented by blue and rate posteriors by green violins. Distributions are shown only for the top models within each partitioning strategy, listed in Table 3 and indicated in each panel (Hom=homoplasy, Ana=anatomy, PF=PartitionFinder2). Panels are arranged in rows corresponding to the dataset, indicated on the right end of the row (CEA-Ord). Rate multipliers are represented in multiplicative (log_{10}) scale. Note that violin widths are scaled with reference to their own partitions. a) C-Hom(Ord)-2, b) C-Ana(Ord)-22 and c) C-ParFin1(Ord)-10.



SM2 Figure 2. a-c. Heatmaps of joint posterior distributions representing tree length variation as a function of APRV (rate multipliers) or ACRV (character rates). Marginal densities are represented by the smoothed histograms on top and to the right of each heatmap, with 95% HPD highlighted in red. Model details in Table 3. a) C-Hom(Ord)-2 b) C-Ana(Ord)-22 c) C-ParFin1(Ord)-10. d. Joint posterior distributions of tree lengths and rate multipliers represented as semi-log plots (y-axis in linear and x-axis in log scale) of the C-Hom(Ord)-2.