Homoplasy-based partitioning outperforms alternatives in Bayesian analysis of discrete morphological data

BRUNNO B. ROSA[1,3], GABRIEL A. R. MELO[1] and MARCOS S. BARBEITOS[2,3]

[1]*Laboratório de Biologia Comparada de Hymenoptera, Departamento de Zoologia, Universidade Federal do Paraná, Caixa Postal 19020, Curitiba, Brazil, 81530-980.*

[2]*Laboratório de Evolução dos Organismos Marinhos, Departamento de Zoologia, Universidade Federal do Paraná, Curitiba, Brazil, 81530-980.*

[3]*Corresponding authors: brunnobueno27@gmail.com / msbarbeitos@gmail.com*

ABSTRACT

Bayesian analysis of morphological data is becoming increasingly popular mainly (but not only) because it allows for time-calibrated phylogenetic inference using relaxed morphological clocks and tip dating whenever fossils are available. As with molecular data, recent studies have shown that modeling among character rate variaton (ACRV) in morphological matrices greatly improves phylogenetic inference. In a likelihood framework this may be accomplished, for instance, by employing a hidden Markov model (HMM) to assign characters to rate categories drawn from a (discretized) Γ distribution and/or by partitioning datasets according to rate heterogeneity and estimating per-partition branch lengths, conditioned on a single topology. While the first approach is available in many phylogenetic analysis software, there is still no clear consensus on how to partition data, except perhaps in the simplest cases (e.g. "by codon" partitioning of coding sequences). Additionally, there is a trade-off between improvement in likelihood scores and the number of free parameters in the analysis, which rises quickly with the number of partitions. This trade-off may be dealt with by employing statistics that penalize overfitting of complex models, such as Akaike or Bayesian information criteria (AIC and BIC), or the more recently introduced stepping-stone (SS) method for marginal likelihood approximation. We applied the latter to three distinct matrices of discrete morphological data and demonstrated that sorting characters by homoplasy scores (obtained from implied weighting parsimony analysis) outperformed other partitioning strategies (anatomically-based and PartitionFinder2). The method was in fact so efficient in segregating characters by rates of evolution that no within-partition ACRV modeling was necessary, while among partition rate variation (APRV) was adequately accommodated by rate multipliers. We conclude that partitioning by homoplasy is a powerful and easy-to-implement strategy to address ACRV in complex datasets. We provide some guidelines focusing on morphological matrices, although this approach may be also applicable to molecular datasets.

Key words: TNT, MrBayes, stepping-stone, homoplasy, implied weighting parsimony, Aves, fossil, Hymenoptera.

INTRODUCTION

Although DNA-sequence data have come to dominate phylogenetic inference, Bayesian analysis of morphological data is gaining increasing attention in recent years, especially for estimation of divergence time using both molecular and morphological characters in a total evidence approach (Quental and Marshall 2010; Wiens et al. 2010; Pyron 2011; Ronquist et al. 2012a; Lee et al. 2013, 2014; Dávalos et al. 2014; Giribet 2015; Lee and Palci 2015; Lee 2016; King et al. 2017). Bayesian inference based on morphology-only datasets is expected to continuously increase with possibility of novel studies investigating morphological evolution (Clarke and Middleton 2008; Klopfstein et al. 2015) and of inferring dated phylogenies under morphological clocks (Ronquist et al. 2012a; Lee et al. 2013, 2014; Lee and Palci 2015; Lee 2016).

Bayesian analysis of morphological data has been based on Lewis' (2001) Mk model. It was developed as a generalization of the Jukes-Cantor model (Jukes and Cantor 1969), with the single estimated parameter being the instantaneous rate of transition between character states (Lewis 2001). More recent developments on the analysis of discrete morphological data involve choice of the best distribution for among-character rate variation (Wagner 2012; Harrison and Larsson 2015), nonstationary models (Klopfstein et al. 2015) and choice of the best hyperprior for modeling state frequencies (Wright et al. 2015). Overall the main concern has been in modeling heterogeneity in rates of evolution across characters.

An alternative (but not mutually exclusive) approach is to segregate characters into separate subsets or partitions according to their rates of evolution (Harrison and Larsson 2015; Wright 2015). Data partitioning of molecular data has become commonplace and it has been shown to strongly influence tree topology, branch lengths, clade support and final likelihoods (Brandley et al. 2005; Marshall et al. 2006; Brown and Lemmon 2007; Lemmon and Moriarty 2004; Kainer and Lanfear 2015). Considering that genes and gene regions (e.g. exons vs. introns, codon positions, paired vs. unpaired regions in ribosomal genes) may evolve under distinct rates, sorting them into subsets allows for accommodation of heterogeneity on evolutionary rate. Partitioning of molecular data is currently carried out from a functional perspective, i.e. sorting by gene and gene regions, or under an automated site-rate approach, such as that implemented in PartionFinder (Lanfear et al. 2017). These approaches have been called, respectively, "ad hoc" and "algorithmic" by Kainer and Lanfear (2015).

Methods aimed at proposing partitions for morphological data are still incipient. Clarke and Middleton (2008) were the first to analyze morphological data broken up in subsets based on morpho-functional

knowledge of the studied characters. In their case, the subsets corresponded to characters derived from specific body regions, such as pelvic and pectoral bones, whose set elements are assumed to evolve independently from those of other sets. More recently, Clarke and Middleton's principles were followed by Tarasov and Génier (2015) for inferring relationships among dung beetles and by Lee (2016) to infer a dated phylogeny for mammals applying total-evidence multiple clocks chosen through grouping of partitions with similar relative branch lengths. This approach is comparable to the functional approach used for molecular data and is referred here as application of an anatomical criterion for partitioning of morphological data.

In parallel to the automated site-rate approach applied to molecular data, the current version of PartitionFinder (2.1.1; Lanfear et al. 2017) implemented the ability of analyzing morphological datasets. Using a preliminary version of PartitionFinder, Wright (2015) evaluated a large number of morphology-only datasets, including that of Clarke and Middleton (2008), and found support for the use of partitions in most of them. As in both Clarke and Middleton (2008) and Tarasov and Génier (2015), data partitioning considerably outperformed the results of the unpartitioned datasets. However, the partition schemes returned by PartitionFinder for Clarke and Middleton's dataset in Wright (2015) resulted in lower marginal likelihood scores compared to the original anatomical partitions used by Clarke and Middleton (2008).

Here we introduce a new method based on homoplasy as a criterion for partitioning of discrete morphological characters. We rely on the assumption that levels of homoplasy (that can be efficiently estimated from maximum parsimony trees) can be a good proxy for the evolutionary rate of the characters, because fast evolving characters will also be highly homoplastic. Therefore, partitions containing characters with similar homoplasy levels are expected to be evolving at similar rates. An analogous approach developed by Kjer and Honeycutt (2007) was applied to molecular datasets, with further modifications better suited to phylogenomic matrices proposed by Misof et al. (2014).

Bayesian analyses under models spanning a wide range of complexity were carried out using unpartioned datasets and matrices partitioned by different "algorithmic" (including homoplasy partitioning) and "ad hoc" strategies. Results were compared using Bayes Factors (BF) computed from marginal likelihoods (MgL) obtained by stepping-stone (SS) importance sampling (Xie et al. 2011).

MgL is computed via SS as the weighted average of posterior likelihoods (the probability of the data given the model), being the weights provided by the prior. MgL scores are proportional to the degree of overlap

between prior and posterior densities. Like Akaike's or Bayesian information criteria, SS sampling also penalizes overparameterization. Complex models tend to have flatter prior landscapes due to the higher dimensionality of the parameter space and hence a poorer match between joint prior and posterior distributions. Thus, MgL scores obtained via SS tend to be worse for parameter-rich models than for simpler ones, unless the latter fail to adequately capture the complexity of the data.

In this study we show that homoplasy partitioning was the most efficient way to segregate characters according to their rates of evolution. Under this scheme, there was no need for modeling rate variation among characters within partitions, whereas among-partition variation was adequately accommodated by rate multipliers. Character sorting using alternative strategies was not nearly as accurate, generating partitions with higher rate heterogeneity. Therefore, modeling rate variation under such schemes requires more complex models and priors, resulting in MgL scores significantly smaller than those obtained under homoplasy partitioning.

MATERIALS AND METHODS

Data sets

Three matrices of discrete morphological data were used in the analyses (Table 1). Clarke and Middleton (2008) were the first to establish a partition criterion for discrete morphological data on birds, using the matrix from Clarke's et al. (2006). For this reason, we chose this matrix (henceforth called CEA) for comparisons with Clarke and Middleton's partition proposals. The second matrix (OZL) was obtained from Lee et al. (2014), who used an expanded and slightly modified version of O'Connor and Zhou's (2012) dataset in explorative analyses of bird diversification employing morphological clocks (see Lee et al. 2014 for details). The third matrix contains morphological data of a rare family of aculeate wasps, Scolebythidae (SCO), chosen for the small number of terminals and a large set of fossil taxa. The scolebythid matrix was assembled originally by Carpenter (1999) and has been modified, with addition of characters and fossil taxa, by Engel and Grimaldi (2007) and Engel et al. (2013). We also added data from *Clystopsenella mirabilis* Engel (extinct species from Dominican amber) and *Pristapenesia asiatica* Azevedo et al. (an extant species from Asia), as well as an additional character (Supplementary File SM1).

Partitioning

Two partitioning criteria were applied to all three data sets in order to compare with the unpartitioned scheme. First, we used anatomical partitioning according to Clarke and Middleton (2008). Under this approach, characters are allocated to partitions using morpho-functional criteria, further detailed in the following sections on each dataset. Second, characters were partitioned according to their homoplasy levels i.e., they were mapped onto partitions corresponding to their homoplasy scores. Farris (1969) introduced the consistency index (CI) as a measure of character homoplasy on a given tree. CI is defined as the ratio between the minimum number of state changes in a matrix and the actual number of steps on a given tree. When CI=1, the number of steps on the tree is the same as the minimum number of changes in the matrix. The character therefore has 0 homoplasy. Goloboff (1993) realized that CI penalized multistate over binary characters because the former would, by definition, have more changes on any given matrix if more than two states were found in the terminals. He improved on this measure by introducing a function that corrects for the difference in state numbers.

We used Goloboff's unbiased measure of homoplasy (*f*) implemented in the software TNT (Goloboff et al. 2008), with the default concavity parameter (k=3). The values returned by the implied weight analysis are normalized between 0 and 1, with the lowest value corresponding to no homoplasy. No values are assigned to non-informative characters, and these were assigned to their own partition. Table 2 provides a summary of the main partition schemes and the abbreviation adopted for each scheme by criterion. Additionally, each model has an individual sequential number available in Supplementary Tables S1–S5. The partition schemes for each data set are detailed below.

We also used the newly released version of PartitionFinder 2.1.1 (Lanfear et al. 2017), which now includes a script for partitioning of morphological characters. Datasets were analyzed under all possibilities of available parameters (using the option *--min-subset-size*=1), as follows: (1) Tree topology: neighbor joining (default), parsimony (equal and implied weighting) and maximum likelihood; (2) Branch lengths: linked and unlinked; (3) Models of evolution: multistate+G+A; and (4) Model selection: AIC, AICc and BIC. We tested 24 parameter combinations for each dataset, totaling 72 tested models (Supplementary Table S6).

Single-partition solutions and redundant schemes returned under different settings were not considered. PartitionFinder (PF) returned up to 7 distinct schemes for a single dataset (Supplementary Table S6). Given the

large number of models to be tested in downstream phylogenetic analyses of each scheme (described below), we attempted to reduce the workload by setting a dissimilarity threshold below which schemes were considered too similar to deserve further evaluation. For all criteria (homoplasy, anatomical and PF) each partitioning scheme was represented as a square matrix *M,* defined by M_{ij}=1 if the i^{th} and j^{th} characters were assigned to the same partition and M_{ij}=0 if otherwise. This is analogous to the representation of phylogenetic trees as adjacency matrices, whose cells are set to 1 if corresponding taxa belong to the same clade or to 0 if they belong to different clades. We then computed pairwise Hamming (or editing) distances between all matrices within each dataset. Distances were rescaled between 0 and 1, where 1 is the maximum possible distance between matrices (i.e. when no characters shared the same partition between them) and represented as dendrograms using single linkage as the clustering method (Supplementary Fig. S1). Dissimilarity among schemes had a clear bimodal distribution for CEA and OZL datasets, which had a number of PF schemes well below the arbitrarily established dissimilarity threshold of 1% (red line Supplementary Fig. S1). In such cases, only the PF scheme with the smallest number of partitions within its cluster was further evaluated (highlighted in bold face in Supplementary Fig. S1).

Aves: Clarke et al. dataset.— Under the anatomical criterion we used three partition schemes previously established by Clarke and Middleton (2008). The first scheme contains two partition sets (Supplementary Table S7): pelvic plus cranial plus axial characters (n=122) and pectoral characters (n=82). The second partition scheme has three sets (Supplementary Table S8): cranial plus axial (n=71), pectoral (n=82) and pelvic characters (n=51). The third scheme has four partition sets (Supplementary Table S9): cranial (n=52), axial (n=19), pectoral (n=82) and pelvic characters (n=51). Under the homoplasy criterion eight different values of homoplasy were returned (Supplementary Tables S10 and S11) and characters were sorted in eight corresponding partitions. Under the PF criterion, we selected four schemes that exceeded the dissimilarity threshold explained above, with 2, 4, 8 and 18 partitions, respectively (Supplementary Fig. S1). Also, see the Supplementary Tables S12 to S18 for the partition sets returned by PF.

Aves: O'Connor and Zhou's dataset modified by Lee et al.— Under the anatomical criteria we used the different body sections of the character list in O'Connor and Zhou (2012) (Supplementary Table S19): skull and mandible (n=48), vertebral column and ribs (n=33), thoracic girdle and sternum (n=39), thoracic limb (n=57), pelvic girdle (n=21), pelvic limb (n=47), and integument (n=2). Under the homoplasy criterion 14 distinct

values of homoplasy scores were obtained (Supplementary Tables S20 and S21) and characters were also partitioned according to these values. We also evaluated three PF schemes, with 3, 5 and 23 partitions (Supplementary Fig. S1). Also, see the Supplementary Tables S22 to S27) for the partition sets returned by PF.

Scolebythidae dataset.— Under the anatomical criterion the matrix was divided into three partitions (Supplementary Table S28): head (n=7), mesosomal plus one metasomal character (n=12) and wing characters (n=9). According to the homoplasy criterion the Scolebythidae matrix exhibited five different classes of homoplasy values and hence five partitions (Supplementary Tables S29 and S30). We also tested the two schemes returned by PF, with 2 and 8 partitions, respectively (Supplementary Fig. S1). Also, see the Supplementary Tables S31 and S32 for the partition sets returned by PF.

Parameters and Priors

All possible permutations of the tested parameters and priors were analyzed for each dataset, resulting in a total of 1624 models. For CEA, we tested 1016 different models, 304 for OZL and 304 for the SCO dataset, respectively. Details of the tested parameters are provided below.

Ascertainment bias (Coding).— The implementation of the Mk model (Lewis 2001) in MrBayes allows for correction of the ascertainment bias exhibited by morphological datasets, in which only variable characters are coded. We tested two coding corrections: (1) *lset coding=variable*, which corrects for lack of non-variable characters in the matrix; and (2) *lset coding=informative*, which in addition to lack of non-variable characters also assumes that only parsimony-informative characters have been scored (Ronquist et al. 2011). We evaluated the effect of coding on unpartitioned, homoplasy- and anatomically-partitioned datasets. Informative coding resulted in much better marginal likelihoods in several analyses (as obtained by Tarasov and Génier 2015), whereas others crashed repeatedly before completion. Conversely, mixed-coding trial runs in which non-informative characters (i.e. characters with a single state change) were appropriately scored as variable, and the remainder as informative, returned much lower marginal likelihoods scores than observed for single-coding runs (data not shown). This pattern was consistent across all datasets. The problems arising when characters were scored as informative led us to focus only on variable-coding datasets because (in all completed analyses) the best models were the same for either ascertainment bias correction. We reported all comparisons between

variable and informative coding using Bayes Factors in Supplementary Table S5, but data using informative coding were omitted from figures and tables in the Results section.

Among-partition rate variation (APRV) on branch lengths.— Following Marshall et al. (2006) and Clarke and Middleton (2008), we also evaluated distinct priors of rate variation among partitions implemented in MrBayes: (1) *prset ratepr=fixed*, which is fixed to the average state change rate across partitions; (2) *prset ratepr=variable*, which allows for variation among partitions, while constraining the average rate of change across partitions to 1. Branch lengths remain linked among partitions through rate multipliers which are, by default, drawn from a flat Dirichilet prior; (3) *unlink brlens=(all)*, which allows for branch lengths to be independently estimated across partitions.

Among-character rate variation (ACRV).— In addition to the rate variation among partitions it is also possible to model within-partition rate variation, assuming an underlying Γ distribution whose shape parameter (α) is sampled in MrBayes from a uniform prior in the (0.05,50) interval. MrBayes allows the following settings: (1) *lset rates=equal* and *link shape=(all)*, which assumes no rate variation across characters; (2) *lset rates=gamma* and *link shape=(all)*, which assumes a single Γ distribution for all partitions, also referred to as shared Γ; and (3) *lset rates=gamma* and *unlink shape=(all)*, in which each partition has its own Γ distribution, also referred as per-partition Γ.

Branch length prior (BLP).— We also tested the effect of different exponential priors for branch length using the inverse scale parameter (λ) suggested by Marshall et al. (2006) for morphological characters and previously evaluated by Clarke and Middleton (2008), i.e. λ=5, λ =10, λ=20 and λ=40.

Ordered versus unordered characters

Restrictions to changes between states of multistate characters are sometimes applied in analysis of morphological data, resulting in a preset ordering of the state changes. Here the characters were left unordered in the analyses of all datasets. Considering that Clarke and Middleton (2008) ordered 38 of the 205 characters, we also reanalyzed CEA implementing these same restrictions. For model comparison, however, we calculated MgL using SS (see below) and not harmonic means as done by these authors. The same parameters described above were applied.

Analysis and Model testing

The analyses were conducted in MrBayes 3.2.6 (Ronquist et al. 2012b) through the CIPRES Science Gateway portal (Miller et al. 2010) and the Ohio Supercomputer Center (OSC, 1987). The different models were compared using BFs, interpreted according to Kass and Raftery (1995). Because MrBayes reports the natural log of marginal likelihoods (MgL), Bayes factors are calculated not as the ratio, but the difference between the ln(MgL) of competing models, hence they are also reported as the natural log of the ratio, or ln(BF). Kass & Raftery's statistic (KRS) is 2·ln(BF), or twice the pairwise differences between the MgLs of the most (H_1) and least (H_0) likely hypotheses (or the models with the highest and lowest MgL). Under Kass and Raftery's criteria, if this statistic is between 0 and 2, evidence against H_0 is "not worth more than a bare mention". Values between 2 and 6 indicate "positive evidence" against H_0, from 6 to 10 should be interpreted as "strong evidence" and over 10, "very strong evidence" against the model with lowest MgL.

Marginal likelihoods were computed following the stepping-stone method (Xie et al. 2011). We employed 50 steps in MgL estimation via SS, while the number of generations and sampling frequencies varied according across datasets, as follows. A total of 50 x 10^6 generations and a sampling frequency of 1000 were employed for both CEA and OZL datasets, while 20 x 10^6 generations and a sampling frequency of 100 were used for SCO. All models were compared against a standard analysis of the non-partitioned data following the default parameters (variable coding, λ=10 and equal rates) indicated in the MrBayes 3.2 manual for morphological characters. Burnin was set to the first 25% of all samples.

In order to evaluate the tree topologies, additional MCMC analyses were carried out for the best models (those with KRS under 2) selected in the previous phase. The number of generations applied for each matrix were: 5 x 10^6 for CEA; 5 to 15 x 10^6 for OZL; and 2 x 10^6 for SCO. Four independent runs (four chains each) and a burnin of 25% were applied on all matrices. Convergence was checked in Tracer 1.6 (Rambaut et al. 2013), AWTY (Wilgenbush et al 2004) with recommendations from Nylander et al. (2008), and in MrBayes. Chain temperatures were also empirically adjusted for each matrix in order to ensure proper mixing: *temp=0.025* for OZL matrix; and *temp=0.100* for both CEA and SCO matrices. The trees shown are majority-rule consensus trees (*Contype=Halfcompat*), with the resulting posterior probabilities as branch support.

We also evaluated the number and support of recovered clades in resulting majority-rule consensus tree obtained from each dataset under the best models. We scored the number of clades with support above 0.95 and

above 0.99. Our results were also qualitatively compared to previous works by evaluating changes in positions of major clades (see Supplementary File SM3 for results).

RESULTS

Partitioning criteria and choice of parameters and priors

Best models for each dataset are compiled in the Tables 3 and full results are detailed in the Supplementary Tables S1–S4, in which all tested models are listed together with the respective values of (ln) marginal likelihoods and Kass and Raftery's (1995) statistic (KRS). We detail results for each dataset in the sections below.

CEA

Reanalysis of CEA shows that the best overall model was obtained for homoplasy partitioning, linked branch lengths, λ=10 and no modeling of ACRV within partitions (i.e. equal rates; C-Hom-14; Table 3). The best model for the unpartitioned dataset (C-Unp-6) also employed λ=10, but ACRV was modeled via Γ distribution. This model was strongly rejected in favor of C-Hom-14 (KRS=126.36; Table 3), despite its much smaller number of free parameters. The three best models using anatomical partitioning were significantly more complex because all of them employed unlinked branch lengths. The best model (C-Ana1-34) used a shared Γ distribution to accommodate ACRV, but applying a single rate to each partition (C-Ana1-26) did not significantly worsen the marginal likelihood (KRS=0.90; Table 3). Both models correspond to Clarke and Middleton's (2008) first scheme (2 partitions). Interestingly, the third best model (C-Ana2-26, KRS=1.44) had the same settings as C-Ana1-26, but it was selected under the second scheme (3 partitions; Table 3). Even though anatomical schemes clustered together, C-Ana2 was closer to C-Ana3 than to C-Ana1 (Supplementary Fig. S1). All anatomical models were also significantly worse than C-Hom-14 (KRS > 100; Table 3). However, anatomical partitioning did generate a very strong improvement in model fit over analyses of unpartitioned data (KRS > 26 in all pairwise comparisons, calculations not shown).

Four PartitionFinder2 (PF) schemes were analyzed: C-ParFin-1, C-ParFin-2, C-ParFin-3 and C-ParFin-4, with 2, 4, 8 and 18 partitions, respectively. There were actually seven possible schemes, but similarity among C-ParFin-4, C-ParFin-5, C-ParFin-6 and CParFin-7 (18, 19, 22 and 25 partitions, respectively) exceeded 99%, hence only C-ParFin-4 was analyzed. All PF schemes were clustered in the single-linkage dendrogram, with the

exception of C-ParFin1, which was the most dissimilar of them all, most likely because it had the smallest number of partitions (Supplementary Fig. S1). C-ParFin-1 was returned by PF when using unlinked branch lengths and either AICc or BIC, regardless of starting tree choice (default NJ, ML, MP or Implied-Weights MP). C-ParFin-2 was selected for any given tree by AICc when branches were linked, whereas C-ParFin-4 was chosen by AIC, also using linked branch lengths, for all starting trees, except implied-weights MP. Overall, linking branch lengths when the selection criterion was either AIC or BIC resulted in schemes with larger numbers of partitions (Supplementary Table S6).

There were six best models under PartitionFinder (PF): three of them were recovered for scheme C-ParFin-1 and three for C-ParFin-2. The model with the best marginal likelihood (C-ParFin2-14, MgL=-1600.60) had the same parameters as C-Hom-14, but its partitioning scheme was very strongly rejected in favor of the latter's (KRS=77.18; Table 3). There was Bayesian preference for linked branch lengths and λ=10 in the case of all other top PF models, with the exception of C-ParFin2-32 (unlinked branch lengths and λ=40). Modeling ACRV using equal rates, shared or per-partition Γ distributions had negligible effect on the MgL of the six top models (Table 3). In every case, rejection of unpartitioned and anatomically partitioned models in favor of PF was also very strong (KRS > 44 and KRS > 22, respectively, calculations not shown). A graphical summary of the results for CEA is shown in Supplementary Fig. S2. There was either strong or very strong Bayesian preference for C-Hom-14 (highlighted in red) over all alternative models.

As stated above, characters were so appropriately segregated by homoplasy partitioning that rate variation in CEA was fully captured by rate multipliers. Multipliers varied more than 800-fold across partitions, from 0.1 (or a tenth of the average rate) to ~82, although the 95% highest posterior density (HPD) ranged from 0.1 to ~32. Figure 1a shows the per-partition posterior distributions of these multipliers as violin plots, arranged in growing order of unbiased homoplasy values (*f*), with the exception of partition 8, to which non-informative characters were assigned. Dots correspond to medians and whiskers to 95% HPDs. Due to the multiplicative nature of the parameter, posteriors are shown in a logarithmic scale. When computed on a per-partition basis, distributions were unimodal and symmetrical, suggesting that variation in rate multipliers results simply from the stochastic behavior of the Markov chain. Figure 1a suggests that the median of multipliers adjusted by MrBayes increased exponentially with *f* (hence the linear relationship in log-scale), while the median multiplier of the uninformative partition (8) was, as expected, considerably lower.

All best models for the anatomical schemes had unlinked branch lengths and the top model did require ACRV accommodation (shared Γ; Table 3). This model was obtained under scheme C-Anat1 (two partitions), although character segregation was poor, as indicated by the strong overlap between tree length (blue) and rate (green) posteriors (Fig. 1b). In standard phylogenetic analysis, the shape (α) and rate (β) parameters of Γ are set to the same value (Yang, 1993). As a consequence, the mean of the distribution (α/β) is always 1, leading to posteriors concentrated around this value (Fig. 1b).

PartitionFinder2's top model (C-ParFin2-14) was identical to C-Hom-14 i.e., it used linked branch lengths and equal rates. However, modeling ACRV using either shared (C-ParFin2-22) or per-partition Γ distributions (C-ParFin2-18) did not significantly worsened the marginal likelihoods (Table 3). Although best model partitions did vary with respect to rate multipliers, densities of partitions 2, 3 and 4 overlap and the latter resembles a mixture of distributions of similar masses (Fig. 1c).

OZL

In the analysis of OZL, a single best model was also selected for both unpartitioned and anatomically partitioned datasets. In both cases, λ=10 and ACRV was accommodated by Γ distributions. SS analysis preferred fixed branch lengths under the top anatomical scheme (O-Ana-10). Fixing branch lengths means that a single set of branch lengths will be fit in all partitions, which is tantamount to suppressing the effect of partitioning in the analysis. It rendered O-Ana-10 and the top unpartitoned model (O-Unp-6) identical and their MgLs virtually indistinguishable (-3770.19 vs. -3770.20; Table 3). In other words, anatomical partitioning of OZL was so inadequate that the top model for this dataset was no better than the best model using the unpartitioned matrix. Once more, the best homoplasy scheme outperformed the unpartitioned or anatomical alternatives (KRS > 192; Table 3). OZL's overall best model (O-Hom-14) had the same parameters as that for CEA (C-Hom-14). Likewise, the second overall best model (O-Hom-24) was also obtained when shared Γ and λ=40 were set, but there was positive evidence against it (KRS=6.44; Table 3; Supplementary Table S2).

When using linked branch lengths, OZL's PF schemes selected by either AIC or BIC had large numbers of partitions (23 or more). Unlinking branch lengths and/or using the corrected version of AIC (AICc) returned schemes with 1, 3 or 5 five partitions (Supplementary Table S6). Although the number of partitions in the first group ranged from 23 to 27, they formed a cluster whose normalized dissimilarity was lower than 1% (Supplementary Fig. S1). Hence only the least partitioned scheme (O-ParFin3) was analyzed together with O-

ParFin1 and O-ParFin2 (3 and 5 partitions, respectively). O-ParFin1 selection was independent of the starting tree or branch length modeling when either BIC or AICc were used. O-ParFin2 was always selected under AIC and unlinked branch lengths, regardless of the starting tree and O-ParFin3 was returned under AIC for linked branch lengths and only for the (default) neighbor-joining tree (Supplementary Table S6). Scheme O-ParFin2 yielded a single top model (O-ParFin2-18) employing linked branch lengths, shared Γ distribution and λ=10. Once again, it was very strongly rejected (KRS=81.94) in favor of the overall best model (O-Hom-14; Table 3) but also represented a very strong improvement over unpartitioned and anatomically partitioned data (KRS > 114). All SS results for OZL are summarized in Supplementary Figs. S2**.** Overall patterns were similar to CEA (Table 3).

OZL was the largest and, rate-wise, the most complex matrix analyzed in this study. There were 14 partitions under homoplasy, whose median rate multipliers again increased with *f*, although there was significant overlap in mid-range bins. The top anatomical model for OZL failed to segregate characters with respect to rate variation because there was Bayesian preference for fixed branch lengths. ACRV was accommodated by a shared Γ and it was extensive within each partition, as shown by the evident multi-modality of the rate posteriors (Fig. 1e).

The top PF model for OZL used linked branch lengths and per-partition Γ. Multiplier posteriors showed significant overlap between partitions 1 and 5 and partitions 2 and 3 and, on a multiplicative scale, partitions 3-5 distributions were slightly bimodal. Partition 5 also had long tails, although they were off the 95% HPD (in blue, Fig. 1f). Rate densities within partitions were, as expected, concentrated around 1, although the chain sporadically sampled extremely high values from the prior, leading to heavily skewed distributions (in green, Fig. 1f).

SCO

The smallest matrix (Scolebythidae) had the second largest number of competing best models. There were three when no partitions were applied to the dataset, three under anatomic partitioning, one top model under homoplasy and five under PF. The latter strategy returned two schemes, with 2 and 8 partitions (Supplementary Table S6) and both schemes yielded best models, as listed in Table 3. There was Bayesian preference for λ=10 in all models and for fixed branch lengths in all partitioned analyses; but because fixing branch lengths actually eliminates the effect of partitioning, models varied only with respect to ACRV modeling. The only exception

was homoplasy, whose best model (S-Hom-13) again employed linked branch lengths and λ=5 instead of λ=10. Excluding S-Hom-13, all other models were equivalent to S-Unp-2, if they employed fixed rates, or to S-Unp-6 if they employed shared or per-partition Γ. Marginal log-likelihoods varied from -285.42 to -284.22 (Table 3) thus KRS for the maximum possible ln(BF) was 2.20 units, which is borderline evidence against the worse-fit model. It is interesting to notice that the largest Bayes factor was observed for models which are supposed to be the same (S-Ana-6 and S-ParFin-6). This suggests that the "positive evidence" threshold of 1 log-likelihood unit proposed by Kass and Raftery (1995) may be too liberal when marginal likelihoods are computed via stepping-stone. We conclude that all models aside from S-Hom-13 were roughly equivalent, in spite of differences in ACRV modeling, but once more strongly rejected in favor of homoplasy partitioning (KRS > 23, Table 3).

Per-partition posterior distributions of rate multipliers obtained from S-Hom-13 (Fig. 1g) resembled the ones obtained for CEA (Fig. 1a) although median dispersion was smaller. A single character (25) was assigned to the uninformative partition, hence the large variance of the corresponding distribution. The top PF model (S-ParFin2-6) employed fixed branch lengths and per-partition Γ. Rate posteriors were mostly bimodal with long tails (Fig. 1h), although not as extreme as found in OZL (Fig. 1e). Because the top model under anatomical partitioning used fixed branch lengths and equal rates, there were no comparable results to be shown.

Tree topologies and clade support

A summary of the trees resulting from MCMC analyses implementing the default model (unpartitioned data) and the best models, by partitioning criteria, is shown in Supplementary File SM3. In general, no striking differences in number of recovered clades are observed among the results from all datasets. The homoplasy partitioning schemes, however, tend to increment the number of clades that are recovered in 50% or more trees in the posterior distributions (and hence not collapsed in the majority rule consensus tree, improving its resolution), in particular for OZL and SCO datasets. For CEA, the homoplasy schemes increase the number of clades with 0.99 posterior probability when applied in both unordered and ordered characters (see Supplementary File SM2).

DISCUSSION

Homoplasy partitioning of discrete morphological characters

Data partitioning in conjunction with the choice of adequate priors are decisive in Bayesian modeling of morphological character evolution. Here we showed that implementing homoplasy partitioning in all analyzed datasets had a large effect on MgL when compared to anatomically partitioned or unpartitioned datasets (Table 3 and Supplementary File SM2). Additionally, we also found improved resolution of consensus both in OZL and SCO datasets (Supplementary File SM3). Our results agree with previous empirical studies that have also demonstrated higher performance of partitioned datasets (Clarke and Middleton 2008; Tarasov and Génier 2015; Wright 2015; Gavryushkina et al. 2016; King et al. 2017). Appropriate sorting of characters into partition sets by the homoplasy approach is evidenced by the Bayesian preference for linked branch lengths and equal rates in analyses of all datasets (Tables 3 and 4). The choice of no ACRV means that evolution across characters within homoplasy partitions was so homogeneous that fitting a single rate to each one was sufficient. On the other hand, there was considerable rate heterogeneity among partitions, as evidenced in Figure 1. In homoplasy partitioned matrices, APRV was efficiently accommodated by linking branch lengths, which adds a modest number of parameters to the model as the number of partitions increases. This result contrasts with what was obtained for Nylander et al. (2004) in analysis of combined morphological and molecular datasets. Although the authors did not employ any likelihood penalization method, they found an increase in 600-900+ log-likelihood units when comparing partitioned to unpartitioned datasets, whereas allowing ACRV increased log-likelihood in roughly 3,000 units. Those authors partitioned the molecular data by gene, but not by codon, and did not attempt to partition the morphological matrix, so ACRV was probably necessary to accommodate the rate variation not captured by their schemes. In our study, there was positive to very strong Bayesian support for no ACRV modeling in homoplasy partitioned data (Table 3).

Homoplasy based criterion of data partitioning has been previously applied to molecular datasets by Kjer and Honeycutt (2007). They used similar reasoning i.e., partitioning of sites according to their levels of homoplasy. They calculated Farris' consistency index (CI) for each site based on a set of trees generated from 1,000 bootstrap runs and sorted the returned CI's into six discrete rate classes, each corresponding to a partition. Considering the penalization of multistate characters imposed by the CI (see the section *Partitioning* in

Methods and Materials), the method proposed here, based on Goloboff's unbiased measure of homoplasy, represents an improvement over that of Kjer and Honeycutt. A thorough evaluation of the application of this approach to molecular datasets is beyond the scope of this study, but should be explored in future work, including both nucleotide and aminoacid datasets.

The adequacy of homoplasy levels as proxy for rates of evolution is supported by the evident ordinal correlation between the per-partition median rate multipliers and Goloboff's unbiased homoplasy scores (Figs. 1a, 1d and 1g). If the posterior conforms to a Gaussian (normal) distribution, its mean (μ) is an unbiased estimator of any given parameter's expectation. Variation around the mean arises from the stochastic nature of the chain after reaching stationarity, hence expressing phylogenetic uncertainty or noise. A suitable measure for noise is the standard deviation (σ) of the posterior and the signal-to-noise ratio (SNR) of the Markov process is given by μ/σ (i.e the reciprocal of the coefficient of variation). In signal processing analysis, SNR is usually converted to a log-ratio scale by the formula $10 \cdot \log_{10}(\mu/\sigma)$. When taken in this form, SNR is expressed in decibels (dB). In this scale, SNR > 0 means more signal than noise in the data and the opposite if SNR < 0 (Lockhart et al. 1985). In the context of phylogenetic analysis, SNR thus calculated could be viewed as a measure of the phylogenetic information contained in the data partition. For instance, if the expected rate multiplier of a partition is 0.1 (i.e. μ =0.1) and σ=0.2, noise is twice as high as signal. This means that the chain sampled much more freely from the prior than it would have if μ =1.0, in which case the standard deviation would only be 20% of the mean. In the first case, SNR=-3.01 dB and in the second SNR=6.98 dB. The difference in SNR between the first and second cases is a function of how much the data in the partition updated the prior; thus, SNR is effectively quantifying phylogenetic signal over noise. The number of characters also tends to decrease with the *f* score (see Supplementary Tables S11, S21 and S30) thus exacerbating the drop in the partition's phylogenetic content.

Figures 1a, 1d and 1g show that there is indeed increase in within-partition rate multipliers variances with homoplasy scores. Because more homoplastic characters are less informative, the chain samples more freely from the prior, generating posteriors with wider HPDs. Therefore, their higher medians could be a mere byproduct of decreasing phylogenetic signal. To evaluate the normality of the multiplier posteriors we visually inspected corresponding probability-probability (p-p) plots (Supplementary Figs. S3a-c). Most distributions were very close to normality, being skewed to the left in some cases (blue curve is below the confidence

envelope), with the noticeable exception of OZL's partition 11, which is highly skewed (Supplementary Figs. S3b). We computed SNR for unordered CEA, SCO and OZL and graphed it against *f* scores arranged in increasing order (Fig. 2). Results showed an overall trend of decreasing phylogenetic signal with increasing homoplasy, but noise never overcame signal except for partition 11 in OZL (Fig. 2b) and the non-informative partition (labeled "#" in Fig. 2c) in SCO.

OZL's partition 11 is made up by a single character, the relative size between the alular ("first wing") digit and the metatarsal (numbered 168 in O'Connor & Zhou, 2012). Its corresponding posterior is highly skewed to the left, as it can be clearly seen in Figure 1d. Because it deviates strongly from normality (probably due to lack of adequate convergence), SNR is not accurately estimated for this partition because its mean is lower than expected, hence its departure from the overall pattern. It is interesting to notice that partition 13 also includes a single limb-related character, the distal extent of metatarsal II, relative to metatarsal IV of the pelvic limb (236 in O'Connor & Zhou, 2012). However, its posterior met normality (Supplementary Figs. S3b) and its SNR exceeded 2 dB, although its *f* score is greater than that of partition 11 (0.85 vs. 0.79, Fig. 2b). Thus, low SNR does not seem to be attributable only to partition size. The same pattern was observed for SCO. Character 28 (shape of margin of pterostigma inside the marginal cell) was assigned to partition 5, whose *f* score was 0.57 while character 25 (shape of prosternum) is parsimony non-informative (partition 6, see Supplementary File SM1 for the full character list). Again, the multiplier posterior of the former conforms well to a normal distribution and has the second highest SNR, whereas the latter has negligible phylogenetic signal, although this estimate may not be very accurate for, as character 168 in OZL, it also departs from normality (Fig. 2c and Supplementary Figs. S3c).

This also has implications in deciding when partitions exhibiting similar behavior could be fused into a single partition. TNT's implementation of implied weighting analysis groups all characters with a single state change (singletons or autopomorphies, in parsimony context) into the same category. The corresponding posterior distributions of rate multipliers (the last ones in Figs. 1a,1d and 1g) have medians which are closer to the ones obtained for the least homoplastic partitions (the first ones in each graph) than to any other partition in the same dataset. Fusing the last and first partitions together resulted in even better MgLs than the best models, although evidence against them was not positive in every case (CAE KRS=1.88, OZL KRS=4.40 and SCO

KRS=2.82). At worst, these results suggest that these two partitions may be reduced to one with no significant worsening in model likelihoods.

We conclude that, although rising APRV is in part a byproduct of the reduction in phylogenetic information, there is some signal left even in the most homoplastic or parsimony uninformative partitions. In every case where the posterior significantly departed from normality, it was skewed to the left. In these cases, μ is lower than expected, so corresponding SNRs were likely underestimates, thus corroborating the prevalence of signal. It is been long known in molecular phylogenetic analysis (e.g. Philippe et al. 1994) that saturated characters, such as the third codon positions, evolve at higher rates. Thus, rate of evolution and phylogenetic signal are inherently intertwined and this was effectively captured by homoplasy partitioning, as evidenced in Figure 2. In the following sections, we discuss the performance of this strategy against its competing alternatives.

Anatomical partitioning

The best models under anatomical partitioning of CEA employed unlinked branch lengths and resulted in considerable improvement over unpartitioned data (see Results and Table 3). However, the posterior distributions of the top model's (C-Ana-34) tree lengths and rates showed considerable overlap (Fig. 1b and Supplementary File SM4). All anatomical schemes were strongly rejected in favor of homoplasy (Table 3). Results were so insensitive to the choice of anatomical partitions proposed by Clarke & Middleton (2008) that despite a normalized dissimilarity of over 10% between C-Ana1 and C-Ana2 (Supplementary Fig. S1), the three best models were based on either scheme (Table 3). In the two other datasets (OZL and SCO), tentative anatomical partitioning was even more inefficient because there was Bayesian preference for fixed branch lengths in all cases, thus results were no better than those obtained from the analysis of unpartitioned data (Table 3).

Besides model performance, we must also take into consideration the differences in implementation when comparing anatomical and homoplasy criteria. Use of an anatomical criterion, as implemented by Clarke and Middleton (2008), presupposes *a priori* functional units or anatomical modules, which have been hypothesized to evolve independently from one another. In the words of Clarke and Middleton (2008), these modules are expected to show "disjunct patterns of evolution". A similar reasoning was given by Tarasov and Génier

(2015), who assumed "that characters of the same anatomical region undergo similar evolutionary dynamics". Mapping of homoplasy scores onto the longitudinally arranged characters (Fig. 3) suggests that, both in Aves and Scolebythidae, patterning rates of evolution is not necessarily correlated with anatomy. One could still argue that the major point in such partitioning strategy is that "disjunct evolution" means that branch lengths ought to be independently estimated across modules. Such scheme would still be superior to the unpartitioned analyses, thus simply mapping homoplasy to individual characters is not a fair test. Table 3 shows this to be true in the case of CEA for which there was an improvement of 12-13 log-likelihood units with respect to the unpartitioned analyses when unlinked branch lengths were employed. However this was not the case of OZL and SCO: top models for these datasets employed fixed branch lengths, which is equivalent to no data partitioning. If on the one hand OZL and SCO matrices were not explicitly partitioned according to functional modules, on the other, Clarke and Middleton's (CEA) schemes vary substantially among themselves and yet resulted in remarkably similar marginal likelihoods, as noted in the beginning of this section. Therefore, an additional problem confronting someone trying to implement this criterion derives from the *ad hoc* hypotheses invoked to support the delimitation of the functional modules. We expect that different groups of organisms will require their own specific hypotheses and that different researchers will eventually reach non-coincident partitions due to the profusion of functional hypotheses available. One could always resort to developmental data or attempt a more objective approach to module identification (e.g. Magwene, 2001). At any rate, the implementation of anatomical data partitioning requires more complex premises than the homoplasy criterion.

Homoplasy vs. PartitionFinder

Unlike the other strategies, PartitionFinder 2.1.1 returned several schemes for each analyzed dataset. Scheme selection was highly sensitive to modeling of rate variation among partitions and the criterion used to penalize overparameterization (AIC, AICc and BIC), but much more invariant with respect to the guide tree. Overall, we found a uniform pattern across all datasets: linking branch lengths reduced the number of free parameters, but led to oversplitting under AIC and especially under BIC, whereas starting trees had little influence on scheme choice (Supplementary Table S6). This is rather surprising given that BIC is supposed to be the most conservative criterion in model selection (Brewer et al. 2016). Our results also differed from tests with an earlier version of PartitionFinder2, in which Wright (2015) analyzed 333 datasets selecting linked

branch lengths and found the average number of partitions supported by AICc to be 10, 15 for BIC and 43 for BIC. She does not inform in her thesis the minimum subset size that was allowed, so her results are perhaps not comparable to ours.

The number of partitions in C-ParFin4, 5, 6 and 7, O-ParFin3, 4, 5 and 6 and S-ParFin2 were at least 60% greater than the corresponding homoplasy scheme, or 8 partitions for the small Scolebythidae matrix (28 characters) and over 18 for the two Aves datasets (Supplementary Fig. S1). These partitions invariably formed clusters of high similarity, indicating differential allocation of just a few characters among them. The only exception was O-ParFin6 that was very different from other schemes (Supplementary Fig. S1b). Oversplitting could probably be reduced by increasing the value of the *--min-subset-size* option, which was set to 1. Indeed, all highly partitioned schemes had partitions with a single character. There are no clear guidelines on how to select the minimum subset size, but increasing it may not necessarily lead to better partitioning schemes: all homoplasy schemes employed in this study also had at least one single-character partition and top models under this strategy strongly outperformed the best PF models (Table 3). Top PF models were based on schemes that had 5 partitions or less in the case of CEA and OZL, whereas homoplasy partitioning returned 8 and 14 partitions, respectively. Differences between homoplasy and PF schemes go beyond the number of partitions. For instance C-ParFin3 and the scheme C-Hom (from which the best overall model resulted) both had 8 partitions but were very dissimilar (Supplementary Fig. S1a), demonstrating that these strategies are fundamentally different.

In the case of both CEA and OZL, PF partitioning very strongly improved model likelihood if compared to either unpartitioned or anatomically partitioned datasets (see Results). Our results contrast with those obtained using the preliminary version of PF2 by Wright (2015). When applied to Clarke and Middleton's (2008) dataset (CEA) with ordered characters, PF2 suggested various partition schemes, with the number of subsets varying from 2 to 55. Wright's comparisons between the partition schemes returned by PF2 with those of Clarke and Middleton showed improved MgL associated with the latter models. Wright (2015) then concluded that the scheme chosen by Clarke and Middleton was superior to those of PF2 and that all available biological information should be taken into account for partitioning of morphological datasets. The author used variable coding, unlinked and linked branch lengths, ACRV approximated by per-partition Γ distributions and λ=5 in her analysis of Clarke and Middleton's data. Our MgL values obtained under these same set of

parameters (C-Hom(Ord)-41 and C-Hom(Ord)-29) were compared to those provided by Wright (2015) (see Table 4). Independently of the MgL estimation method, Clarke and Middleton's anatomical schemes still have improved MgL values when compared to those of the preliminary version of PF2. Perhaps an increase in the MgL associated with her PF2 partition schemes could result from a broader exploration of the parameter space. For example, our results show that for CAE better MgL values are associated with λ=10–20, instead of λ=5.

Unlike the Aves datasets, PartitionFinder2 did not improve model fit to the smallest matrix (SCO, 28 characters) over the analysis of unpartitioned data and neither did the anatomical scheme (Table 3). Interestingly, this was the only dataset in which the number of PF partitions exceeded homoplasy. These results suggest that model fit may be particularly sensitive to partitioning in smaller datasets.

Choice of Parameters

According to our results, effective partitioning strategies should result in very strong Bayesian preference for linked branch lengths and equal rates. This was the case in all top models, always recovered for homoplasy partitioned datasets. Our results also show that MrBayes default's value for the inverse scale parameter of the exponential prior distribution (λ=10) resulted in best models in most analyses, regardless of the partitioning strategy (Table 3). An important exception was the top model for Scolebythidae (S-Hom-13) that showed Bayesian preference for λ=5. The exponential distribution mean is the reciprocal of λ, so when λ=10, the mean is 0.1 and twice as large (0.2) when λ=5 (Ronquist et al. 2009). which also makes the prior flatter and less informative. Thus, there may be two reasons that explain the preference for λ=5 in the case of the smallest data matrix: either the data are informative and pull the chain towards longer branches or the data are simply not informative, and there is Bayesian preference for a vaguer prior.

To distinguish among these two hypotheses, we sampled the prior distribution of tree lengths using MrBayes' option *mcmc data=no*, and compared to the posteriors obtained using λ=5 and λ=10. Results show that the data are actually informative (Fig. 4). Most of the posterior mass for λ=10 (cyan) is shifted to the right with respect to the prior, indicating positive preference for longer branches (Table 3). This is reinforced by the much better fit between prior and posterior when λ=5, which explains the improved marginal likelihood.

We suspected that the longer branches in this and all other analyses employing the same strategy (see Results and Supplementary File SM4) may be due to the segregation of the most homoplastic characters in its

singleton partitions (see Supplementary Tables S11, S21 and S30). The influence of these higher multipliers on branch lengths was likely exacerbated in the small Scolebythidae matrix and hence the difference in λ choice.

We ran analyses with λ=5 and λ=10, after excluding partition 5 (which has a single character, but is informative according to Fg. 2c) and fusing partitions 1 and 6 (respectively, the homoplasy-free and parsimony-uninformative partitions; see first section of Discussion), and comparison of resulting marginal likelihoods indicated positive Bayesian preference for λ=10 (KRS=2.30), corroborating our initial hypothesis. This suggests that smaller values of λ may provide a better model fit in smaller matrices with high homoplastic characters. There is, however, no reason why any particular rate prior should be appropriate for all datasets. The preference for λ=10 may be peculiar to the tested matrices or simply because it makes the prior vaguer and hence more conservative. Evaluating alternative branch length priors is beyond the scope of this study, but one should also consider this possibility when applying homoplasy partitioning to his/her dataset.

Conclusions

Use of homoplasy as a criterion for data partitioning is an efficient approach to explore discrete morphological characters under Bayesian inference. The models implementing homoplasy partitioning schemes showed significantly better fit for the three datasets analyzed than all other alternatives. They also resulted in increased resolution of the tree topologies, in particular for OZL, the largest and most complex matrix analyzed in this study. In the case of the smallest matrix (SCO) homoplasy was indeed the only strategy that effectively partitioned the data, resulting in superior model likelihood with improved tree resolution and branch support. This criterion is easily implemented, requiring no *a priori* premises, and can be applied to datasets irrespective of their numbers of characters. It is also more straightforward than an algorithmic alternative (PartitionFinder2) that has a much larger parameter space that must be surveyed by the user and whose schemes were outperformed in every comparison with homoplasy partitioning. Morphological datasets frequently contain less than 100 characters and our results show that a full analysis of a matrix twice as large (CEA) may be accomplished in a matter of hours. Thus, we believe homoplasy partitioning to be a promising approach in model-based phylogenetic analyses.

ACKNOWLEDGEMENTS

We would like to thank Eduardo Almeida, Rodrigo Feitosa, Peter Foster, Luke Harmon, Mike Lee, April Wright and an anonymous reviewer for comments on an earlier draft of the manuscript. Financial support has been provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico (grants 309641/2016-0; 140263/2015-2).

LITERATURE CITED

Brandley M.C., Schmitz A., Reeder T.W. 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. Syst. Biol. 54:373–390.

Brewer M.J., Butler A., Cooksley S.L. 2016. The relative performance of AIC, AICC and BIC in the presence of unobserved heterogeneity. Methods Ecol. Evol. 7:679–692.

Brown J.M., Lemmon A.R. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst. Biol. 56:643–655.

Carpenter J.M. 1999. What do we know about chrysidoid (Hymenoptera) relationships? Zool. Scr. 28:215–231.

Clarke J.A., Middleton K.M. 2008. Mosaicism, modules, and the evolution of birds: results from a Bayesian approach to the study of morphological evolution using discrete character data. Syst. Biol. 57:185–201.

Clarke J.A., Zhou Z., Zhang F. 2006. Insight into the evolution of avian flight from a new clade of Early Cretaceous ornithurines from China and the morphology of *Yixianornis grabaui*. J. Anat. 208:287–308.

Dávalos L.M., Velazco P.M., Warsi O.M., Smits P.D., Simmons N.B. 2014. Integrating incomplete fossils by isolating conflicting signal in saturated and non-independent morphological characters. Syst. Biol. 63:582–600.

Engel M.S, Grimaldi D.. 2007. Cretaceous Scolebythidae (Hymenoptera) and phylogeny of the family. Am. Mus. Novit. 3568.

Engel M.S., Ortega-Blanco J., McKellar R.C. 2013. New scolebythid wasps in Cretaceous amber from Spain and Canada, with implications for the phylogeny of the family (Hymenoptera: Scolebythidae). Cretac. Res. 46:31–42.

Farris J.S. 1969. A successive approximations approach to character weighting. Syst Zool. 18:374–385.

Gavryushkina A., Heath T.A., Ksepka D.T., Stadler T., Welch D., Drummond A.J. 2017. Bayesian total-evidence dating reveals the recent crown radiation of penguins. Syst. Biol. 66:57–73.

Giribet G. 2015. Morphology should not be forgotten in the era of genomics - a phylogenetic perspective. Zool. Anz. 256:96–103.

Goloboff P. 1993. Estimating character weights during tree search. Cladistics. 9:83–91.

Goloboff P., Farris S., Nixon K. 2008. TNT, a free program for phylogenetic analysis. Cladistics. 24:774–786.

Harrison L.B., Larsson H.C.E. 2015. Among-character rate variation distributions in phylogenetic analysis of discrete morphological characters. Syst. Biol. 64:307–324.

Jukes T.H., Cantor C.R. 1969. Evolution of protein molecules. In: Munro H.N., editor. Mammalian Protein Metabolism. New York: Academic Press, p. 21–132.

Kainer D., Lanfear R. 2015. The effects of partitioning on phylogenetic inference. Mol. Biol. Evol. 32:1611–1627.

Kass R.E., Raftery A.E. 1995. Bayes factors. J. Am. Stat. Assoc. 18:773–795.

King B., Qiao T., Lee M.S.Y., Zhu M., Long J.A. 2017. Bayesian morphological clock methods resurrect placoderm monophyly and reveal rapid early evolution in jawed vertebrates. Syst. Biol. 66:499–516.

Kjer K.M., Honeycutt R.L. 2007. Site specific rates of mitochondrial genomes and the phylogeny of Eutheria. BMC Evol. Biol. 7:8 doi:10.1186/1471-2148-7-8.

Klopfstein S., Vilhelmsen L., Ronquist F. 2015. A nonstationary Markov model detects directional evolution in hymenopteran morphology. Syst. Biol. 64:1089–1103.

Lanfear R., Frandsen P.B., Wright A.M., Senfeld T., Calcott B. 2017. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34:772–773.

Lee M.S.Y. 2016. Multiple morphological clocks and total-evidence tip-dating in mammals. Biol. Lett. 12:20160033.

Lee M.S.Y., Palci A. 2015. Morphological phylogenetics in the genomic age. Curr. Biol. 25:R922–R929.

Lee M.S.Y., Soubrier J., Edgecombe G.D. 2013. Rates of phenotypic and genomic evolution during the Cambrian explosion. Curr. Biol. 23:1889–1895.

Lee M.S.Y., Cau A., Naish D., Dyke G.J. 2014. Morphological clocks in paleontology, and a mid-Cretaceous origin of crown Aves. Syst. Biol. 63:442–449.

Lemmon A., Moriarty E. 2004. The importance of proper model assumption in Bayesian phylogenetics. Syst. Biol. 53:265–277.

Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913–925.

Lockhart G.B., Cheetham B.M.G. 1985. Basic Digital Signal Processing: Butterworths Basic Series. London: Butterworth-Heinemann.

Magwene, P. L. 2001. New Tools for Studying Integration and Modularity. Evolution 55 (9): 1734–1745.

Marshall D., Simon C., Buckley T. 2006. Accurate branch length estimation in partitioned Bayesian analyses requires accommodation of among-partition rate variation and attention to branch length priors. Syst. Biol. 55:993–1003.

Miller M.A., Pfeiffer W., Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gatew. Comput. Environ. Work. GCE 2010.

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Yiyuan, Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B.M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong, T.K.F., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Yunhui, Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Yingrui, Xu, X., Zhang, Yong, Yang, H., Wang, Jian, Wang, Jun, Kjer, K.M., Zhou, X. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346:763–767.

Nylander, J.A.A., Ronquist F., Huelsenbeck J.H., Nieves-Aldrey J.L. 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53:47–67.

Nylander J.A.A., Wilgenbusch J.C., Warren D.L., Swofford D.L. 2008. AWTY (are we there yet?): A system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics. 24:581–583.

O'Connor J.K., Zhou Z. 2012. A redescription of *Chaoyangia beishanensis* (Aves) and a comprehensive phylogeny of Mesozoic birds. J. Syst. Palaeontol. 2019:1–18.

Ohio Supercomputer Center. 1987. Columbus OH: Ohio Supercomputer Center. Available from: http://osc.edu/ark:/19495/f5s1ph73/

Philippe H., Sörhannus U., Baroin A., Perasso R., Gasse F., Adoutte A. 1994. Comparison of molecular and paleontological data in diatoms suggests a major gap in the fossil record. J. Evolution Biol. 7:247–65.

Pyron R.A. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. Syst. Biol. 60:466–481.

Quental T.B., Marshall C.R. 2010. Diversity dynamics: Molecular phylogenies need the fossil record. Trends Ecol. Evol. 25:435–441.

Rambaut A., Suchard M., Drummond A. 2013. Tracer 1.6. Available from: http://tree.bio.ed.ac.uk/software/tracer/

Ronquist F., Huelsenbeck J., Teslenko M. 2011. MrBayes Version 3.2 Manual: Tutorials and Model Summaries. 1–103.

Ronquist F., Klopfstein S., Vilhelmsen L., Schulmeister S., Murray D.L., Rasnitsyn A.P. 2012a. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Syst. Biol. 61:973–999.

Ronquist, F., Van Der Mark P., Huelsenbeck J. 2009. Bayesian Phylogenetic Analysis Using MrBayes. In: Lemey P., Salemi M., Vandamme A.M. editors. The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. New York: Cambridge University Press, 210–66.

Ronquist F., Teslenko M., Van Der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. 2012b. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.

Tarasov S., Génier F. 2015. Innovative Bayesian and parsimony phylogeny of dung beetles (Coleoptera, Scarabaeidae, Scarabaeinae) enhanced by ontology-based partitioning of morphological characters. PLoS One 10(3): e0116671.

Wagner P.J. 2012. Modelling rate distributions using character compatibility: implications for morphological evolution among fossil invertebrates. Biol. Lett. 8:143–146.

Wiens J.J., Kuczynski C.A., Townsend T., Reeder T.W., Mulcahy D.G., Sites J.W. 2010. Combining phylogenomics and fossils in higher-level squamate reptile phylogeny: Molecular data change the placement of fossil taxa. Syst. Biol. 59:674–688.

Wilgenbusch J.C., Warren D.L., Swofford D.L. 2004. AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. Available from: http://ceb.csit.fsu.edu/awty

Wright A.M. 2015. Estimating phylogenetic trees from discrete morphological data. Ph.D. Dissertation. University of Texas at Austin.

Wright A.M., Lloyd G.T., Hillis D.M. 2015. Modeling character change heterogeneity in phylogenetic analyses of morphology through the use of priors. Syst. Biol. 65:602–611.

Xie W., Lewis P.O., Fan Y., Kuo L., Chen M.H. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Syst. Biol. 60:150–160.

Yang, Z. 1993. Maximum-Likelihood Estimation of Phylogeny from DNA Sequences When Substitution Rates Differ over Sites. Mol. Biol. Evol. 10:1396–1401.

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 each dataset, indicated on the right end of the row (CEA, OZL, SCO, 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-14 b) C-Ana1-34 c) C-Parfin2-14 d) O-Hom-14 e) O-Ana-10 f) O-ParFin-18 g) S-Hom-13 h) O-ParFin8-18. i) C-Hom(Ord)-2 j) C-Ana(Ord)-22 k) C-ParFin1(Ord)-10.

Figure 2. Signal to noise ratio (SNR) plots correlating the amount of phylogenetic information in each partition to the corresponding homoplasy scores (*f*). Scores are in ordinal and not in ratio scale because no numeric values are assigned by TNT to the parsimony non-informative partition (labeled #). Negative SNRs (below the red line) denote more noise than signal in the partition. a) CEA b) OZL c) SCO.

Figure 3. Homoplasy scores obtained from implied weighting parsimony mapped onto the corresponding anatomic partitions for the three analyzed data matrices. See text and Supplementary Tables S4–S14 for details.

Figure 4. Prior and posterior tree length densities sampled via MCMC using the best (S-Hom-13, λ=5) and second best (S-Hom-14, λ=10) phylogeteic models, applied to the homoplasy partitioned Scolebythidae (SCO) dataset. The analysis employing S-Hom-14 shows that the prior distribution (cyan) was strongly updated by the data, for most of the posterior's density (pink) is shifted towards the prior's right tail. Setting the inverse scale parameter to 5 doubles the prior's average tree length (purple distribution), greatly improving its fit to the posterior (in green).

LEGEND TO TABLES

Table 1. Summary of the analyzed datasets and corresponding references.

Table 2. Criteria, number of partitions, acronyms and references for each scheme.

Table 3. Best models by criterion for all datasets analyzed with variable coding and unordered characters. Best model overall under each scheme in bold face.

Table 4. Comparisons between stepping-stone marginal likelihoods of models by partitioning criteria for Clarke and Middleton's (2008) dataset (here referred as Clarke's et al. dataset).

Table 1. Summary of the analyzed datasets and corresponding references.

Taxa	Matrix acronym	No. terminals	No. characters	References
Aves	CEA	25 (5 extant; 20 extinct)	205	Clarke et al. (2006), Clarke and Middleton (2008)
Aves	OZL	65 (4 extant; 61 extinct)	247	O'Connor and Zhou (2012), Lee et al. (2014)
Hexapoda, Scolebythidae	SCO	18 (6 extant; 11 extinct)	28	Carpenter (1999), Engel and Grimaldi (2007), Engel et al. (2013), this study [a]

[a] See Supplementary File SM1 for characters and matrix.

Table 2. Criteria, number of partitions, acronyms and references for each scheme.

Matrix	Criterion	No. partitions	Acronym	Reference
CEA	Anatomy	2	C-Ana1	Clarke and Middleton (2008)
	Anatomy	3	C-Ana2	Clarke and Middleton (2008)
	Anatomy	4	C-Ana3	Clarke and Middleton (2008)
	Homoplasy	8	C-Hom	This study
	Partition Finder	2	C-ParFin1	This study
	Partition Finder	4	C-ParFin2	This study
	Partition Finder	8	C-ParFin3	This study
	Partition Finder	18	C-ParFin4	This study
	Partition Finder	19	C-ParFin5	This study
	Partition Finder	22	C-ParFin6	This study
	Partition Finder	25	C-ParFin7	This study
OZL	Anatomy	7	O-Ana	This study[a]
	Homoplasy	14	O-Hom	This study
	Partition Finder	3	O-ParFin1	This study
	Partition Finder	5	O-ParFin2	This study
	Partition Finder	23	O-ParFin3	This study
	Partition Finder	25	O-ParFin4	This study
	Partition Finder	26	O-ParFin5	This study
	Partition Finder	27	O-ParFin6	This study
SCO	Anatomy	3	S-Ana	This study
	Homoplasy	6	S-Hom	This study
	Partition Finder	2	S-ParFin1	This study
	Partition Finder	8	S-ParFin2	This study
	Partition Finder	1	S-ParFin3	This study

[a] Based on anatomical regions in O'Connor and Zhou (2012) (see text for details).

Table 3. Best models by criterion for all datasets analyzed with variable coding and unordered characters. Best model overall under each scheme in bold face.

CEA

Model	Partitions	APRV[b]	ACRV	λ	MgL	KRS[c]	KRS[d]
C-Unp-6	1	N/A	Shared Γ	10	-1625.19	126.36	0.00
C-Unp-7	1	N/A	Shared Γ	20	-1625.46	126.9	0.54
C-Unp-2[a]	1	N/A	Equal rates	10	-1626.05	128.08	1.72
C-Ana1-34	2	Unlinked branch lengths	Shared Γ	10	-1612.56	101.1	0.00
C-Ana1-26	2	Unlinked branch lengths	Equal rates	10	-1613.01	102	0.90
C-Ana2-26	3	Unlinked branch lengths	Equal rates	10	-1613.28	102.54	1.44
C-Hom-14	**8**	**Linked branch lengths**	**Equal rates**	**10**	**-1562.01**	**0.00**	**0.00**
C-ParFin2-14	4	Linked branch lengths	Equal rates	10	-1600.60	77.18	0.00
C-ParFin2-32	4	Unlinked branch lengths	Per-partition Γ	40	-1600.77	77.52	0.34
C-ParFin1-14	2	Linked branch lengths	Equal rates	10	-1601.08	78.14	0.96
C-ParFin1-22	2	Linked branch lengths	Shared Γ	10	-1601.19	78.36	1.18
C-ParFin1-18	2	Linked branch lengths	Per-partition Γ	10	-1601.29	78.56	1.38
C-ParFin2-22	4	Linked branch lengths	Shared Γ	10	-1601.55	79.08	1.90

OZL

Model	Partitions	APRV[b]	ACRV	λ	MgL	KRS[c]	KRS[d]
O-Unp-6	1	N/A	Shared Γ	10	-3770.20	192.92	0.00
O-Unp-2[a]	1	N/A	Equal rates	10	-3812.92	280.36	85.44
O-Ana-10	7	Fixed branch lengths	Shared Γ	10	-3770.19	194.90	0.00
O-Hom-14	**14**	**Linked branch lengths**	**Equal rates**	**10**	**-3672.74**	**0.00**	**0.00**
O-ParFin2-18	5	Linked branch lengths	Per-partition Γ	10	-3713.71	81.94	0.00

SCO

Model	Partitions	APRV[b]	ACRV	λ	MgL	KRS[c]	KRS[d]
S-Unp-2[a]	1	N/A	Equal rates	10	-284.44	23.58	0.00
S-Unp-6	1	N/A	Shared Γ	10	-284.97	24.64	1.06
S-Ana-2	3	Fixed branch lengths	Equal rates	10	-284.50	23.70	0.00
S-Ana-10	3	Fixed branch lengths	Shared Γ	10	-285.00	24.70	1.00
S-Ana-6	3	Fixed branch lengths	Per-partition Γ	10	-285.42	25.54	1.84
S-Hom-13	**6**	**Linked branch lengths**	**Equal rates**	**5**	**-271.65**	**0.00**	**0.00**
S-ParFin2-6	8	Fixed branch lengths	Per-partition Γ	10	-284.22	23.14	0.00
S-ParFin1-10	2	Fixed branch lengths	Shared Γ	10	-284.35	23.40	0.08
S-ParFin2-10	8	Fixed branch lengths	Shared Γ	10	-284.39	23.48	0.10
S-ParFin1-6	2	Fixed branch lengths	Per-partition Γ	10	-284.40	23.50	0.24
S-ParFin1-2	2	Fixed branch lengths	Equal rates	10	-284.47	23.64	0.24

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] "Fixed branch lengths": likelihood estimation was conditional on a single topology and set of branch lengths across partitions; "Linked branch lengths": branch lengths were allowed to vary across partitions, but remained linked by rate multipliers; "Unlinked branch lengths": branch lengths corresponding to each partition were independently estimated.

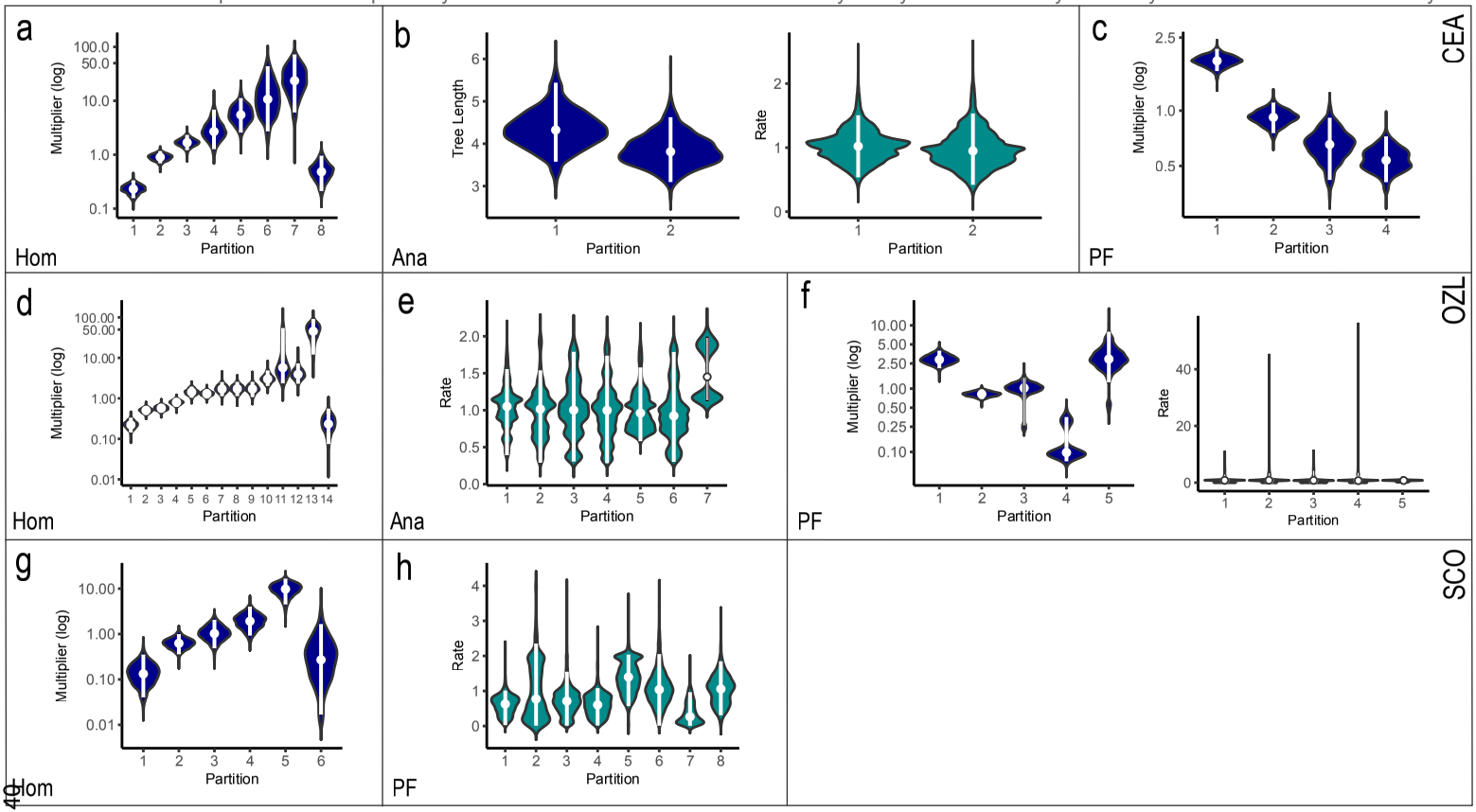
[c] Kass & Raftery's statistic computed with respect to the best model overall.

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

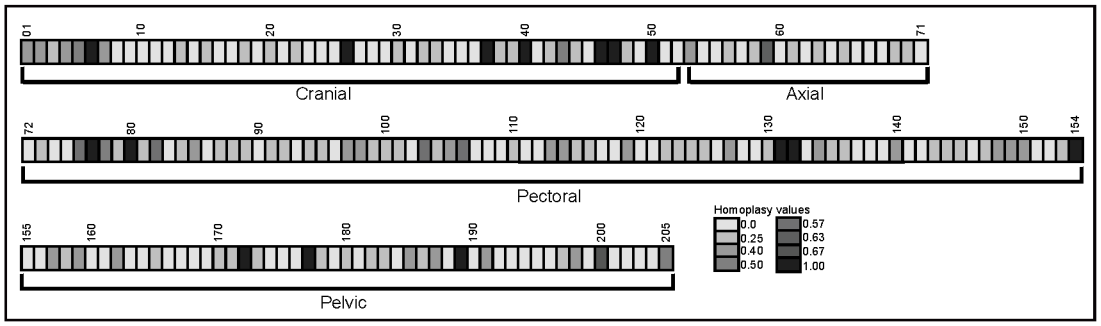
Table 4. Comparisons between stepping-stone marginal likelihoods of models by partitioning criteria for Clarke and Middleton's (2008) dataset (here referred as Clarke's et al. dataset).

Character ordering	Partitioning criteria	MgL by SS[a]
Unordered	Unpartitioned	-1639.44 to -1637.94
Unordered	Anatomy	-1642.41 to -1628.27
Unordered	Homoplasy	-1633.15 to **-1574.67**
Ordered	Unpartitioned	-1615.13 to -1608.08
Ordered	Anatomy	-1610.93 to -1612.16
Ordered	Homoplasy	-1602.92 to **-1548.56**
Ordered	PartitionFinder2	-1734.4 to -1623.5[b]

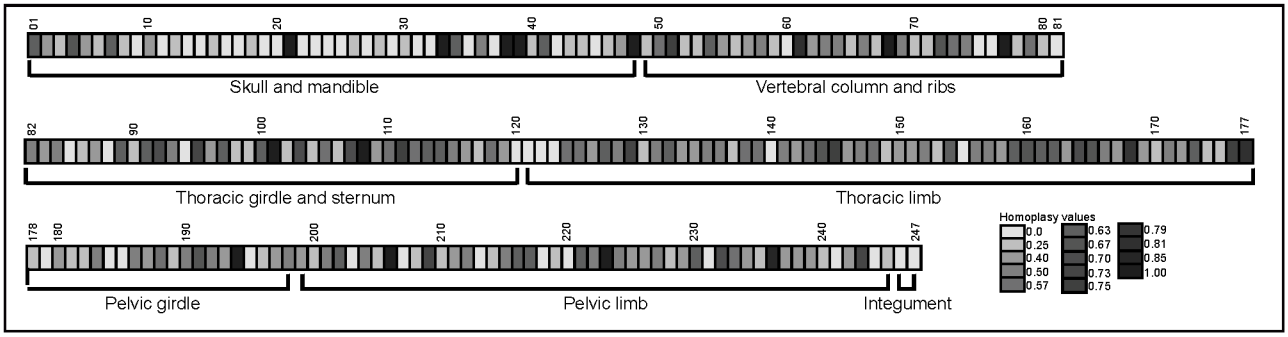
[a] The following set of parameters was implemented in the analyses: coding variable, APRV unlinked and linked, ACRV per partition gamma, and $\lambda = 5$.
[b] MgL values taken from Wright (2015).



CEA



OZL



SCO