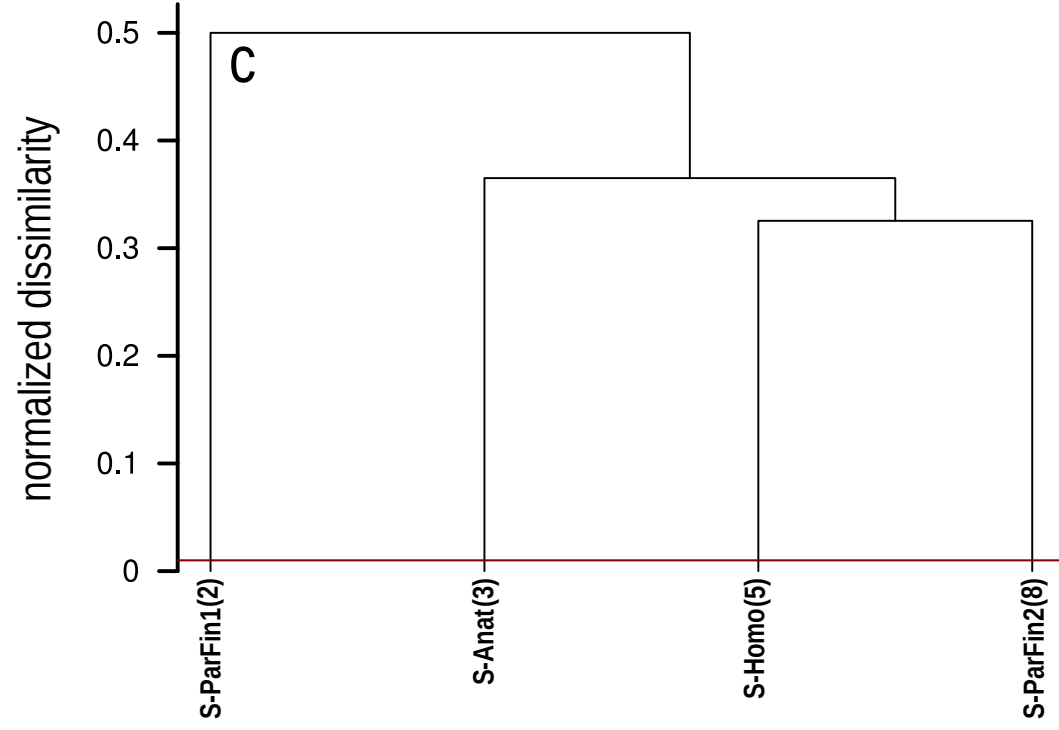
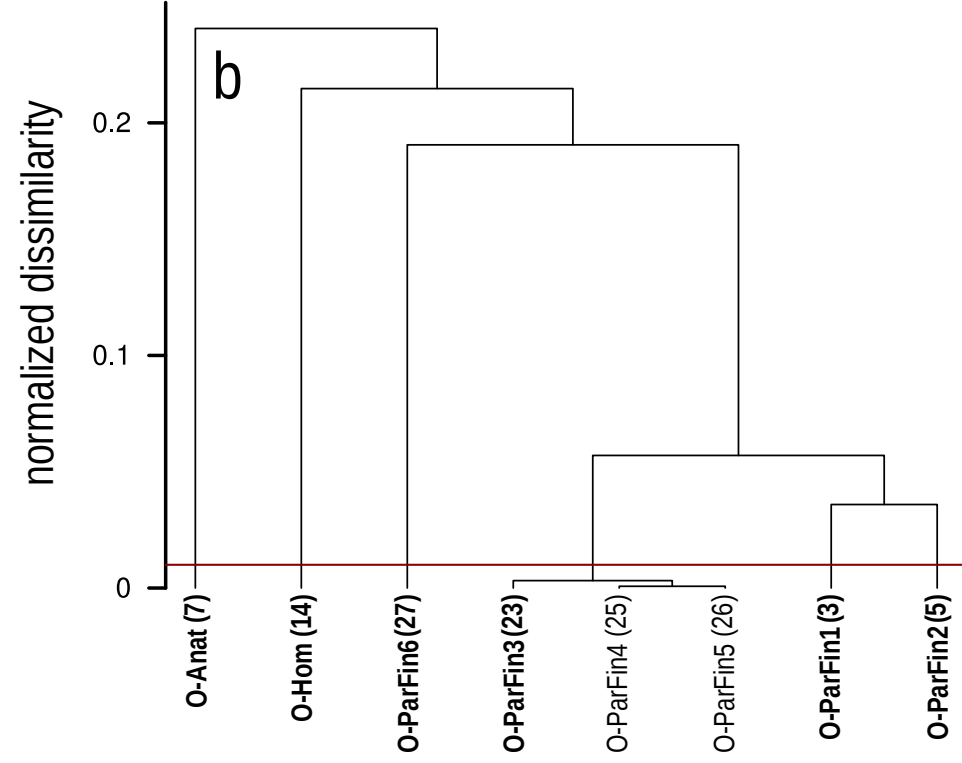
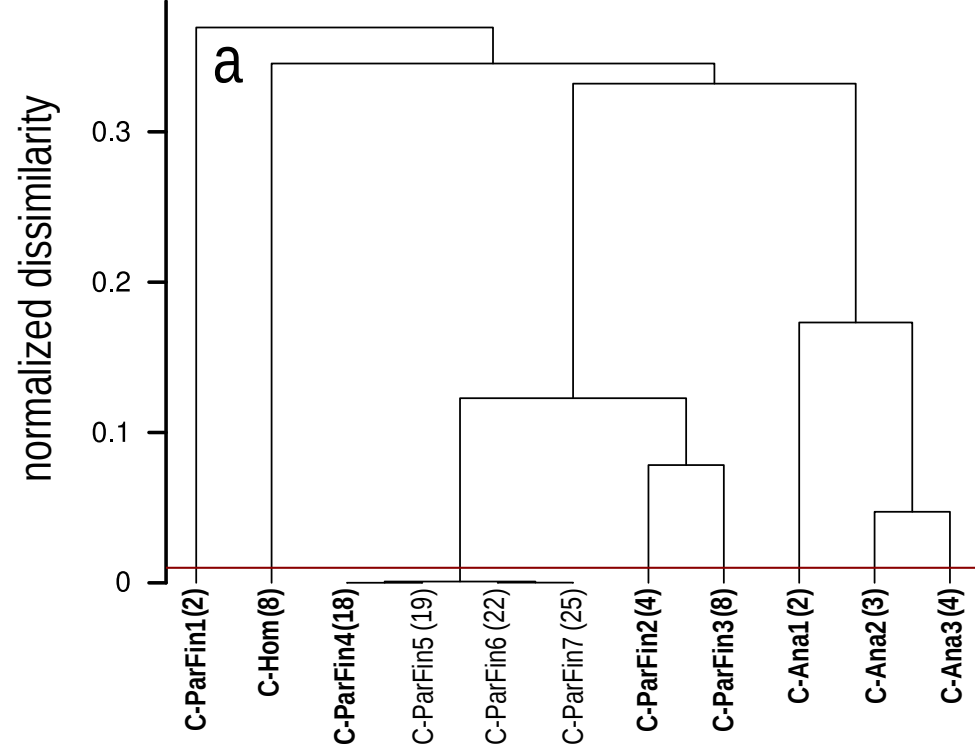
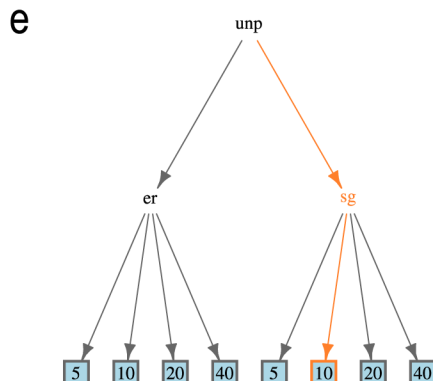
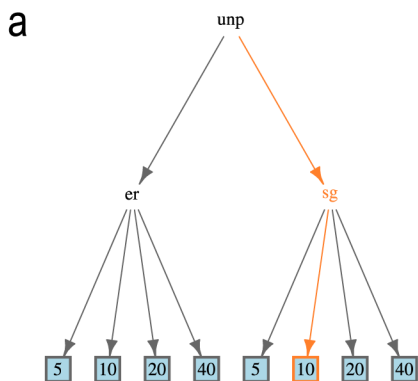


Supplementary Material Figure S1. Single linkage dendrograms of dissimilarities among partition schemes. Each scheme was represented as binary matrix and pairwise dissimilarities were estimated as Hamming distances (see the Methods section for details). Distances were normalized (re-scaled) to the 0-1 range, where 0 represents identical schemes and 1 the maximum theoretical dissimilarity for each dataset. Each scheme is identified by the acronym of the corresponding strategy (hom=homoplasy, ana=anatomy, pf=PartitionFinder2) and numbered in ascending order of the corresponding number of partitions (in parenthesis). Red lines indicate 1% of the maximum possible distance. Analyzed schemes are in bold face. **a.** Partition schemes for the CEA dataset. Note that PF schemes 4-7 fell below the dissimilarity threshold, thus only the scheme with the smallest number of partitions (pf 4) was used in downstream analyses. **b.** Partition schemes for the OZL dataset. Relative similarity of PF schemes 3-5 also exceeded 99% and hence only pf 3 was selected. **c.** Partition schemes for the OZL dataset. All schemes were analyzed.

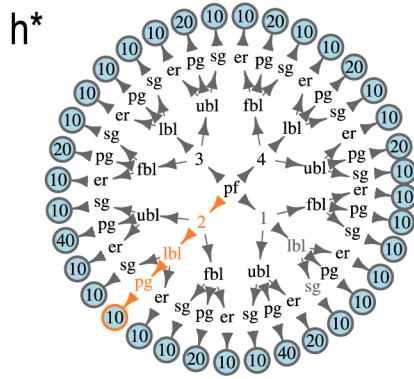
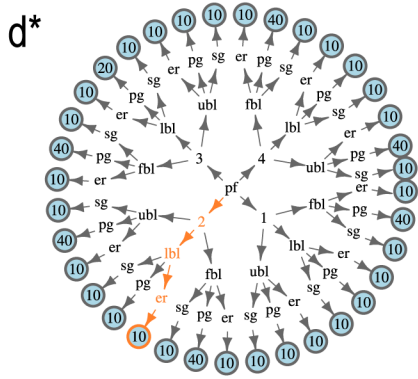
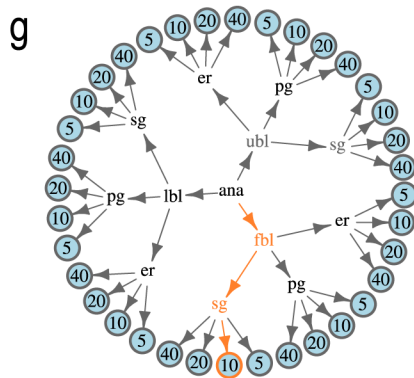
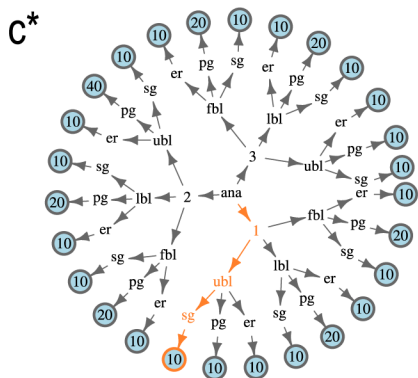
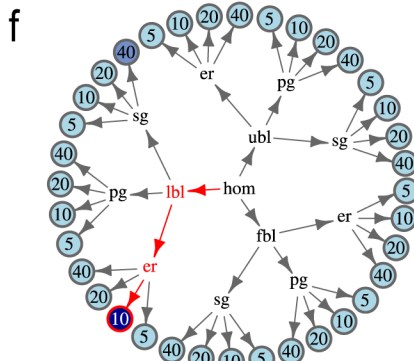
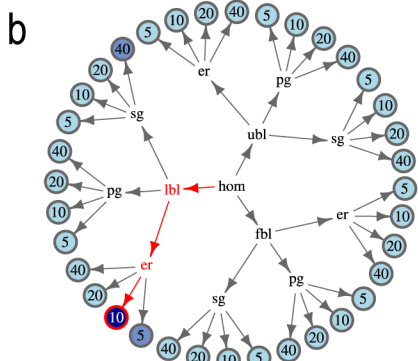


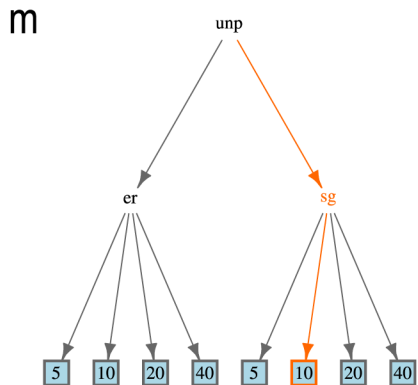
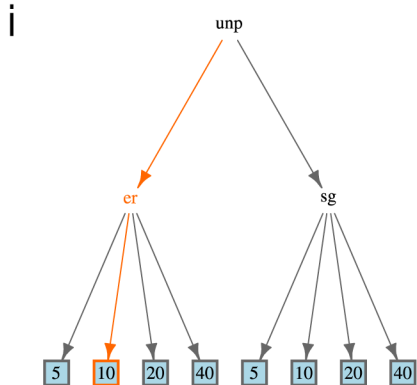
Supplementary Material Figure S2. Comparisons of the stepping-stone (SS) estimation of marginal likelihood for all tested partitioning strategies and models. Root vertices (node) indicate strategy (unp=unpartitioned, ana=anatomic, hom=homoplasy, pf=Partition Finder). Starting from the root towards the terminals, numbers (if present) represent partitioning scheme when more than one was tested. Internal nodes labeled “fbl” (fixed branch lengths), “lbl” (linked branch lengths) and “ubl” (unlinked branch lengths) indicate how APRV was modeled. Nodes labeled “er” (equal rates), “sg” (shared gamma) and “pg” (per-partition gamma), correspond to among characters rate variation (ACRV) accommodation and numbers within leaves are the exponential distribution inverse scale parameter λ (5, 10, 20 and 40). Due to space constraints, only the value of λ corresponding to top model is shown in some graphs (indicated by * close to their labels). Leaves are color-coded according to the level of significance of Kass & Raftery’s (1995) statistic (KRS), reported in the blue scale. Best model overall is highlighted in red and best model within each partitioning strategy in orange. Figs 1a-1d correspond to Clarke et al.’s matrix (CEA), 1e-1h to O’Connor and Zhou’s matrix (OZL), 1i-1l to the Sclebythidae matrix (SCO) and 1m-1p to ordered CEA. See Table 3 for further details on each model.



KRS

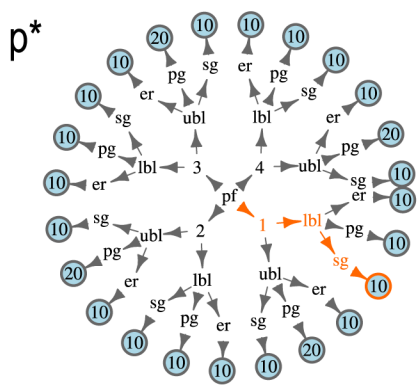
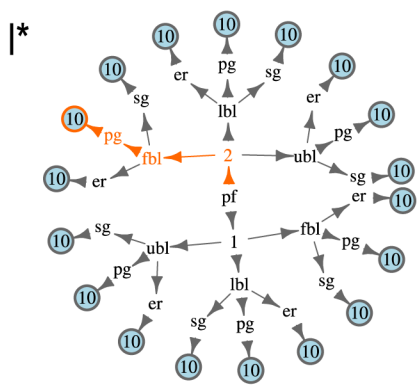
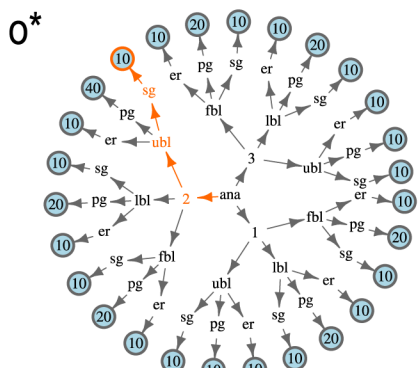
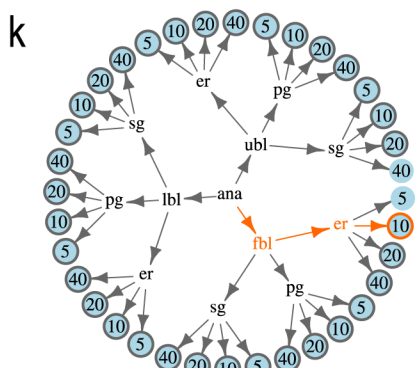
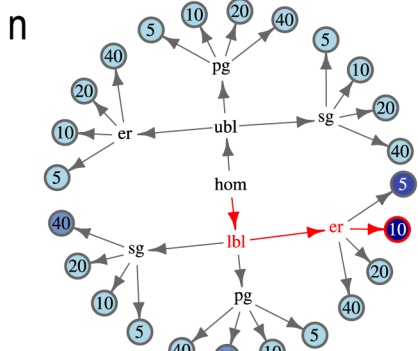
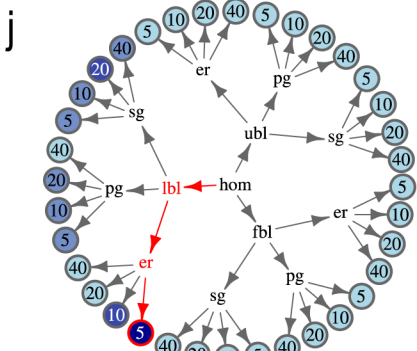
- Best model
- 0-2
- 2-6
- 6-10
- >10





KRS

- Best model
- 0-2
- 2-6
- 6-10
- >10



Supplementary Material Figure S3. Probability-probability (p-p) scatterplots of the per-partition posterior distributions of rate multipliers. Partitions's numbers are in headers. The blue lines represent correlations between the observed (x-axis) and the expected (y-axis) cumulative distribution functions. The gray slope represents perfect agreement between the posterior and the expected normal distribution with identical mean and standard deviation. The orange envelope is the 95% confidence interval. **a.** CEA. **b.** OZL. **c.** SCO.

