# Congruence test of molecular clock calibration hypotheses based on Bayes factor comparisons

Carmelo Andújar<sup>1</sup>\*, Víctor Soria-Carrasco<sup>2</sup>, José Serrano<sup>1</sup> and Jesús Gómez-Zurita<sup>2\*</sup>

<sup>1</sup>Departamento de Zoología y Antropología Física, Facultad de Veterinaria, Universidad de Murcia, 30071 Murcia, Spain; and <sup>2</sup>Animal Biodiversity and Evolution, Institut de Biologia Evolutiva (CSIC-Univ. Pompeu Fabra), Pg. Marítim de la Barceloneta 37, 08003 Barcelona, Spain

#### Summary

1 Molecular clock calibration is a crucial step for placing phylogenetic reconstructions in the temporal framework required to test evolutionary hypotheses and estimate evolutionary rates. In general, most authors agree that the best approach is to incorporate multiple calibrations to avoid the risk of bias associated with a single dating source. However, the indiscriminate inclusion of as many calibration points as possible can lead to tree shape distortion and an overestimation of the variation in evolutionary rates among branches due to errors in the geological, paleontological or paleogeographic information used for dating.

**2** We present a test of congruence among calibration hypotheses to assist their filtering prior to molecular clock analysis, which we have called Bayes Factor Cluster Analysis (BFCA). This is a heuristic method based on the comparison of pairwise calibrations hypotheses by Bayes factors that allows identifying sets of congruent calibrations.

**3** We have tested BFCA through simulation both using BEAST and MCMCTree programs and also to analyse a real case of multiple calibration hypotheses to date the evolution of the genus *Carabus* (Coleoptera: Carabidae).

**4** The analyses of simulated data showed the predictability of change in Bayes factors when comparing alternative calibration hypotheses on a particular tree topology, and thus the suitability of BFCA in identifying unreliable calibrations, especially in cases with limited variation in evolutionary rates among branches. The exclusion of inconsistent calibrations as identified by BFCA produced significant changes in the estimation of divergence times and evolutionary rates in the genus *Carabus*, illustrating the importance of filtering calibrations before analyses.

5 The method has been implemented in an open-source R package called bfca to simplify its application.

**Key-words:** molecular clock, Bayes factors, calibration hypotheses, Bayes Factor Cluster Analysis, substitution rate

#### Introduction

The molecular clock hypothesis states that molecular changes occur at an approximately constant rate over time, and therefore, molecular divergence between species is proportional to time as their evolutionary separation (Zuckerkandl & Pauling 1965). In phylogenetics, this rate of change is typically inferred by extrapolating from the known age of a particular node, in a process known as calibration. Nonetheless, heterogeneity in rates of molecular evolution is common among genes (Wolfe, Sharp & Li 1989; Aguileta, Bielawski & Yang 2006; Pons et al. 2010) and species (Bousquet et al. 1992; Thomas et al. 2006; Bromham 2009; but see criticism by Schwartz & Mueller 2010). Thus, the need to accommodate such rate heterogeneity has stimulated the development of a wealth of relaxed clock methods, including maximum-likelihood smoothing (Sanderson 1997, 2003) and Bayesian estimates based on calibration priors (Thorne & Kishino 2002; Yang & Rannala 2006;

\*Correspondence author. E-mails: candujar@um.es, j.gomez-zurita@ ibe.upf-csic.es Drummond & Rambaut 2007). Among these, the latter are perceived as superior for their ability to realistically account for uncertainties in calibrations by using statistical distributions with soft bounds (Inoue, Donoghue & Yang 2010). However, besides biological sources of heterogeneity in molecular evolutionary rates and regardless of the approach employed, the application of the molecular clock can still be flawed because of methodological hindrance in accurate branch length estimation (Hedges & Kumar 2003) and, especially, of the unreliability of selected calibration ages and their correct application to the appropriate cladogenetic events (Smith & Peterson 2002; Graur & Martin 2004; Near, Meylan & Shaffer 2005; Ho & Phillips 2009).

Calibration of phylogenetic trees with paleontological or geological data is a process subject to multiple difficulties and potential sources of error. These include incompleteness of the fossil record, taxonomic misidentification, dating errors either of fossil strata or specific geologic events, suitability of specific biogeographic hypotheses, incomplete taxon sampling or even the assignation of calibrations to the appropriate nodes in the phylogeny (Benton & Ayala 2003; Bromham & Penny 2003; MEE312151Dispatch: 18.12.13Journal CodeManuscript No.No. of pages: 17

CE: Raja S. PE: Durga Near & Sanderson 2004; Heads 2005, 2010; Ho & Phillips 2009). The concurrence of so many sources of uncertainty has favoured the idea of using simultaneously as many calibrations points as possible to overcome the potential biases associated with individual calibrations (Smith & Peterson 2002; Soltis *et al.* 2002; Conroy & van Tuinen 2003; Graur & Martin 2004). However, these approaches produce compromise solutions that may artificially distort the shape of the tree (branch lengths and topology), thus calling for methods to evaluate the incongruence of calibrations.

There is indeed a growing interest in the development of procedures to assess congruence of calibration hypotheses when multiple calibrations are available. An example of a simplistic approach conditioned by the clock-like behaviour of data is the regression of calibration hypotheses to a linear equation with slope informing of the substitution rate and removal of outliers a posteriori (Gómez-Zurita 2004). Other methodology allowing for relaxed clocks is the cross-validation method described by Near, Meylan & Shaffer (2005), extended by Noonan & Chippindale (2006), Burbrink & Lawson (2007) and Clarke, Warnock & Donoghue (2011). This method, mostly applied to fossil calibrations, has been criticized for producing biased results due to the taphonomic bias in the fossil record (i.e. the increasing probability of fossil preservation towards the present; Marshall 2008; Dornburg et al. 2011). To deal with this problem, Marshall (2008) proposed a novel way to assess potential calibrations by taking an inverse approach to the cross-validation, based on the selection of a single fossil providing the oldest evolutionary time-scale. Marshall's method relies on calculations of an empirical scaling factor for each fossil using relative branch lengths of a given ultrametric, non-calibrated tree. Recently, Dornburg et al. (2011) proposed a Bayesian extension of this method by calculating distributions of scaling factors over the credible interval of branch length and topological estimates, and by assessing the overlap of the 95% highest posterior density (HPD) intervals of available potential calibrations to the one selected by the scaling factors of Marshall (2008). Finally, other methods, like those proposed by Sanders & Lee (2007) or Pyron (2010), evaluate the accuracy of a calibration relative to others assumed as reliable.

In this study, we present a novel approach, which we have called Bayes Factor Cluster Analysis (BFCA), to objectively select a group of congruent calibration hypotheses from a set of potential calibration scenarios. This methodology relies on the use of Bayes Factor (BF) comparisons (Kass & Raftery 1995; Suchard, Weiss & Sinsheimer 2001) and allows for the evaluation of both paleontological and geological calibration hypotheses introduced in Bayesian phylogenetic analyses in the form of probability density functions, as implemented in BEAST (Drummond & Rambaut 2007) and MCMCTree (Yang & Rannala 2006). We have explored the effect of incongruence among node age calibrations on BF values and evaluated the performance of the method using simulated data. Finally, we have applied BFCA to investigate several alternative calibration scenarios to date a phylogenetic tree of the beetle genus Carabus inferred from the widely used nd5 gene, providing a

real example of the effect of incorporating inappropriate calibrations in the estimation of divergence times and evolutionary rates. The method has been implemented in the open-source R package *bfca*. We have also developed a set of tools in the form of Perl scripts to assist in the generation of input files required for the multiple runs of BEAST and MCMCTree, as well as the posterior processing of log files. This software is freely available from https://bitbucket.org/visoca/bfca and http://www.ibe. upf-csic.es/soft/softwareanddata.html.

#### Materials and methods

#### BAYES FACTOR CLUSTER ANALYSIS PROCEDURE

Probabilistic phylogenetic methods aim at finding the tree topology that maximizes a likelihood function for a particular data set under an evolutionary model (Felsenstein 1981). Choices affecting this model of evolution and topological constraints have an effect on likelihood estimates (Sullivan & Joyce 2005). This is the basis of different phylogenetic tests routinely conducted to compare competing hypotheses, such as the likelihood ratio test or BFs (Huelsenbeck, Hillis & Nielsen 1996; Suchard, Weiss & Sinsheimer 2001). The latter are defined in a Bayesian framework as the ratio of marginal likelihoods (i.e. the likelihood of the data under a particular model after integrating across all possible parameter values) from two competing hypotheses and have been interpreted as the relative success of each hypothesis at predicting the data (Kass & Raftery 1995; Brown & Lemmon 2007).

In principle, if several calibrations are applied simultaneously in an analysis, the more reciprocally inconsistent they are among each other and, critically, with the underlying genetic variation accounting for branch lengths, the higher will be their effect on the optimal tree shape, and consequently on the tree likelihood. Thus, BF comparisons could be potentially applied to all possible combinations of available age calibration hypotheses to investigate their mutual consistency and their fit with the data. Such an exhaustive procedure would allow identifying the most inclusive combination of calibrations resulting in differences in marginal likelihood within an acceptance threshold defined a priori by the researcher, typically with a value of  $2\ln BF > 2$  (Kass & Raftery 1995). However, the number of combinations increases exponentially with the number of calibration hypotheses, and this procedure quickly becomes computationally intractable. To overcome this limitation, we propose a method based on analyses of pairwise calibrations and the posterior selection of the largest subset of calibrations without positive evidence of incongruence affecting any of its paired elements. The method involves four well-defined steps:

**1** Estimation of rate variation. The model of rate variation applying to the data – strict clock (SC) or uncorrelated log-normal relaxed clock (ULN) – is selected previously based on BFs from analyses without calibration information. If a ULN clock model is favoured, the standard deviation of the log-normal distribution of branch rates (ULNSD; *ucld.stdev* in BEAST; *sigma2\_gamma* in MCMCTTEE) is fixed in all subsequent pairwise analyses to a value estimated previously using a single arbitrary calibration for the root (in our case, a normal distribution with mean = 100, SD = 0·1). This value, based exclusively on sequence information, provides a limit to rate variation, which may otherwise compensate potential discordance among calibration priors by introducing extra variation of among-branch rates.

**2** *Pairwise analyses of calibration points.* All possible pairwise combinations from a collection of available calibration hypotheses are successively used in independent calibration Bayesian analyses. **4** *Clustering and selection of hypotheses.* Bayes factors indicating positive evidence in favour of one pairwise calibration hypothesis over another are used to identify which calibrations are incongruent. Only combinations of calibrations without any such conflicting pairs are considered, and the most inclusive subset of mutually concordant hypotheses is selected for calibration analyses. Here, we considered a  $2\ln BF > 2$  as an initial reference threshold for positive evidence as proposed by Kass and Raftery (1995), although it can be modified in *bfca* for flexibility in the analyses. We explored the effect of varying this threshold with simulated data (see below).

#### PROOF-OF-PRINCIPLE

We used RATEEVOLVER 1.1 (Ho 2005) to generate an ultrametric tree of nine taxa and spanning 20 time units (Fig. 1a; from Drummond et al. 2006), and from this tree, we simulated a 1000 nucleotide-long alignment under a GTR model using sEQ-GEN 1.3.2 (Rambaut & Grassly 1997). We used these simulated ultrametric data (i) to demonstrate our fundamental assumption that congruence of the tree calibration constraints quantitatively affect marginal-likelihood scores and (ii) to evaluate approximately the minimum degree of incongruence required between two calibration hypotheses to reflect positive evidence using BF criteria (i.e.  $2\ln BF > 2$ ). Bayesian calibration analyses with varying calibration age densities were carried out both with BEAST 1.5.4 and MCMCTree under three different clock relaxation scenarios: strict clock, ULN allowing moderate clock relaxation (ULNSD fixed to 0.3) and ULN allowing high clock relaxation (ULNSD fixed to 0.6). We kept constant the age distribution of one node (A, E or G; see Fig. 1a) and varied the calibration density of another node, which was contemporaneous, younger or older than the corresponding constant-calibrated node. Depending on the statistical distribution used for calibration, the mean or the offset assigned to the variable prior distribution varied across a range of 23 values around the real age of the node (Tables S1-S4, Supporting information). Calibration age density distributions for both constant and varying nodes were implemented in three different fashions - including normal, exponential and log-normal distributions in BEAST, and normal, narrow and wide skew-normal distributions in MCMCTree - in order to explore their effect in the analyses. BEAST analyses consisted of two MCMC runs of 30 million generations, sampling every 1000 generations and were conducted using a GTR substitution model, a Yule tree prior, and constraining the ucld.stdv parameter to 0.3 or 0.6 for tests using relaxed clocks. MCMCTree analyses included two MCMC runs of 100 000 generations sampling every ten generations and were carried out using an HKY substitution model - the most complex model allowed by the software for exact likelihood calculation - a Yule tree prior and constraining the sigma2\_gamma parameter to 0.09 or 0.36 for the ULN clock cases (other details on priors in Tables S5 and S6, Supporting information). We discarded the initial 10% of samples as burn-in, and checked that effective sampling size for the posterior likelihood was always above 200, to ensure that analyses ran long enough to obtain stable estimations. Additionally, we compared the results of the two independent MCMC runs for each analysis, obtaining differences in marginal likelihood below 0.5 in 99.9% of cases (Fig.

S1, Supporting information). For each data set, marginal likelihoods were computed, and lnBF estimated relative to the best marginal-likelihood value obtained for any of the 23 ages assigned to the variable node. LnBF values were plotted against variable node age to confirm their increase with escalation of incongruence between node ages, as well as to recognize the range of overlap below the threshold for rejection of hypothesis congruence.

#### VALIDATION OF BAYES FACTOR CLUSTER ANALYSIS PERFORMANCE THROUGH SIMULATION

We investigated the performance of BFCA over a range of analytical conditions. We produced nine calibration scenarios, each of them with five nodes calibrated using a variety of density functions as may be carried out in a real-tree calibration exercise (Fig. 1b). Within each of these scenarios, three nodes were constrained to be mutually concordant and two to be incongruent relative to their real node ages.

Bayes Factor Cluster Analyses were carried out for the nine calibration scenarios with 90 simulated data sets generated based on an ultrametric tree of nine taxa and 20 units of time span (Fig. 1a; from Drummond et al. 2006). We used RATEEVOLVER 1.1 (Ho 2005) using a ULN clock model to generate 30 trees with moderate among-branch rate variation (ULNSD = 0.3, ULNmod) and 30 trees with high among-branch rate variation (ULNSD = 0.6, ULNhig). From each of these trees, 1000 nucleotide-long alignments were simulated using sEQ-GEN 1.3.2 (Rambaut & Grassly 1997) under a GTR model. Similarly, 30 additional data sets were simulated from the original ultrametric tree to investigate the performance of the method under strict clock (SC) evolution. In all cases, phylogenetic dating was carried out with BEAST and MCMCTree as indicated above (involving a total of 32 400 runs). When using a ULN relaxed clock model, we first estimated the mean ULNSD using a normal distribution ( $\mu = 20, \sigma = 0.1$ ) as an arbitrary calibration prior on the root node, and thereafter, we fixed the estimated value in subsequent analyses using pairwise calibrations (see above). We quantified: (i) the number of times that the BFCA recovered the three calibrations defined a priori as congruent and rejected the incongruent information (positive result); (ii) the number of cases where some of the a priori congruent calibrations were rejected (negative, type  $\rho$  result); and (iii) the number of cases when one or more incongruent calibrations were not rejected and were thus included in the final set along with all the congruent calibrations (uninformative, type  $\sigma$  result). We examined both the effect of the clock model on the rate of recovery of positive results and also the effect of varying lnBF thresholds (ranging from 0.2 to 4).

#### BAYES FACTOR CLUSTER ANALYSIS OF A REAL CALIBRATION EXAMPLE: THE BEETLE GENUS CARABUS

After the reliability of the BFCA method was assessed on simulated data, we explored its utility on real data investigating the evolutionary rate of the *nd5* gene in the genus *Carabus* (Coleoptera: Carabidae). The *nd5* gene of *Carabus* counts with more than 3000 entries in public sequence databases (February 2013) and has been used repeatedly to date the evolution of this genus under several available calibration scenarios. However, the use of calibrations based on fossil and/or biogeographic data has led to inconsistent results (Prüser & Mossakowski 1998; Su *et al.* 1998; Tominaga *et al.* 2000; Osawa, Su & Imura 2004; Sota *et al.* 2005; Andújar, Serrano & Gómez-Zurita 2012). Thus, calibration of the *nd5* tree of *Carabus* represents an adequate test case to apply our BFCA approach.



**Fig. 1.** (a) Rooted binary tree used for simulating sequence evolution. Time-scale represents arbitrary time units. (b) Hypothetical scenarios explored to test the performance of Bayes Factor Cluster Analysis (BFCA). Calibration priors were modelled as normal (Nr), exponential (Ex) or log-normal (Ln) distributions for BEAST analyses and using the skew-normal (SN) with varying asymmetry for MCMCTree. Parameters defining distributions are indicated on each node.  $\mu$ : mean,  $\sigma$ : standard deviation, *a*: lower bound, *b*: upper bound, ofs: offset. Normal distributions marked with an asterisk were truncated to avoid analytical problems due to possible exploration of negative height values.

We built a data set using 37 *nd5* sequences from public nucleotide sequence databases (Benson *et al.* 2010) and 21 newly generated homologous sequences from specimens selected to allow for the inclusion of several calibration points. Beetle DNA was extracted, purified and sequenced as indicated in Andújar, Serrano & Gómez-Zurita (2012) (Data S1, Supporting information). Species data, GenBank accession numbers and source for each sequence are indicated in Table S7 (Supporting information).

Alignment of protein-coding *nd5* sequences did not require gaps, and the resulting manually aligned 904 nucleotide-long matrix of *nd5* sequences from 58 species was used for subsequent phylogenetic analyses and the application of the BFCA test. The evolutionary model best fitting the *nd5* data matrix was calculated with jModelTest (Posada 2008) and selected under the Akaike information criterion (Posada & Buckley 2004). Maximum-likelihood (ML), parsimony (MP) and Bayesian inference (BI) analyses were conducted to check the reliability of nodes used in subsequent calibration tests (details in Data S2, Supporting information).

A suite of calibration hypotheses representing different taxonomic splits in *Carabus* was available from the literature. We selected 16 such hypotheses which span the breakup of Gondwana, the formation of the Canary Islands, the evolution of the Western Mediterranean and the tectonic separation of the Japanese archipelago from the mainland (with two alternative time frameworks), as well as one taxonomically reliable Tertiary fossil (Table 1; Data S3, Supporting information). Calibration density functions on the age of the corresponding cladogenetic events were modelled in all subsequent calibration analyses in BEAST as described in Table 1.

To select the optimal clock and nd5 data partition models for the data, we conducted BF comparisons evaluating six different schemes in BEAST 1.5.4: (i) SC without partitioning [NP], (ii) SC and two partitions considering first and second codon positions together [2P], (iii) SC and each codon position as a different partition [3P], (iv) ULN and NP, (v) ULN and 2P, and (vi) ULN and 3P. We used a GTR + I + G substitution model with a gamma distribution estimated using four rate categories. A speciation Yule process was used as tree prior, and a normal distribution ( $\mu = 100, \sigma = 0.1$ ) was set as an arbitrary calibration prior for the root. Two independent MCMC chains were run for 50 million generations, sampling every 2000th generation. Log files were combined after removing 20% of the initial values as burn-in, and marginal-likelihood scores were estimated as described before. The estimated difference in marginal likelihood (lnBF) was interpreted as requiring at least a ten units increase per additional free parameter (p) (ratio  $\ln BF/\Delta p$  higher than 10) before accepting a more complex model (Pagel & Meade 2004; Miller, Bergsten & Whiting 2009). We assumed one extra parameter for the ULN clock relative to the SC as suggested by Drummond et al. (2006).

The selected clock and partition models were used to perform a BFCA of the 16 calibration hypotheses (four analyses including pairs of calibrations affecting the same node were excluded because they represented alternative time frameworks for the evolution of the Japanese archipelago; Table 1). The obtained subset of congruent calibration hypotheses along with the preferred clock and partition models were used to infer the phylogeny of the genus *Carabus* and to estimate the molecular rate of evolution for the *nd5* gene. In this case, we ran two independent runs in BEAST as above (details in Table S8, Supporting information), but increased the number of gamma categories to ten to improve the accuracy in the estimation of the evolutionary rate. Additionally, we evaluated the effect of different calibration strategies on rate estimation, using each calibration individually and also the combination of all calibrations simultaneously. Rate values are indicated as their mean and 95% HPD interval.

#### Results

### EVALUATION OF INCONGRUENCE USING BAYES FACTORS

A simulated case representative of the tree in Fig. 1a was investigated to evaluate the interaction of calibrations with increasing incongruence. The eight combinations of constant vs. variable node age priors produced very similar results independently of the methodology used, generally representing an increasing difference in marginal-likelihood scores as the calibration for the age of the variable node departed from its true value, but also depending on the age interval allowed by the corresponding probability density functions (Figs S2 and S3, Supporting information). Figure 2 shows an example of results obtained in BEAST when the calibration prior of node A was kept constant and that of node B varied along the age gradient.

Any combination of normal (Nr) and exponential (Ex) or narrow skew-normal (SNn) calibrations produced curves with InBF values growing both sides asymmetrically from a minimum lnBF value ( $\approx 0$ ) centred on the real age value of the variable node when its constrained age slid both towards the past or the present. This trend represents the increase in lnBF value as the variable calibration prior departs from the real value until it eventually exceeds the threshold selected to indicate positive evidence in favour of one hypothesis. The usage of wide calibration distributions, as we did for log-normal (Ln) and wide skew-normal (SNw) distributions, resulted in higher chances of congruence between calibrations, as expected. These wide calibrations showed an asymmetrical effect in hypotheses concordance, because of their shape including a hard bound and a long tail. Consequently, in cases where both calibrations were modelled with wide distributions, no trend towards incongruence was observed for the tested range of age differences.

The main difference between clock models was the noticeable and expected decrease in difference between marginal-likelihood values when a relaxed clock was used. Relaxed clocks broadened the range for hypothesis congruence to the point that, when high rate variation (ULNSD = 0.6) was allowed, incongruence was not reached in most cases. Otherwise, SC and ULN with moderate among-branch rate variation (ULNSD = 0.3) produced similar results, but expectedly, the calibration incongruence for positive evidence favouring one hypothesis over another was clearly broadened with the relaxed clock (Fig. 2). In the analytical conditions used for this example, the range of age differences between nodes where the test did not discriminate a pair of calibrations as inconsistent varied between 1.7 and 7.7 time units (depending on the relative positions of nodes; see Discussion) in the SC case. This range of age differences varied between 4.5 and 22.4 time units in the ULN allowing moderate clock relaxation (Tables S9 and S10, Supporting information). Thus, clock relaxation accommodated part of the incongruence and, as a consequence, the difference in age between calibrations required to produce positive evidence for one hypothesis over another increased

| C*Split hereen work charmin<br>(Neurosciences)Volumic encregence<br>(According Field)143Linform (a = 0, b = 145)003.14143.44.145055F2Carry Marchines(According Field)Messimin deposit of<br>(Carry Marchines)Messimin deposit of<br>(Carry Marchines)3.5Lage normal ( $\mu = 35$ , $3.4-1450$ 9.03.14143.44.14509.0F2Carry MarchinesMessimin deposit of<br>(Carry Marchines)3.5Lage normal ( $\mu = 35$ , $3.5-346$ 3.9-34753.91Radiation of <i>Lapraembers</i> .1Mandren minimal<br>(monoprone)3.5The above0.03.14143.9-34753.9-34752.21.1Radiation of <i>Lapraembers</i> .1Mandren minimal<br>( $\mu = 35$ , $\mu = 0, h = 100$ )10.033.5The above3.9-34753.9-34752.21.1Radiation of <i>Lapraembers</i> .1Mandren minimal<br>( $\mu = 35$ , $\mu = 0, h = 100$ )10.033.5The above3.5The above3.72.1-22.1-21.1Radiation of <i>Dimener</i> .11Mandren minimal<br>( $\mu = 36$ , $\mu = 1, h = 100$ )13.3The above3.5The above3.73.7-3463.7-3562.71.1Radiation of <i>Dimener</i> .11Mandren minimal<br>( $\mu = 36, m = 10, m = 100$ )13.0The above3.7-141.72.02.01.2Radiation of <i>Dimener</i> .11Mandren minimal<br>( $\mu = 36, m = 10, m = 100$ )13.01.10-10001.70-10001.70-10001.70-10001.2Radiation of <i>Dimener</i> .11Mandren minimal<br>( $\mu = 36, m = 10, m = 1000$ )1.70-10001.70   | Calibration<br>hypothesis | Evolutionary event: node   | Calibration event  | Age of<br>event<br>(Ma) | Priors on<br>node ages  | 95% probability<br>density | 95% HPD<br>effective prior<br>age (Final analysis) | 95% HPD<br>posterior age<br>(Final analysis) |
|---|---------------------------|--|--|-------------------------|---|----------------------------|--|--|
| P:2Cardward<br>control (Amounded)Mestimin deposit of<br>and (Amounded)SLegenomia ( $\mu = S_1$ S + 158.5S 66 41891Radiation of <i>Layouxabus</i> 13Find decombination35Truncated Normal ( $\mu = S_1$ 35.55.46319.587322Radiation of <i>Layouxabus</i> 13Find decombination35Truncated Normal ( $\mu = S_1$ 155.54.6319.587323Radiation of <i>Layouxabus</i> 13Find decombination35Truncated Normal ( $\mu = S_1$ 155.54.6319.587324Split between subcursFind decombination35Truncated Normal ( $\mu = S_1$ 319.5587323210Radiation of <i>Layouxabus</i> 13Initial decombination35Truncated Normal ( $\mu = S_1$ 319.5586319.558732110Radiation of <i>Layouxabus</i> 13Initial decombination35Truncated Normal ( $\mu = S_1$ 319.416.9612bRadiation of <i>Layouxabus</i> 13Initial decombination15Normal ( $\mu = S_1 = 0, 1, 0 = 1000$ )155.54613bRadiation of <i>Layouxabus</i> 13Initial decombination15Normal ( $\mu = S_1 = 0, 1, 0 = 1000$ )155.54613bRadiation of <i>Layouxabus</i> 13Initial decombination15Normal ( $\mu = S_1 = 0, 1, 0 = 1000$ )155.54613bRadiation of <i>Layouxabus</i> 13Initial decombination15Normal ( $\mu = S_1 = 0, 1, 0 = 1000$ )155.54613bRadiation of <i>Layouxabus</i> 13Initial decombi  | Č                         | Split between two Canarian<br>endemic species: <i>Carabus</i><br>( <i>Nesaeocarabus</i> ) <i>coarctatus</i> and <i>C</i> . | Volcanic emergence<br>of Gran Canaria                            | 14.5                    | Uniform $(a = 0, b = 14.5)$   | 0.03-14.14                 | 3.44-14.50   | 5.29-10.48                                   |
| 11Reduction of Lipocenduc: 1)Fundal disconnection<br>control disconnection<br>335Tensted Normal ( $\mu = 35$ ,<br>$\sigma = 1, \sigma = 0, 1, b = 1000$ )39 553-46319-5873212Radiation of Lipocenduc: 13Tapar from minimod<br>Tapar from minimod35Tuneated Normal ( $\mu = 35$ ,<br>$\sigma = 1, \sigma = 0, 1, b = 1000$ )155-546319-5871213Radiation of Lipocenduc: 13Tapar from minimod<br>Tapar from minimod35Tuneated Normal ( $\mu = 35$ ,<br>$\sigma = 1, \sigma = 0, 1, b = 1000$ )155-546319-5872214Split between subgrant<br>Isoconduc and<br>Disport from minimod35Tuneated Normal ( $\mu = 35$ ,<br>$\sigma = 1, \sigma = 0, 1, b = 1000$ )155-546319-5872213Radiation of Dimense: 11Initial disconnection of<br>Accordanci and<br>Disport from minimod15Normal ( $\mu = 15, \sigma = 1$ )1304-16.9613Radiation of Dimense: 13Initial disconnection of<br>Accordanci and<br>Disport from minimod15Normal ( $\mu = 15, \sigma = 1$ )1304-16.9613Radiation of Dimense: 14Initial disconnection of<br>Activation of Dimense: 14Initial disconnection of<br>Activation of Dimense: 141304-16.9614Split between chargesJapan from minimod15Normal ( $\mu = 15, \sigma = 1$ )1304-16.9615Radiation of Dimense: 14Dipant from minimod15Normal ( $\mu = 15, \sigma = 1$ )1304-16.9616Split between chargesDipant from minimod15Normal ( $\mu = 15, \sigma = 1$ )1304-1                | F2                        | Carabus (Autocarabus)  | Messinian deposit of   | 5                       | Log-normal ( $\mu = 25$ ,   | 5.4-158.5                  | 5.06-14.18   | 9.93-19.02                                   |
| 12Radiation of Leprocendors 12Tradition of Leprocendors 12Tradition of Compensation<br>apon from minimal<br>33Transated Normal ( $\mu = 35$ , $1, 55, 546$ 3, 95,85811,13Radiation of Onemperas: 13Apon from minimal<br>apon from minimal3Transated Normal ( $\mu = 35$ , $1, 55, 546$ 3, 95,8582314Split between subgenusTradificionnection of<br>Alocardov and<br>Alocardov and<br>Midition of <i>Demogenes</i> . 1313Transated Normal ( $\mu = 35$ , $n = 1, \sigma = 0, 1, \rho = 1000$ )13, 55,54613Radiation of <i>Demogenes</i> . 13Japon from minimal<br>Alocardov and<br>Midition of <i>Demogenes</i> . 13Japon from minimal<br>a = 1, $\sigma = 0, 1, \rho = 1000$ 13, 04,169613Radiation of <i>Demogenes</i> . 13Japon from minimal<br>Midition of <i>Demogenes</i> . 13Japon from minimal<br>a = 1, $\sigma = 0, 1, \rho = 1000$ 13, 04,169614Split between subgenusJapon from manihand<br>Midition of <i>Leprocendus</i> . 13Japon from manihand<br>Midition of <i>Demogenes</i> . 13Japon from manihand<br>a = 1, $\sigma = 0, 1, 0, 1000$ 13, 04,169615Radiation of <i>Demogenes</i> . 13Japon from manihand<br>Midition of <i>Leprocendus</i> . 13Japon from manihand<br>Midition manihand15Normal ( $\mu = 15, \sigma = 1$ )13,04,169615Radiation of <i>Demogenes</i> . 13Japon from manihand5.33Exponential ( $\mu = 0, 0, 0, 0, 10, 0, 10, 0, 10, 10, 10, 0, 0, 10, 1$ | Iſ                        | cancellatus Iossil<br>Radiation of Damaster. J1  | Cantal (France)<br>Final disconnection of<br>Ianan from mainland | 3.5                     | $\sigma = 1.5, \text{ offset} = 5)$<br>Truncated Normal ( $\mu = 3.5, \sigma = 1, \sigma = 0.1, h = 1000$ ) | 1.55-5.46                  | 3.19-5.87  | 2.29 4.32                                    |
| 13Radiation of Ohomoprene: 13Final disconnection of<br>and disconnection of<br>biand disconnection of<br>   | J2                        | Radiation of Leptocarabus: J2  | Final disconnection of<br>Isnan from mainland                    | 3.5                     | Truncated Normal ( $\mu = 3.5$ ,<br>$\sigma = 1$ , $a = 0.1$ , $b = 1000$ )                                 | 1.55-5.46                  | 3.19-5.85  | 1.26-3.13                                    |
| 14Split between subgenus<br><i>Inconders</i> and<br><i>Inconders</i> and<br><i>Inconders</i> and<br><i>Inconders</i> and<br><i>Inconders</i> and<br><i>Inconders</i> and<br>  | J3                        | Radiation of Ohomopterus: J3   | Final disconnection of<br>Ianan from mainland                    | 3.5                     | Truncated Normal $(\mu = 3.5, \sigma = 1.00)$   | 1.55-5.46                  | 3.47–5.86  | 2.37-4.26                                    |
| JIbRadiation of Domaster, J1Initial disconnection of<br>Japan from mainandJ3Normal ( $\mu = 15, \sigma = 1$ )304-16-96-J2bRadiation of Leptocarduse; J2Initial disconnection of<br>Japan from mainandJ5Normal ( $\mu = 15, \sigma = 1$ )1304-16-96J3bRadiation of Leptocarduse; J3Initial disconnection of<br>Japan from mainlandJ5Normal ( $\mu = 15, \sigma = 1$ )1304-16-96J4bSplit between subgemsInitial disconnection of<br>  | <b>J</b> 4                | Split between subgenus<br>Isiocarabus and  | Final disconnection of<br>Japan from mainland                    | 3.5                     | Truncated Normal ( $\mu = 3.5$ ,<br>$\sigma = 1, a = 0.1, b = 1000$ )                                       | 1.55–5.46                  | 1  | I  |
| 12bRadiation of Leptocarchus: J2Japan from mainland<br>light if sconnection of<br>J3b15Normal $\mu = 15, \sigma = 1$ 13-04-16-96J3bRadiation of Leptocarchus: J3Injuil disconnection of<br>Japan from mainland15Normal $\mu = 15, \sigma = 1$ 13-04-16-96J4bSplit between subgenus<br>faiocarabus and Ohmoprens: J4Injuil disconnection of<br>Japan from mainland15Normal $\mu = 15, \sigma = 1$ 13-04-16-96MilSplit between Carchus (Mesocarchus)<br>right between Carchus (Enzycarchus)Opening Gibra lar strait<br>5-335-33Exponential $(\mu = 0.5, offset = 5.3)$ 5-31-7.14M3Split between Carchus (Enzycarchus)<br>area from Corsta and North<br>Africe neodous (Enzycarchus)Opening Gibra lar strait<br>   | JIb                       | Oriomopie us. J+<br>Radiation of Damaster: J1  | Initial disconnection of   | 15                      | Normal ( $\mu = 15, \sigma = 1$ )   | 13.04-16.96                | Ι  | I  |
| J3bRadiation of Ohomoprense: J3Japan from manified<br>Japan from manified<br>J4bJ304-16.96J4bSplit between subgenues<br>faceradeus and Ohomoprense: J3Japan from manified<br>   | J2b                       | Radiation of <i>Leptocarabus</i> : J2  | Japan from mainland<br>Initial disconnection of                  | 15                      | Normal $\mu = 15$ , $\sigma = 1$ )  | 13-04-16-96                | I  | I  |
| J4bSplit between subgenus<br><i>layan trom mainland</i> Japan from mainlandJapan from from mainlandJapan from mainlandJapan from from from from from from from from  | J3b                       | Radiation of <i>Ohomopterus</i> : J3   | Japan from mainland<br>Initial disconnection of                  | 15                      | Normal ( $\mu = 15, \sigma = 1$ )   | 13.04-16.96                | I  | I  |
| M1Split between <i>Carabus (Mescarabus)</i><br>infjersis and Iberian <i>Mescarabus)</i> Split between <i>Carabus (Mescarabus)</i><br>opening Gibraltar strait5.33Exponential ( $\mu = 0.5$ , offset = 5.3)5.31-7.14-M2Split between <i>Carabus (Mescarabus)</i><br>infjersis and Iberian <i>Mescarabus</i> )Opening Gibraltar strait5.33Exponential ( $\mu = 0.5$ , offset = 5.3)5.31-7.14M3Split between <i>Carabus (Eurycarabus)</i><br>  | J4b                       | Split between subgenus   | Japan from mainland<br>Initial disconnection of                  | 15                      | Normal $(\mu = 15, \sigma = 1)$   | 13.04-16.96                | 12.09-16.04  | 12.76–16.50                                  |
| M2riffensis and Iberian Mesocardues M1<br>genei from Corica and North<br>African Eurycarduss)Dening Gibraltar strait<br>5.335.33Exponential ( $\mu = 0.5$ , offset = 5.3)5.31–7.145.30-5.925.3M3Split between Carduss (Eurycarduss)<br>   | MI                        | Istocarabus and Onomopterus: J4<br>Split between Carabus (Mesocarabus)   | Japan Irom mainland<br>Opening Gibraltar strait                  | 5.33                    | Exponential ( $\mu = 0.5$ , offset = 5.3)   | 5.31-7.14                  | I  | I  |
| African Eurycarabus: M2African Eurycarabus: M2Cpening Gibraltar strait5.33Exponential ( $\mu = 0.5$ , offset = 5.3)5.31-7.145.30-5.905.3M4Split between two Carabus<br>melancholicus subspecies: M3Opening Gibraltar strait5.33Exponential ( $\mu = 0.5$ , offset = 5.3)5.31-7.145.30-5.905.3M4Split between two Carabus<br>melanchorax)Opening Gibraltar strait5.33Exponential ( $\mu = 0.5$ , offset = 5.3)5.31-7.145.30-5.905.3M5Split between two Carabus<br>morbillosus subspecies: M4Opening Gibraltar strait5.33Exponential ( $\mu = 0.5$ , offset = 5.3)5.31-7.14-M5Split between two Carabus<br>morbillosus subspecies: M5Opening Gibraltar strait5.33Exponential ( $\mu = 0.5$ , offset = 5.3)5.31-7.14-M3Split between two Carabus<br>morbinorax)Opening Gibraltar strait5.33Exponential ( $\mu = 0.5$ , offset = 5.3)5.31-7.14-Macrothorax)rugosus subspecies: M5Opening Gibraltar strait5.33Exponential ( $\mu = 0.5$ , offset = 5.3)5.31-7.14-Macrothorax)rugosus subspecies: M5Separation of New Zealand from85Normal ( $\mu = 8.5$ , $\sigma = 3$ )79.12-90.88-Macripamborus: N2Australia-AntarcticaAustralia-Antarctica5.3T9.12-90.88-   | M2                        | riffensis and Iberian Mesocarabus. M1<br>Split between Carabus (Eurycarabus)<br>genei from Corsica and North               | Opening Gibraltar strait   | 5.33                    | Exponential ( $\mu = 0.5$ , offset = 5.3)   | 5.31–7.14                  | 5.30-5.92  | 5.30-6.87                                    |
| M4melancholters subspectes: M3<br>Materioriters subspectes: M4Defining Gibraltar strait5:33Exponential ( $\mu = 0.5$ , offset = 5:3)5:31-7.14-M5Split between two Carabus<br>morbillosus subspecies: M4Opening Gibraltar strait5:33Exponential ( $\mu = 0.5$ , offset = 5:3)5:31-7.14-M5Split between two Carabus<br>morbillosus subspecies: M4Opening Gibraltar strait5:33Exponential ( $\mu = 0.5$ , offset = 5:3)5:31-7.14-M5Split between two Carabus<br>mosus subspecies: M5Opening Gibraltar strait5:33Exponential ( $\mu = 0.5$ , offset = 5:3)5:31-7.14-NZSplit between two Carabus<br>mosus subspecies: M5Searation of New Zealand from<br>Australia-Antarctica85Normal ( $\mu = 85$ , $\sigma = 3$ )79:12-90:88-  | M3                        | African Eurycarabus: M2<br>Split between two Carabus<br>(Rhabdotocarabus)  | Opening Gibraltar strait   | 5.33                    | Exponential ( $\mu = 0.5$ , offset = 5.3)   | 5.31–7.14                  | 5.30-5.90  | 5.30-6.49                                    |
| M5Split between two CarabusOpening Gibraltar strait $5.33$ Exponential ( $\mu = 0.5$ , offset = $5.3$ ) $5.31-7.14$ $-$ Macrothorax)(Macrothorax) $(Macrothorax)$ $5.31-7.14$ $-$ nugosus subspecies: M5Split between Pamborus andSeparation of New Zealand from $85$ Normal ( $\mu = 85$ , $\sigma = 3$ ) $79.12-90.88$ $-$ NZSplit between Pamborus: NZAustralia-AntarcticaAustralia-Antarctica $-$   | M4                        | metancholtcus subspectes: M3<br>Split between two Carabus<br>(Macrothorax)<br>morbillows subspecies: M4                    | Opening Gibraltar strait   | 5.33                    | Exponential ( $\mu = 0.5$ , offset = 5.3)   | 5.31–7.14                  |  | I  |
| NZ Split between <i>Pamborus</i> : NZ Split between <i>Pamborus</i> : NZ Australia-Antarctica $35$ Normal ( $\mu = 85$ , $\sigma = 3$ ) 79.12–90.88 – <i>Maoripamborus</i> : NZ Australia-Antarctica  | M5                        | Split between two Carabus<br>(Marothorax)<br>murosus subserveries: M5  | Opening Gibraltar strait   | 5.33                    | Exponential ( $\mu = 0.5$ , offset = 5.3)   | 5.31-7.14                  |  | 1  |
|   | ZN                        | Split between Pamborus and<br>Maoripamborus: NZ  | Separation of New Zealand from<br>Australia-Antarctica           | 85                      | Normal ( $\mu = 85$ , $\sigma = 3$ )  | 79.12-90.88                |  | I  |

6 C. Andújar et al.



**Fig. 2.** Results of the initial evaluation of incongruence limits for each of the nine paired combinations of prior distributions on nodes A and B from Fig. 1a (Nr: normal; Ex: exponential; Ln: log-normal) under SC, ULN with fixed ULNSD = 0·3 and ULN with fixed ULNSD = 0·6. Constraints on node A remain constant while the age constraint of node B vary along a range of values (*X* axes). Node A prior: Nr ( $\mu = 5$ ,  $\sigma = 0.5$ ); Ex ( $\mu = 0.5$ , offset = 5); Ln ( $\mu = 12$ ,  $\sigma = 1$ , offset = 5). Node B prior: Nr ( $\mu = y$ ,  $\sigma = 0.5$ ); Ex ( $\mu = 0.5$ , offset = *y*); Ln ( $\mu = 12$ ,  $\sigma = 1$ , offset = *y*), whereby y = 0.5, 1, 1·5, 2, 2·5, 3, 3·5, 4, 4·5, 5, 5·5, 6, 6·5, 7, 7·5, 8, 8·5, 9, 9·5, 10, 11, 13. *Y* axes: lnBF. Vertical line indicates the real age of the variable node (node B). The remaining examples are given in Supporting information Figs S2 and S3.

almost threefold. Finally, when high clock relaxation was allowed, the same trend was accentuated and much higher age differences between calibrations were required to produce positive evidence for a hypothesis based on lnBF values.

### BAYES FACTOR CLUSTER ANALYSIS ON SIMULATED DATA

The performance of the BFCA method was evaluated quantitatively based on the rate of correct selection of congruent calibrations and rejection of incongruent ones on a simulated phylogeny of known relationships and ages of branching events. Test performance was investigated under different calibration scenarios using different probability density functions as priors (Fig. 1b), as well as simulating strict or relaxed clocks, the latter with either moderate or high rate variation among branches (ULNmod and ULNhig, respectively).

Each of the nine calibration scenarios was investigated for 90 iterations of simulated data (thirty for each clock model;

810 BFCA tests in total). Results distinguishing between SC, ULNmod and ULNhig cases for a threshold of lnBF > 1 are summarized in Table 2 for BEAST and Table 3 for MCMCTREE. Calibration scenarios simulated under SC produced 95-2% of positive results in BEAST and 88-2% in MCMCTREE. The scenario that presented more difficulties for the correct identification of congruent calibrations was 3a, producing 87-0% and 66-6% of positive results in BEAST and MCMCTREE, respectively (the others ranged 93–100% in BEAST and 76-6–100% in MCMCTREE).

Simulations under the ULNmod clock model generated 40.4% of positive results, 55.6% of type  $\sigma$  uninformative results and 4% of type  $\rho$  negative results in BEAST. In the MCMCTree analyses, we obtained 27.3% of positive results, and 72.3% and 0.4% of  $\sigma$  and  $\rho$  results, respectively. Interestingly, both programs identified the same most problematic scenarios (scenarios 3a, 1a and 1c). For both methodologies, there was a strong negative correlation (Pearson's r = -0.89, P < 0.001) between the proportion of positive results and the estimated mean of the ULNSD parameter. Thus, when the estimated

| 1             | ŠČ       |        |     |      |      |      |      |      | Moon III NGD | nr.  | Nmod  | _     |    |    |    |    |    |    |          | n   | ,Nhig  |      |    |    |    |    |    |
|---------------|----------|--------|-----|------|------|------|------|------|--------------|------|-------|-------|----|----|----|----|----|----|----------|-----|--------|------|----|----|----|----|----|
|               | a l      | lb lc  | 5   | a 2t | ) 2c | 36   | 1 3t | ) 3c | (ULNmod)     | la   | 1b    | lc    | 2a | 2b | 2c | 3а | 3b | 3с | (ULNhig) | la  | 1b     | lc   | 2a | 2b | 2c | 3a | 3b |
| +             | т<br>    | +      | +   | +    | +    | +    | +    | +    | 0.1016       | +    | I     | I     | +  | +  | +  | +  | +  | +  | 0.3892   | 0   | +      | 0    | +  | +  | 0  | 0  | +  |
| +             | J        | +      | +   | +    | +    | +    | +    | +    | 0.1555       | I    | +     | +     | I  | +  | +  | I  | T  | +  | 0.4500   | 0   | I      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | -        | +      | +   | +    | +    | +    | +    | +    | 0.1723       | +    | +     | +     | +  | +  | +  | +  | +  | +  | 0.5141   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | -،<br>ب  | +      | +   | Ι    | +    | +    | +    | +    | 0.1894       | +    | +     | +     | +  | +  | +  | +  | +  | +  | 0.5829   | 0   | 0      | 0    | +  | 0  | 0  | 0  | +  |
| +             | ۲.       | +      | +   | +    | +    | +    | +    | +    | 0.2317       | 0    | Ι     | +     | I  | Ι  | +  | +  | Ι  | +  | 0.6071   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | '<br>-   | +      | +   | +    | +    | +    | +    | +    | 0.2484       | +    | +     | +     | +  | +  | +  | +  | +  | +  | 0.6327   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| т             | ,<br>+   | +      | +   | +    | +    | +    | +    | +    | 0.2979       | 0    | +     | 0     | +  | +  | 0  | 0  | +  | +  | 0.6378   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| т             | '<br>+   | +      | +   | +    | +    | +    | +    | +    | 0.3026       | +    | +     | 0     | +  | +  | +  | +  | +  | +  | 0.6655   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | '<br>+   | +      | +   | +    | +    | +    | +    | +    | 0.3422       | +    | 0     | 0     | +  | +  | +  | +  | +  | +  | 0.6834   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | ۰.<br>۲  | +      | +   | +    | +    | +    | +    | +    | 0.3443       | +    | +     | +     | +  | +  | Ţ  | 0  | 0  | 0  | 0.7072   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | - ر<br>ب | +      | +   | +    | +    | +    | +    | +    | 0.3641       | 0    | 0     | 0     | +  | +  | 0  | 0  | +  | +  | 0.7077   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | ۔<br>۲   | +      | +   | +    | +    | +    | +    | +    | 0.3880       | 0    | 0     | +     | +  | +  | 0  | 0  | +  | 0  | 0.7169   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | ۔<br>۲   | +      | +   | +    | +    | +    | +    | +    | 0.3985       | 0    | +     | 0     | +  | +  | 0  | 0  | 0  | 0  | 0.7274   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | ۔<br>ب   | +      | +   | +    | +    | Ι    | +    | +    | 0.3993       | 0    | +     | 0     | +  | +  | 0  | 0  | +  | +  | 0.7453   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | -        | +      | +   | +    | +    | +    | +    | +    | 0.4057       | +    | 0     | +     | +  | +  | +  | +  | +  | 0  | 0.7749   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| 1             | ,        | +      | +   | +    | +    | I    | +    | I    | 0.4084       | +    | 0     | 0     | +  | +  | +  | 0  | 0  | 0  | 0.7767   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | - ۱<br>ب | +      | +   | +    | +    | T    | T    | +    | 0.4344       | 0    | 0     | 0     | +  | +  | 0  | 0  | 0  | 0  | 0.7883   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| 1             | , F      | +      | +   | +    | +    | Ι    | +    | I    | 0.4526       | 0    | 0     | 0     | +  | 0  | 0  | 0  | +  | 0  | 0.7900   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | -،<br>ب  | +      | +   | +    | +    | +    | +    | +    | 0.4574       | 0    | +     | 0     | +  | +  | +  | 0  | 0  | 0  | 0.8174   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | -،<br>ب  | +      | +   | +    | +    | +    | +    | +    | 0.4584       | 0    | 0     | 0     | 0  | 0  | 0  | 0  | +  | 0  | 0.8209   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | ۰.<br>۲  | +      | +   | +    | +    | +    | +    | +    | 0.4607       | 0    | 0     | 0     | +  | 0  | 0  | 0  | +  | +  | 0.8330   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | ۰.<br>۲  | +      | +   | +    | +    | +    | +    | +    | 0.4996       | 0    | 0     | 0     | 0  | 0  | 0  | 0  | 0  | 0  | 0.8469   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | ۲.       | +      | Ι   | Ι    | +    | +    | +    | +    | 0.5085       | 0    | 0     | 0     | +  | 0  | 0  | 0  | 0  | +  | 0.8947   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | ۰.<br>۲  | +      | +   | +    | +    | +    | +    | +    | 0.5260       | 0    | 0     | 0     | 0  | 0  | 0  | 0  | 0  | 0  | 0.8993   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | -،<br>ب  | +      | +   | +    | +    | +    | +    | +    | 0.5460       | 0    | 0     | 0     | 0  | 0  | 0  | 0  | 0  | 0  | 0.9037   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | ا<br>ب   | +      | +   | +    | +    | +    | +    | +    | 0.5558       | 0    | 0     | 0     | 0  | 0  | 0  | 0  | 0  | 0  | 0.9149   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | -،<br>ب  | +      | +   | +    | +    | +    | +    | +    | 0.5577       | 0    | 0     | 0     | 0  | 0  | 0  | 0  | 0  | 0  | 0.9490   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | ۰.<br>۲  | +      | +   | +    | +    | +    | +    | +    | 0.5730       | 0    | 0     | 0     | 0  | 0  | 0  | 0  | 0  | 0  | 0.9803   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | ۰.<br>۲  | +      | +   | +    | +    | +    | +    | +    | 0.5749       | 0    | 0     | 0     | 0  | 0  | 0  | 0  | 0  | 0  | 1.0798   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| +             | -،<br>ب  | +      | +   | +    | +    | +    | +    | +    | 0.5957       | 0    | 0     | 0     | 0  | 0  | 0  | 0  | 0  | 0  | 1.1053   | 0   | 0      | 0    | 0  | 0  | 0  | 0  | 0  |
| ositives 2    | 82       | 29 3(  | 3   | 9 28 | 30   | ) 2( | 29   | 28   |              | 6    | 10    | ×     | 19 | 17 | Ξ  | ×  | 14 | 13 |          | 0   | -      | 0    | Ч  | -  | 0  | 0  | 2  |
| % Positives 9 | 15.2%    | (257/2 | 70) |      |      |      |      |      |              | 40.4 | % (10 | 9/270 | e  |    |    |    |    |    |          | 2.6 | % (7/2 | 270) |    |    |    |    |    |

Table 2. Summary of the results for the 810 BFCA runs to test method performance through simulation in BEAST. Columns gather the results under a strict clock model (SC), and uncorrelated log-normal relaxed

**Table 3.** Summary of the results for the 810 BFCA runs to test method performance through simulation in MCMCTRE. Columns gather the results under a strict clock model (SC), and uncorrelated log-normal relaxed clock models with moderate (ULNmod) and high among-branch rate variation (ULNhig). Columns 1a to 3c indicate the nine different calibration scenarios tested (as shown in Fig. 1b). Positive results are

|             | SC   |        |        |    |    |    |    |    |    | Mean     | nr   | Nmoc   | Ŧ      |    |    |    |    |    |    | CISIN 111 MOSIM | NLI | Vhig |    |    |    |    |    |    |    |
|-------------|------|--------|--------|----|----|----|----|----|----|----------|------|--------|--------|----|----|----|----|----|----|-----------------|-----|------|----|----|----|----|----|----|----|
|             | la   | 1b     | lc     | 2a | 2b | 2c | 3а | 3b | 3c | (DLNmod) | 1a   | 1b     | lc     | 2a | 2b | 2c | 3a | 3b | 3c | (ULNhig)        | la  | 1b   | lc | 2a | 2b | 2c | 3a | 3b | 3с |
|             | +    | +      | +      | +  | +  | +  | +  | +  | +  | 0.1591   | +    | I      | +      | +  | +  | +  | +  | +  | +  | 0.4677          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | +  | +  | +  | +  | 0.2195   | +    | +      | +      | +  | +  | +  | +  | +  | +  | 0.4724          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | I      | +      | +  | Ι  | +  | +  | +  | +  | 0.2490   | +    | +      | +      | +  | +  | +  | +  | +  | +  | 0.5030          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | Ι    | +      | +      | +  | +  | Т  | 1  | +  | +  | 0.2915   | +    | 0      | 0      | +  | +  | +  | +  | +  | 0  | 0.5292          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | T  | +  | T  | +  | 0.3245   | 0    | +      | 0      | +  | 0  | 0  | 0  | +  | +  | 0.5420          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | +  | Ι  | +  | +  | 0.3266   | +    | +      | 0      | +  | +  | +  | +  | +  | +  | 0.5803          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | T    | +      | +      | +  | +  | Т  | I  | +  | +  | 0.3526   | 0    | +      | 0      | 0  | 0  | 0  | 0  | +  | 0  | 0.6006          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | +  | I  | I  | +  | 0.3811   | 0    | 0      | 0      | 0  | 0  | 0  | 0  | +  | +  | 0.6085          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | I    | +      | +      | +  | +  | T  | +  | +  | +  | 0.3870   | 0    | 0      | 0      | +  | +  | 0  | 0  | 0  | 0  | 0.6434          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | T  | I  | +  | +  | 0.3885   | 0    | +      | 0      | +  | +  | 0  | 0  | 0  | 0  | 0.6458          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | I    | +      | +      | +  | Ι  | +  | I  | +  | +  | 0.3913   | 0    | +      | 0      | +  | +  | 0  | 0  | 0  | 0  | 0.6615          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | +  | T  | T  | +  | 0.3922   | 0    | 0      | 0      | 0  | 0  | 0  | 0  | 0  | +  | 0.6806          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | +  | +  | +  | +  | 0.3950   | 0    | 0      | 0      | +  | +  | +  | 0  | 0  | 0  | 0.6819          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | Ι  | +  | +  | +  | +  | 0.4084   | 0    | 0      | 0      | 0  | 0  | 0  | 0  | +  | 0  | 0.6834          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | +  | +  | +  | +  | 0.4106   | 0    | 0      | 0      | +  | 0  | 0  | 0  | +  | +  | 0.6998          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | +  | +  | +  | +  | 0.4280   | 0    | 0      | 0      | 0  | 0  | 0  | 0  | 0  | 0  | 0.7051          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | +  | +  | +  | +  | 0.4492   | 0    | 0      | 0      | 0  | 0  | 0  | 0  | +  | 0  | 0.7098          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | I    | +      | +      | +  | +  | +  | +  | +  | +  | 0.4501   | 0    | 0      | 0      | +  | 0  | 0  | 0  | +  | 0  | 0.7185          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | I    | +      | +      | T  | +  | +  | I  | +  | +  | 0.4593   | 0    | 0      | 0      | 0  | 0  | 0  | 0  | +  | 0  | 0.7370          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | +  | +  | +  | +  | 0.4621   | 0    | 0      | 0      | +  | 0  | 0  | 0  | 0  | 0  | 0.7623          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | +  | +  | +  | +  | 0.4651   | 0    | +      | 0      | +  | 0  | 0  | 0  | 0  | 0  | 0.7731          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | +  | +  | +  | +  | 0.4767   | 0    | 0      | 0      | 0  | 0  | 0  | 0  | 0  | 0  | 0.7811          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | +  | +  | +  | +  | 0.4949   | 0    | 0      | 0      | 0  | 0  | 0  | 0  | 0  | 0  | 0.7920          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | +  | I  | +  | +  | 0.4997   | 0    | 0      | 0      | 0  | 0  | 0  | 0  | 0  | 0  | 0.7979          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | I    | +      | +      | +  | +  | +  | I  | +  | +  | 0.5244   | 0    | 0      | 0      | 0  | 0  | 0  | 0  | 0  | 0  | 0.8336          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | +  | +  | +  | +  | 0.5299   | 0    | 0      | 0      | 0  | 0  | 0  | 0  | 0  | 0  | 0.8478          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | +  | +  | +  | +  | 0.5358   | 0    | 0      | 0      | 0  | 0  | 0  | 0  | 0  | 0  | 0.8515          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | I  | +  | +  | +  | 0.5584   | 0    | 0      | 0      | 0  | 0  | 0  | 0  | 0  | 0  | 0.8705          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | +  | +  | +  | +  | 0.5683   | 0    | 0      | 0      | 0  | 0  | 0  | 0  | 0  | 0  | 0.9171          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
|             | +    | +      | +      | +  | +  | +  | I  | +  | +  | 0.5729   | 0    | 0      | 0      | 0  | 0  | 0  | 0  | 0  | 0  | 0.9597          | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| Positives   | 23   | 29     | 30     | 29 | 27 | 24 | 19 | 27 | 30 |          | 2    | ×      | С      | 14 | 6  | 9  | S  | 13 | ×  |                 | 0   | 0    | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| % Positives | 88.2 | % (238 | 8/270) |    |    |    |    |    |    |          | 26.3 | 1% (71 | (/270) |    |    |    |    |    |    |                 | 0%  |      |    |    |    |    |    |    |    |
|             |      |        |        |    |    |    |    |    |    |          |      |        |        |    |    |    |    |    |    |                 |     |      |    |    |    |    |    |    | l  |

ULNSD mean was below 0.35, the proportions of positive results were 75.7% and 81.5% in BEAST and MCMCTREE, respectively. The proportion of positive results fell to 39.2% and 17.1% for ULNSD values between 0.35 and 0.46. Finally, simulated ULNhig data sets produced a very low proportion of positive results (2.6% in BEAST, 0% in MCMCTREE), with a high proportion of type  $\sigma$  uninformative results (97.0% in BEAST and 99.3% in MCMCTREE). Following the same previous trend, positive results were only obtained in cases where the estimated mean of ULNSD was relatively low (<0.59).

Test stringency had an effect on the percentage of recovery of positive results but depending on analytical conditions and the method of choice. Lower lnBF thresholds yielded a lower proportion of positive results, whereas increasing the threshold above lnBF > 1.5 resulted already in 99% of positive results when an SC was applied for both BEAST and MCMCTree analyses (Table 4). Increased clock relaxation progressively reduced the threshold maximizing the proportion of positive results, from lnBF = 1.0 (ULNSD < 0.35) to 0.3 (ULNSD > 0.46) in the case of BEAST analyses and  $\ln BF = 0.7$  to 0.3 in those using MCMCTree, the latter approach being more sensitive to small differences in likelihood values to discriminate among incongruent hypotheses (Table 4). The best result in this series of analyses using ULNmod relaxed clocks was 89% of positive results with a threshold  $\ln BF = 0.7$  in MCMCTree analyses for low ULNSD values (Table 4). For ULNSD > 0.46, BFCA failed to produce positive results nearly always with  $\ln BF \ge 1$ (6%), but it reached a percentage of correct answers above

50% for a threshold  $0.3 < \ln BF < 0.4$  (Table 4). Finally, increasing the acceptance threshold had a clear effect reducing the proportion of type  $\rho$  negative results, which fell below 5% depending on analytical conditions (e.g.  $\ln BF < 1.0$  for SC under BEAST).

### BAYES FACTOR CLUSTER ANALYSIS FOR ND5 DATA IN CARABUS

Phylogenetic analyses of the 904 nt nd5 matrix of Carabus and using different phylogenetic inference methods resulted in very similar topologies, with the only difference found at the deepest nodes, which were not resolved under MP. However, even in this case, no topological incongruence was observed for highly supported nodes (Fig. 3). Nodes used for calibration received posterior probabilities between 0.92 and 1.00. Only the split represented by J4 received low bootstrap support and moderate posterior probability (0.71), but this split between Isiocarabus and Ohomopterus was confirmed by phylogenetic analyses using multiple loci (Sota & Ishikawa 2004). The BF comparison of Bayesian analyses for six different partitioning schemes and strict vs. relaxed clock models identified the 2P data partitioning scheme and SC model as optimal for nd5 data in Carabus (Table 5). Under this partitioning scheme, relaxed clock analyses yielded an ucld.stdev mean value = 0.089 and a truncate distribution abutting 0 (95% HPD interval  $3.65 \times 10^{-6}$ -0.20), which indeed suggests a negligible bias against the clock-like behaviour of data.

**Table 4.** BFCA results obtained for simulated data using BEAST and MCMCTREE and different acceptance threshold values. The results are shown as (x/y/z) representing the percentages of positive (only congruent calibration points are included in the BFCA solution; x), of negative type  $\rho$  results (some congruent calibration is excluded from the BFCA solution; y) and uninformative type  $\sigma$  results (some incongruent calibration is not excluded from the BFCA solution; z), respectively

|           | BEAST    |                 |                        |                 | MCMCTree | 2               |                     |                |
|-----------|----------|-----------------|------------------------|-----------------|----------|-----------------|---------------------|----------------|
| Threshold | SC       | ULNSD<br>< 0.35 | 0·35 < ULNSD<br>< 0·46 | ULNSD<br>> 0.46 | SC       | ULNSD<br>< 0.35 | 0·35 < ULNSD < 0·46 | ULNSD > $0.46$ |
| 0.1       | 7/93/0   | 4/96/0          | 11/89/0                | 11/89/0         | 1/99/0   | 6/94/0          | 18/82/0             | 12/88/0        |
| 0.2       | 27/73/0  | 14/86/0         | 24/74/1                | 43/54/2         | 10/90/0  | 15/85/0         | 45/50/4             | 31/62/7        |
| 0.3       | 48/52/0  | 24/76/0         | 48/44/8                | 63/24/12*       | 20/80/0  | 28/70/2         | 67/25/9             | 54/23/23*      |
| 0.4       | 60/40/0  | 34/66/0         | 68/22/10               | 60/6/34         | 30/70/0  | 57/39/4         | 69/11/20*           | 51/5/44†       |
| 0.5       | 75/25/0  | 50/50/0         | 77/9/14*               | 49/1/50†        | 41/59/0  | 67/28/6         | 61/3/37†            | 42/0/58        |
| 0.6       | 84/16/0  | 62/36/2         | 71/7/22                | 31/1/68         | 49/51/0  | 78/17/6         | 50/2/49             | 26/0/74        |
| 0.7       | 90/10/0  | 70/22/8         | 67/2/31†               | 22/0/78         | 60/40/0  | 89/6/6*         | 43/1/56             | 17/0/83        |
| 0.8       | 90/10/0  | 73/18/9         | 59/1/40                | 16/0/84         | 69/31/0  | 87/4/9†         | 38/1/61             | 12/0/88        |
| 0.9       | 94/6/0   | 74/13/12        | 47/0/53                | 11/0/89         | 79/21/0  | 83/4/13         | 27/1/72             | 8/0/92         |
| 1         | 95/5/0†  | 76/12/12*       | 40/0/60                | 6/0/94          | 88/12/0  | 81/2/17         | 18/0/82             | 6/0/94         |
| 1.1       | 97/3/0   | 74/11/14        | 34/0/66                | 2/0/98          | 92/8/0   | 80/2/19         | 11/0/89             | 4/0/96         |
| 1.2       | 97/3/0   | 73/9/18         | 29/0/71                | 1/0/99          | 93/7/0   | 74/0/26         | 6/0/94              | 2/0/98         |
| 1.3       | 99/1/0   | 74/7/19         | 27/0/73                | 1/0/99          | 95/5/0†  | 72/0/28         | 3/0/97              | 2/0/98         |
| 1.4       | 99/1/0   | 73/6/21         | 24/0/76                | 1/0/99          | 98/2/0   | 69/0/31         | 2/0/98              | 1/0/99         |
| 1.5       | 99/1/0   | 73/4/22†        | 21/0/79                | 1/0/99          | 99/1/0*  | 65/0/35         | 1/0/99              | 1/0/99         |
| 1.6       | 100/0/0* | 69/4/27         | 17/0/83                | 1/0/99          | 99/1/0   | 57/0/43         | 1/0/99              | 1/0/99         |
| 1.7       | 100/0/0  | 66/4/30         | 14/0/86                | 1/0/99          | 99/1/0   | 50/0/50         | 0/0/100             | 0/0/100        |
| 1.8       | 100/0/0  | 66/4/30         | 11/0/89                | 1/0/99          | 99/1/0   | 48/0/52         | 0/0/100             | 0/0/100        |
| 1.9       | 100/0/0  | 66/3/31         | 8/0/92                 | 1/0/99          | 99/1/0   | 48/0/52         | 0/0/100             | 0/0/100        |
| 2         | 100/0/0  | 64/3/32         | 7/0/93                 | 0/0/100         | 99/1/0   | 48/0/52         | 0/0/100             | 0/0/100        |

\*Threshold maximizing the number of positive results.

<sup>†</sup>Threshold with type  $\rho$  negative results  $\leq 5\%$  (low probability of rejecting a congruent prior).



Fig. 3. Ultrametric tree obtained with BEAST for the *nd5 Carabus* data set with the set of calibration hypothesis C, F, J1, J2, J3, J4b, M2 and M3, selected using BFCA. Numbers on nodes indicate support with Bayesian/ML/MP analyses, respectively. Bars correspond to the 95% Highest Posterior Density intervals for node ages.

 Table 5. Bayes factor comparisons for selection of clock model and partitioning scheme based on BEAST analyses for the nd5 gene in the genus

 Carabus

| Partition/clock scheme | р  | Marginal likelihood | NP/SC  | NP/ULN | 2P/SC | 2P/ULN | 3P/SC | 3P/ULN |
|------------------------|----|---------------------|--------|--------|-------|--------|-------|--------|
| NP/SC                  | 12 | -10058-19           | _      | 1.53*  | 35.25 | 32.19  | 19.80 | 18.95  |
| NP/ULN                 | 13 | -10056.65           | 1.53†  | _      | 39.00 | 35.25  | 20.84 | 19.82  |
| 2P/SC                  | 22 | -9705.69            | 352.49 | 350.96 | _     | 1.55   | 4.34  | 4.12   |
| 2P/ULN                 | 23 | -9704.15            | 354.04 | 352.50 | 1.55  | _      | 4.65  | 4.38   |
| 3P/SC                  | 32 | -9662.25            | 395.93 | 394.40 | 43.44 | 41.89  | _     | 1.93   |
| 3P/ULN                 | 33 | -9660.33            | 397.86 | 396.33 | 45.37 | 43.82  | 1.93  | _      |

p, Total number of free parameters required for each model and partitioning scheme; SC, strict clock; ULN, uncorrelated log-normal clock; NP, no codon partitioning; 2P, two partitions considering first and second codon positions together; 3P, three partitions considering each codon position as a different partition.

\*Above the diagonal:  $\ln BF/\Delta p$  (where  $\Delta p$ : difference in total number of free parameters between two models).

†Below the diagonal: lnBF.

|   | 2        |
|---|----------|
|   | 2        |
|   | 2        |
|   | 4        |
|   | 5        |
|   |          |
|   | _        |
|   | /        |
|   |          |
|   | Q        |
|   | 2        |
|   |          |
|   | 1        |
|   | 2        |
|   | 2        |
|   | 3        |
|   | 4        |
|   | 5        |
|   |          |
|   | 7        |
|   | /        |
|   |          |
|   | 9        |
|   | 0        |
|   |          |
|   | 1        |
|   | 2        |
|   | 2        |
|   | 5        |
|   | 4        |
|   | 5        |
|   | 6        |
|   | 7        |
|   | /        |
|   |          |
|   | 9        |
|   | 0        |
|   |          |
|   | 1        |
|   | 2        |
|   | 2        |
|   | 5        |
|   | 4        |
|   | 5        |
|   | 6        |
|   | 7        |
|   | /        |
|   |          |
|   | 9        |
|   |          |
|   | 4        |
|   |          |
|   | 2        |
|   | 3        |
|   | 1        |
|   | 4        |
|   | 5        |
|   | 6        |
|   | 7        |
|   | <i>′</i> |
|   |          |
|   | 9        |
| ; |          |
|   | -        |
| ) | 1        |
| ) | 2        |
| ) | 3        |
|   | Δ        |
|   | -r       |
|   |          |
| ) | 5        |
| ) | 5<br>6   |

Table 6. Values of marginal likelihood (above diagonal) and lnBF comparisons (below diagonal) relative to the best marginal-likelihood value for each pair of calibration hypotheses proposed for the *Carabus nd5* The pair of calibration hypothesis that results in the best marginal-likelihood value is underlined. Codes for calibration hypotheses are as in the main text. Hypotheses included in the most inclusive set of

| congru  | ent calibra | ations selecte | d by BFCA a | re marked wi | th an asterisk |          |          |          |          |          |          |          |             |          |          | _        |
|---------|-------------|----------------|-------------|--------------|----------------|----------|----------|----------|----------|----------|----------|----------|-------------|----------|----------|----------|
|         | C*          | F*             | J1*         | J1b          | J2*            | J2b      | J3*      | J3b      | J4       | J4b*     | MI       | M2*      | <u>M3</u> * | M4       | M5       | NZ       |
| Č       |             | -9706.54       | -9706-46    | -9714.49     | -9706.38       | -9721.21 | -9706.58 | -9715-52 | -9706.56 | -9706-55 | -9706.18 | -9706.00 | -9706.25    | -9707.04 | -9707.16 | -9713.11 |
| *<br>Ц  | 0.555       |                | -9706-67    | -9706-76     | -9706.75       | -9707.35 | -9706.82 | -9707.07 | -9706.80 | -9706.52 | -9706.18 | -9706.54 | -9706.61    | -9706.82 | -9707.29 | -9706.14 |
| $J1^*$  | 0.482       | 0.684          |             | n/a          | -9706.80       | -9718.68 | -9706.49 | -9712.99 | -9706.78 | -9706.13 | -9706.69 | -9706.59 | -9706.56    | -9707.24 | -9707.67 | -9710.55 |
| JIb     | 8.513       | 0.781          | n/a         |              | -9707.44       | -9707.52 | -9712.10 | -9706.24 | -9743.10 | -9722.17 | -9725.13 | -9717.94 | -9717.05    | -9709.42 | -9708.64 | -9706.78 |
| J2*     | 0.394       | 0.769          | 0.815       | 1.456        |                | n/a      | -9706.94 | -9707.65 | -9706.57 | -9706.59 | -9706.90 | -9706.42 | -9706.59    | -9706.20 | -9706.12 | -9706.24 |
| J2b     | 15.232      | 1.366          | 12.698      | 1.543        | n/a            |          | -9718.56 | -9707.66 | -9751.53 | -9730.02 | -9731.15 | -9723.55 | -9722.59    | -9714.10 | -9712.97 | -9711.03 |
| J3*     | 0.598       | 0.840          | 0.507       | 6.123        | 0.961          | 12.583   |          | n/a      | -9706.87 | -9706.26 | -9706.60 | -9706.52 | -9706.19    | -9706.70 | -9707.11 | -9710.04 |
| J3b     | 9.539       | 1.090          | 7.014       | 0.258        | 1.667          | 1.677    | n/a      |          | -9731.92 | -9725.47 | -9724.91 | -9718.06 | -9717.13    | -9709.66 | -9709.10 | -9706.88 |
| J4      | 0.578       | 0.818          | 0.794       | 37.120       | 0.590          | 45.552   | 0.887    | 25.944   |          | n/a      | -9706.68 | -9710.23 | -9711.72    | -9721.03 | -9725-66 | -9744.05 |
| J4b*    | 0.568       | 0.537          | 0.151       | 16.189       | 0.611          | 24.038   | 0.276    | 19.492   | n/a      |          | -9708.76 | -9706.22 | -9705.98    | -9708.04 | -9708.95 | -9721.39 |
| MI      | 0.196       | 0.199          | 0.710       | 19.152       | 0.916          | 25.168   | 0.622    | 18.925   | 0.697    | 2.780    |          | -9707.62 | -9708.73    | -9714.79 | -9716.39 | -9722.51 |
| M2*     | 0.015       | 0.560          | 0.605       | 11.963       | 0.441          | 17.574   | 0.534    | 12.080   | 4.248    | 0.238    | 1.642    |          | -9706.13    | -9709.49 | -9711.27 | -9716-29 |
| M3*     | 0.266       | 0.631          | 0.580       | 11.070       | 0.607          | 16.607   | 0.211    | 11.149   | 5.741    | 0.000    | 2.748    | 0.148    |             | -9708.61 | -9709.83 | -9715.02 |
| $M_{4}$ | 1.058       | 0.841          | 1.262       | 3.443        | 0.222          | 8·120    | 0.721    | 3.676    | 15.049   | 2.055    | 8.812    | 3.505    | 2.633       |          | -9706.08 | -9707.45 |
| M5      | 1.177       | 1.311          | 1.689       | 2.657        | 0.136          | 6.985    | 1.132    | 3.120    | 19.676   | 2.971    | 10.409   | 5.291    | 3.850       | 0.102    |          | -9706.84 |
| NZ      | 7.134       | 0.160          | 4.571       | 0.804        | 0.261          | 5.045    | 4.055    | 0.899    | 38-071   | 15-411   | 16-533   | 10.308   | 9.036       | 1.465    | 0.855    |          |

Thus, the BFCA procedure was applied to nd5 data under 2P/SC and considering the set of 16 potential node calibration hypotheses, including alternative age constraints for nodes related to the origin of the Japanese fauna. Table 6 summarizes the marginal-likelihood values for each of the resulting 116 pairwise combinations of calibration hypotheses and their InBF value relative to the optimal result. BFCA identified a single group of eight calibration constraints, namely C, F, J1, J2, J3, J4b, M2, M3, which consistently showed lnBF values of pairwise analyses below the threshold  $\ln BF > 1$ . Indeed, any combination of nine or more calibration points (or any other combination of eight hypotheses) included pairwise analyses with  $\ln BF > 2.6$ , strongly arguing against combinability (Kass & Raftery 1995). Thus, this group of eight dating node constraints was selected as the most inclusive set of concordant calibration hypotheses.

The calibrated Bayesian nd5 phylogeny of Carabus using the set of calibrations selected by BFCA resulted in a mean evolutionary rate for nd5 of 0.0154 (95% HPD: 0.0112-0.0198) substitutions per site per Ma per lineage (subs./s./Ma/l.), equivalent to 3.08% divergence between two lineages per Ma, and a time for the most recent common ancestor (TMRCA) of Carabus and Calosoma of 36.2 Ma (95% HPD: 27.64-45.72). For comparative purposes, the evolutionary rate and age were also estimated with the simultaneous use of all calibration constraints using alternative Pliocene or Miocene scenarios for Japanese taxa. The estimated mean rates were remarkably lower than those obtained based on the BFCA selection of calibration hypotheses: 0.0129 (95% HPD: 0.0095-0.0162) subs./ s./Ma/l. or 0.0073 (95% HPD: 0.0056-0.0091) subs./s./Ma/l., respectively, and the TMRCA of Carabus and Calosoma were remarkably higher: 77.7 Ma (95% HPD: 71.94-83.65) and 84.7 Ma (95% HPD: 76.22-94.96), respectively. Calibration hypotheses individually analysed produced rate estimates ranging from 0.0021 (95% HPD: 0.0012-0.0032; constraint J2b) subs./s./Ma/l. to 0.8254 (95% HPD: 0.1602-1.4033; constraint C) subs./s./Ma/l and TMRCA of Carabus and Calosoma ranging from 268.8 (95% HPD: 150.60-399.27; constraint J2b) to 0.9 Ma (95% HPD: 0.33-2.12; constraint C) (Table S11, Supporting information).

#### Discussion

#### IMPORTANCE OF USING RELIABLE CALIBRATIONS

A major advance in the process of phylogenetic dating has been to include calibration data as probability distributions that account for the different sources of uncertainty in the age of calibrations (Inoue, Donoghue & Yang 2010). Yet, the selection and appropriate use of calibrations has been reported as probably the most persistent problem in dating analyses (Heads 2005) with a major effect on estimated ages (Inoue, Donoghue & Yang 2010). Because of potential bias of using single calibration points and favoured by the development of relaxed clock models, the simultaneous use of all information available to calibrate a tree has been recommended as a way to minimize these problems (Yang & Rannala 2006; Ho &

Phillips 2009). However, the indiscriminate application of all available calibration data can lead to biased consensus solutions in which incongruence is accommodated at the cost of inferring highly artificial levels of rate variation, producing both incorrect local and global rates even when relaxed clock approaches are used. Although some recent work indicates that the incorporation of incorrect calibrations has a limited effect when relaxed clocks are used in combination with soft bounds distributions (Yang & Rannala 2006), systematic studies assessing the impact of the incongruent calibrations are still lacking. However, it is plausible that the inclusion of extremely incongruent calibrations will have an effect on the estimation of divergence times and substitution rates, especially when taxonomic sampling and genetic data are limited.

Thus, calibration analyses could, in principle, be improved by filtering the data to identify sets of calibration points that are consistent with one another, addressing possible dating errors by mutual corroboration of independent data. Thus, our main contribution to the calibration problem through BFCA is the development of a heuristic test that enables the identification of these sets of congruent calibrations taking into account the uncertainty associated with each one of them, which is incorporated with a particular probability density function.

#### BAYES FACTOR CLUSTER ANALYSIS PERFORMANCE

We have explored the effects on marginal-likelihood scores of the simultaneous application of pairs of calibrations with varying degrees of conflict under different scenarios of clock relaxation. Our main observation is that, as predicted, conflicting calibrations affecting relative branch lengths indeed produce a progressive decrease on marginal likelihood in comparison with unconstrained (or optimally constrained) analyses, that is, marginal likelihood is negatively correlated with the magnitude of the conflict.

Our simulations using pairs of calibration points showed some interesting trends relevant to understand BFCA. One such trend is that the sensitivity of the test depends on the relative position of the calibrated nodes. Thus, for pairs of calibrations represented by normal, exponential or narrow skew-normal functions, deeper nodes seem to admit wider departure from their real age before favouring one hypothesis over another. Furthermore, we also observed that the more separated the nodes are on the tree, the harder it is to detect the incongruence. Nonetheless, the most important trend is related with the use of relaxed vs. strict clock models. In this regard, we found that the discrimination power of BFCA is highly dependent on the level of among-branch rate variation (i.e. the value of ULNSD). When moderate levels of rate variation were allowed (ULNSD = 0.3), the calibration needed a departure from the real value almost three times higher than in the case of SC to yield values of lnBF > 1. High levels of rate variation (ULNSD = 0.6) worsened the trend, allowing the accommodation of any calibration in the range of incongruence examined. We demonstrated that calibration conflicts are accommodated by increasing the variation of evolutionary rates among branches when using relaxed clocks, and therefore, such variation is inversely correlated with the discrimination power of the BFCA.

We also evaluated the performance of BFCA through simulation under different clock models and calibration scenarios. We distinguished two situations to describe BFCA departures from the expected result: negative (type  $\rho$ ) and uninformative (type  $\sigma$ ) results. In the first case, correct calibrations are recognized as conflictive, whereas in the second case, incongruent calibrations are not recognized as such and are included in the BFCA solution along with the correct calibrations. As expected from our simulations using pairs of calibrations, the clock model is the main factor conditioning the power of the method. Thus, BFCA performance is very high for data sets simulated under a SC, identifying the right set of congruent calibration hypotheses in 95.0% and 89.2% of cases in our tests with BEAST and MCMCTree, respectively (threshold of  $\ln BF > 1$ ). For simulations with moderate rate variation, the proportion of positive results is maximized with lnBF thresholds between 0.3 and 1, depending on the estimated ULNSD of the data and the program used. The optimum threshold for the data sets showing ULNSD < 0.35 allows detecting the correct set of calibrations in 76% of cases in BEAST and 89% of cases in MCMCTree, while maintaining a relatively low proportion of negative (type  $\rho$ ) results (12% and 2% of cases, respectively). These observations highlight some of the limitations of BFCA when dealing with alignments that show high variation in substitution rates among taxa. The power of the BFCA is low under these conditions because the high rate variation among the branches of the tree together with the limited information of the sequences (1000 bp in the simulated analyses) allow the accommodation of incongruent calibrations with a negligible cost in likelihood. Ongoing investigations show that this undesirable effect is highly dependent on the length of the sequences, in agreement with the findings of Yang & Rannala (2003). These preliminary analyses indicate that incongruent **3** calibrations have a higher effect on the likelihood of analyses when using longer, more informative, sequences, and consequently providing a best performance of BFCA with relaxed clocks models (C. Andújar et al. unpublished data).

Expectedly, the proportion of positive results as deduced from simulated data depends on the lnBF threshold used (Table 4). In the case of our SC simulated data sets, increasing the threshold to 1.5 resulted in the recovery of 99% of correct answers, and therefore, thresholds between 1 and 1.5 (Kass & Raftery 1995) would be recommended for data sets under strict clock evolution and independently of the software used. In cases where a relaxed clock applies, there seems to be a compromise threshold between a situation where congruent priors are excluded with high probability (negative results) for low acceptance thresholds, and one where incongruent priors start to be included in the solution (uninformative results) for high acceptance thresholds. Thus, for simulated data sets with ULNSD > 0.35, the recovery of positive results with  $\ln BF = 1$ is very low, but it improves to 50–60% for  $0.3 < \ln BF < 0.5$ , while for ULNSD < 0.35, the situation does not seem to differ notably from the SC scenario except for a better performance

for slightly lower acceptance thresholds, in the range of  $\ln BF = 1.0$ . It might then be advisable to use  $\ln BF$  thresholds as low as  $\ln BF = 0.5$  for data sets with high among-branch rate variation. As indicated above, higher thresholds reduce negative and increase uninformative results. The proportion of negative results can be interpreted as the assumed risk to exclude a correct calibration, and, consequently, higher thresholds will ensure the recovery of all the correct calibrations (but at the cost of increasing the probability of including incorrect ones). For example, in the case of ULNSD > 0.35 data sets, the threshold of 0.5 implies a proportion of negative results between 9% and 0% that needs to be taken into account. It is important to emphasize that the threshold customization as implemented in the R package bfca (threshold parameter) is an important feature that allows exploring BFCA results under different analytical conditions, an advisable procedure when dealing with complicated cases where among-branch rate variation is significant.

#### BAYES FACTOR CLUSTER ANALYSIS AND ND5 EVOLUTIONARY RATE IN CARABUS

The study of the evolutionary rate of the *nd5* gene in the genus *Carabus* illustrates some of the problems mentioned above and stresses the suitability of the BFCA to identify consistent calibration data. While the data fitted a clock-like evolution, available calibration hypotheses analysed individually produced evolutionary rate estimates ranging from 0.0021 to 0.8254 subs./s./Ma/l and TMRCA of *Carabus* and *Calosoma* ranging from 268.8 to 0.9 Ma (Table S11, Supporting information). Facing this range of potential results spanning two to three orders of magnitude, one recurrent solution has been averaging using all points (Yang & Rannala 2006; Ho & Phillips 2009), which would result in a somewhat intermediate value.

Among 16 available calibration hypotheses for *Carabus*, the use of BFCA identified eight as mutually congruent, including (i) the age F of the fossil *C. cancellatus*, (ii) J1, J2 and J3, placing the diversification of Japanese species of subgenera *Damaster, Leptocarabus* and *Ohomopterus* around 3.5 Ma, (iii) J4b, concordant with the Miocene split of mainland *Carabus* lineages and Japanese *Ohomopterus*, (iv) M2 and M3 corresponding to the split at the end of the Messinian of African and European species of *Eurycarabus* and subspecies of *Carabus melancholicus* and (v) C, assigning 14.5 Ma as the maximum age for the split of *Nesaeocarabus* within the Canary Islands. Fig. 3 shows an ultrametric tree calibrated simultaneously using the selected calibrations for these eight nodes.

Calibration hypotheses discarded by BFCA can be reinterpreted in the light of the dated phylogeny of the group. The allopatric distribution of European *Mesocarabus* lineages and North African *C. (Mesocarabus) riffensis* as well as the allopatric ranges of European and North African subspecies of both *C. (Macrothorax) morbillosus* and *C. (Macrothorax) rugosus* has been interpreted as the result of Messinian vicariance (e.g. Prüser & Mossakowski 1998). However, our results for *Mesocarabus* indicate that the split occurred earlier, in the Late Miocene (M1: ~10.5 Ma; 95% HPD: 7.6–13.5 Ma), and are consistent with a Betic-Riffian origin of the group (Andújar et al. 2012), or more recently in the case of Macrothorax, which started to diverge at the Plio-Pleistocene border (M4: 3 Ma; 95% HPD: 1.8-4.3 Ma; and M5: 2.5 Ma; 95% HPD: 1.6–3.6 Ma), well after the refilling of the Mediterranean Sea. Particularly for the latter, post-Messinian transmarine dispersal is not at odds with the natural history of these large beetle species. In contrast to most other Carabus species, C. morbillosus has relatively long and innervated wings (Ortuño & Hernández 1992), probably enhancing their dispersal power. Supporting the idea of recent dispersal, highly similar mtDNA haplotypes were found between Tunisian and Sardinian individuals of this species (Prüser & Mossakowski 1998). For faunas in remote continental islands of the Southern Hemisphere, it is tempting to invoke continental drift as the explanation for allopatric distributions, and this was the case for Australian Pamborus and Maoripamborus from New Zealand (Sota et al. 2005). Nonetheless, dating this event with an Upper Cretaceous age interval as modelled here is highly inconsistent with most other calibration points, as revealed by BFCA, and would dramatically reduce the evolutionary rate for nd5 to 0.0044 subs./s./Ma/l. (95% HPD: 0.0031-0.0059), slower than the slowest mtDNA protein-coding genes in Coleoptera (nad4L; Pons et al. 2010). The effect of sequence saturation and limitations of the evolutionary model in estimating the age of this node (older than others here studied) may affect BF estimations favouring the exclusion from the optimum solution. It is worth noting that the BFCA do not prevent for the caution that should be taken when estimating the ages of nodes beyond the limit imposed by the saturation of markers.

The evolutionary rate for *nd5* in *Carabus* estimated after BFCA testing, 0.0154 (95% HPD: 0.0112–0.0198) subs./s./ Ma/l., is higher than values previously reported for this or other mtDNA genes of *Carabus*, but remarkably similar to the estimated rate for this gene in Coleoptera (0.0168 subs./s./Ma/ 1.; 95% HPD: 0.0086–0.0279; Pons *et al.* 2010). The discrepancies with previous attempts can be explained because of the use of flawed calibration points (e.g. tectonic vicariance for *Maoripamborus*; Sota *et al.* 2005), by constraints applied to wrong nodes (e.g. Japanese radiations of *Carabus* defined by their sister relationship with continental relatives; Su *et al.* 1998; Tominaga *et al.* 2000; Osawa, Su & Imura 2004), or by using inappropriate corrections of genetic distances among species (e.g. underestimated *nd1* divergences for *Rhabdotocarabus*; Prüser & Mossakowski 1998).

### BFCA AS A TEST FOR SELECTING CONGRUENT CALIBRATIONS

Our BFCA approach proved to be a useful tool to objectively select concordant calibration constraints, with evidence supplied by BF comparisons. BFCA exploits the analytical strengths of Bayesian methodologies allowing the application of relaxed clock models and the incorporation of both fossil and geologic data in the form of calibrations defined by probability density functions. Thus, BFCA overcomes the limitations of methods that rely on point calibrations, such as the

cross-validation fossil method of Near, Meylan and Shaffer (2005) or Marshall's (2008) method to deal with taphonomic biases on the fossil record. There are other parametric approaches, like the Bayesian extension of Marshall's method by Dornburg et al. (2011), based on the selection of the best fossil to compare with the complete pool of calibrations. However, these methods are limited to calibration exercises based on fossils. This limitation does not condition BFCA given that parametric prior distributions can represent any source of dating information, from specific fossils to a geographical event relevant in phylogenetic terms (e.g. vicariance). There are yet other methods that use probabilistic distributions as priors and evaluate the accuracy of a proposed calibration with respect to others, which are assumed to be reliable (Sanders & Lee 2007; Pyron 2010). Unfortunately, information about the reliability of calibration hypotheses, the crux of tree calibration practice, is often lacking. Instead, BFCA has the advantage that it operates on the assumption that reliable independent calibrations must be necessarily congruent among them and will tend to outnumber erroneous ones, especially if they have been selected following rigorous procedures (Parham et al. 2011). BFCA can be applied with both strict and relaxed clock models, and it is able to identify incongruent calibration hypotheses even in cases with moderate rate variation among branches. Besides, the applicability of BFCA with relaxed clocks can take advantage of the possibility of including priors of among-branch rate variation based on information from previous analyses of similar molecular markers, restricting the artefact overestimation in rate variation required to accommodate incongruence among calibrations. Lastly, it is worth noting that, although we centred our analyses in the use of certain models implemented in BEAST and MCMCTree, the method is equally applicable to the results produced by any other Bayesian phylogenetic inference or dating method, as long as they are amenable of including calibration information in some way.

One practical drawback of BFCA is that it is computationally demanding because of the number of independent analyses required. However, the application of BFCA remains feasible even for problems involving many calibration points as long as parallel computing using small high-performance clusters is available. This can be achieved because the method is based on pairwise analyses instead of carrying out BF comparisons among all possible combinations of hypotheses, which quickly becomes computationally intractable. For example, the analysis of 16 calibration hypotheses in the case of *Carabus* required 120 independent BEAST runs with the BFCA approach, in contrast to the 65 519 runs that would be needed when considering all possible combinations of calibrations.

It should be noted that the BFCA is a congruence-based method that should be used with three or more calibrations, and evidence for congruence is only admitted if at least three calibrations are recovered as congruent. Another aspect that may potentially affect the BFCA is the way in which calibrations are incorporated by the Bayesian software as node prior ages into the analyses. For example, it is known that the interaction among the calibration age and the tree prior may result in effective joint priors different from the defined calibration densities (Heled & Drummond 2012). How this may affect the reliability of calibration analyses and the evaluation of calibration hypotheses with the BFCA approach should be addressed in the future. Lastly, accuracy in the estimation of marginal likelihoods could potentially have an influence in our methodology. Harmonic estimators such as the variant used here (Suchard, Weiss & Sinsheimer 2001) have been shown to lead to erroneous results when used to compare between evolutionary models which differ in the number of their parameters (Lartillot & Philippe 2006; Baele et al. 2012a,b). However, BFCA results are based on the comparison of analyses where the number of parameters remains constant, on the evaluation of identical models with different calibration priors, and therefore, this bias is not known to affect our approach. Particularly in the case of the simulations used here, HME estimates always reached good convergence, as reflected also in the negligible differences between pairs of runs for the same analysis (Fig. S1, Supporting information), and very convincingly in the clear trends observed in the simulations related to our Proof-of-principle. In any case, it will be of particular interest to further investigate the behaviour of the BFCA by using more accurate methods for the estimation of marginal likelihoods, such as thermodynamic integration (Lartillot & Philippe 2006) or stepping-stone sampling (Xie et al. 2011).

Bayes Factor Cluster Analysis addresses several of the limitations for selection of calibration hypotheses and clock calibration. However, it is important to keep in mind that results of BFCA are much dependent on the quality of calibration hypotheses and decisions on prior modelling for the analyses, which rely on previous choices independent of the BFCA method. For example, uncertain calibrations modelled with wide prior distributions, such as the broad log-normal distribution that we used, have little effect on the shape of the tree and have, therefore, high chances of being included in the final BFCA solution. Moreover, a biased selection of incorrect calibration constraints (e.g. due to taphonomic bias in the case of fossils) or their systematic incorrect placement in a phylogeny will produce erroneous divergence times and evolutionary rate estimations, independently of the method employed for filtering hypotheses (Parham et al. 2011). An example of the latter situation is shown by the effects of alternative ages applied to putative vicariance events in the case of the geologically complex history of the Japanese archipelago for the genus Carabus. Other important factors that will also require further scrutiny are the effect of saturation due to multiple substitutions, topological relationships among taxa, or the influence of different proportions of congruent and incongruent calibrations, but they are all amenable to systematic study using our new analytical BFCA tool and software.

#### Acknowledgements

This work was supported by the Spanish Ministry of Science and Innovation (CGL2006/06706 and CGL2009-10906 to C.A. and J.S. and CGL2008-00007 to J.G.-Z., the latter also with support of the European Regional Development Fund). CA received support from an FPU predoctoral studentship (Spanish Ministry of Education). Anna Papadopoulou and Ignacio Ribera (Institute of Evolutionary Biology, Barcelona, CSIC) discussed with us many of the ideas

developed here, read a preliminary version of this work and provided with very constructive criticism to significantly improve the study. Carlos Ruiz and Paula Arribas (University of Murcia) also helped with their comments and advice. Achille Casale (University of Sassari) helped with the identification of some *Carabus* taxa. Thanks are due to Obdulia Sánchez, Ana Asensio, José Luis Lencina (University of Murcia) and Gwenaelle Genson (CBGP Montpellier) for their technical assistance. Most of computational analyses were performed using the Ben Arabi supercomputer of the Fundación Parque Científico de Murcia (Murcia, Spain) and the Iceberg HPC cluster of the University of Sheffield (Sheffield, UK).

#### References

- Aguileta, G., Bielawski, J.P. & Yang, Z. (2006) Evolutionary rate variation among vertebrate [beta] globin genes: implications for dating gene family duplication events. *Gene*, **380**, 21–29.
- Andújar, C., Serrano, J. & Gómez-Zurita, J. (2012) Winding up the molecular clock in the genus *Carabus* (Coleoptera: Carabidae): assessment of methodological decisions on rate and node age estimation. *BMC Evolutionary Biology*, 12, 40.
- Andújar, C., Gómez-Zurita, J., Rasplus, J.Y. & Serrano, J. (2012) Molecular systematic and evolution of the subgenus *Mesocarabus* Thomson 1875 (Coleoptera, Carabidae, *Carabus*) based on mitochondrial and nuclear DNA. *Zoological Journal of the Linnean Society*, **166**, 787–804.
- Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M.A. & Alekseyenko, A.V. (2012a) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, 29, 2157–2167.
- Baele, G., Li, W.L.S., Drummond, A.J., Suchard, M.A. & Lemey, P. (2012b) Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Molecular Biology and Evolution*, 30, 239–243.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Sayers, E.W. (2010) GenBank. Nucleic Acids Research, 38, 46–51.
- Benton, M.J. & Ayala, F.J. (2003) Dating the tree of life. Science, 300, 1698–1700. Bousquet, J., Strauss, S.H., Doerksen, A.H. & Price, R.A. (1992) Extensive variation in evolutionary rate of rbcL gene sequences among seed plants. Proceedines of the National Academy of Sciences USA. 89, 7844–7848.
- Bromham, L. (2009) Why do species vary in their rate of molecular evolution? *Biology Letter*, 5, 401–404.
- Bromham, L. & Penny, D. (2003) The modern molecular clock. *Nature Reviews Genetics*, 4, 216–224.
- Brown, J.M. & Lemmon, A.R. (2007) The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Systematic Biology*, 56, 643–655.
- Burbrink, F.T. & Lawson, R. (2007) How and when did old world ratsnakes disperse into the New World? *Molecular Phylogenetics and Evolution*, 43, 173– 189.
- Clarke, J.Y., Warnock, R. & Donoghue, P. (2011) Establishing a time-scale for plant evolution. *New Phytologist*, **192**, 266–301.
- Conroy, C.J. & van Tuinen, M. (2003) Extracting time from phylogenies: positive interplay between fossil and genetic data. *Journal of Mammalogy*, 84, 444–455.
- Dornburg, A., Beaulieu, J.M., Oliver, J.C. & Near, T.J. (2011) Integrating fossil preservation biases in the selection of calibrations for molecular divergence time estimation. *Systematic Biology*, 60, 519–527.
- Drummond, A.J. & Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology, 7, 214.
- Drummond, A.J., Ho, S.Y.W., Phillips, M.J. & Rambaut, A. (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4, 699–710.
- Felsenstein, J. (1981) Evolutionary trees from DNA sequences: a maximum-likelihood approach. Journal of Molecular Evolution, 17, 368–376.
- Gómez-Zurita, J. (2004) Molecular systematics and time-scale for the evolution of *Timarcha*, a leaf-beetle genus with a disjunct Holarctic distribution. *Molecular Phylogenetics and Evolution*, **32**, 647–665.
- Graur, D. & Martin, W. (2004) Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends in Genetics*, 20, 80–86.
- Heads, M. (2005) Dating nodes on molecular phylogenies: a critique of molecular biogeography. *Cladistics*, 21, 62–78.
- Heads, M. (2010) Evolution and biogeography of primates: a new model based on molecular phylogenetics, vicariance and plate tectonics. *Zoologica Scripta*, **39**, 107–127.
- Hedges, S.B. & Kumar, S. (2003) Genomic clocks and evolutionary timescales. *Trends in Genetics*, 19, 200–206.
- Heled, J. & Drummond, A.J. (2012) Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Systematic Biology*, **61**, 138–149.

Ho, S.Y.W. (2005). RateEvolver Version 1.1. University of Oxford, ????, UK.

- Ho, S.Y.W. & Phillips, M.J. (2009) Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology*, 58, 367–380.
- Huelsenbeck, J.P., Hillis, D.M. & Nielsen, R. (1996) A likelihood-ratio test of monophyly. *Systematic Biology*, 45, 546–558.
- Inoue, J., Donoghue, P. & Yang, Z. (2010) The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Systematic Biology*, **59**, 74–89.
- Kass, R.E. & Raftery, A.E. (1995) Bayes factors. Journal of the American Statistical Association, 90, 773–795.
- Lartillot, N. & Philippe, H. (2006) Computing Bayes factors using thermodynamic integration. Systematic Biology, 55, 195–207.
- Marshall, C.R. (2008) A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibrations points. *American Naturalist*, **171**, 726–742.
- Miller, K.B., Bergsten, J. & Whiting, M.F. (2009) Phylogeny and classification of the tribe Hydaticini (Coleoptera: Dytiscidae): partition choice for Bayesian analysis with multiple nuclear and mitochondrial protein-coding genes. *Zoo-logica Scripta*, 38, 591–615.
- Near, T.J., Meylan, P.A. & Shaffer, H.B. (2005) Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. *American Naturalist*, 165, 137–146.
- Near, T.J. & Sanderson, M.J. (2004) Assessing the quality of molecular divergence time estimates by fossil calibrations and fossil-based model selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359, 1477–1483.
- Newton, M. & Raftery, A. (1994) Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society: Series B*, **56**, 3–48.
- Noonan, B.P. & Chippindale, P.T. (2006) Vicariant origin of Malagasy reptiles supports late cretaceous Antarctic land bridge. *American Naturalist*, 168, 730–741.
- Ortuño, V.M. & Hernández, J.M. (1992) The hindwing in the Iberian Carabini (Coleoptera, Caraboidea). Boletim da Sociedade Portuguesa de Entomologia, 1, 33–42.
- Osawa, S., Su, Z.-H. & Imura, Y. (2004) Molecular Phylogeny and Evolution of Carabid Ground Beetles. Springer-Verlag, Tokyo.
- Pagel, M. & Meade, A. (2004) A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Systematic Biology*, 53, 571–581.
- Parham, J.F., Donoghue, P.C., Bell, C.J., Calway, T.D., Head, J.J., Holroyd, P.A. et al. (2011) Best practices for justifying fossil calibrations. Systematic Biology, 61, 346–359.
- Pons, J., Ribera, I., Bertranpetit, J. & Balke, M. (2010) Nucleotide substitution rates for the full set of mitochondrial protein-coding genes in Coleoptera. *Molecular Phylogenetics and Evolution*, **56**, 796–807.
- Posada, D. (2008) jModelTest: phylogenetic model averaging. *Molecular Biology* and Evolution, 25, 1253–1256.
- Posada, D. & Buckley, T.R. (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, 53, 793–808.
- Prüser, F. & Mossakowski, D. (1998) Low substitution rates in mitochondrial DNA in Mediterranean carabid beetles. *Insect Molecular Biology*, 7, 121–128.
- Pyron, R.A. (2010) A likelihood method for assessing molecular divergence time estimates and the placement of fossil calibrations. *Systematic Biology*, **59**, 185– 194.
- Rambaut, A. & Drummond, A.J. (2007) Tracer v1.5. Available from http://beast. bio.ed.ac.uk/Tracer
- Rambaut, A. & Grassly, N.C. (1997) Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Computer Applications in the Biosciences*, 13, 235–238.
- Sanders, K.L. & Lee, M.S.Y. (2007) Evaluating molecular clock calibrations using Bayesian analyses with soft and hard bounds. *Biology Letters*, 3, 275– 279.
- Sanderson, M.J. (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution*, 14, 1218–1231.
- Sanderson, M.J. (2003) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*, **19**, 301– 302.
- Schwartz, R.S. & Mueller, R.L. (2010) Variation in DNA substitution rates among lineages erroneously inferred from simulated clock-like data. *PLoS One*, 5, e9649.

- Smith, A.B. & Peterson, K.J. (2002) Dating the time of origin of major clades: molecular clocks and the fossil record. *Annual Review of Earth and Planetary Sciences*, 30, 65–88.
- Soltis, P.S., Soltis, D.E., Savolainen, V., Crane, P.R. & Barraclough, T.G. (2002) Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. *Proceedings of the National Academy of Sciences USA*, 99, 4430–4435.
- Sota, T. & Ishikawa, R. (2004) Phylogeny and life-history evolution in *Carabus* (subtribe Carabina: Coleoptera, Carabidae) based on sequences of two nuclear genes. *Biological Journal of the Linnean Society*, **81**, 135–149.
- Sota, T., Takami, Y., Monteith, G.B. & Moore, B.P. (2005) Phylogeny and character evolution of endemic Australian carabid beetles of the genus *Pamborus* based on mitochondrial and nuclear gene sequences. *Molecular Phylogenetics* and Evolution, 36, 391–404.
- Su, Z.H., Tominaga, O., Okamoto, M. & Osawa, S. (1998) Origin and diversification of hindwingless *Damaster* ground beetles within the Japanese islands as deduced from mitochondrial ND5 gene sequences (Coleoptera, Carabidae). *Molecular Biology and Evolution*, 15, 1026–1039.
- Suchard, M.A., Weiss, R.E. & Sinsheimer, J.S. (2001) Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular Biology and Evolution*, 18, 1001–1013.
- Sullivan, J. & Joyce, P. (2005) Model selection in phylogenetics. Annual Review of Ecology, Evolution, and Systematics, 36, 445–466.
- Thomas, J.A., Welch, J.J., Woolfit, M. & Bromham, L. (2006) There is no universal molecular clock for invertebrates, but rate variation does not scale with body size. *Proceedings of the National Academy of Sciences USA*, **103**, 7366– 7371.
- Thorne, J.L. & Kishino, H. (2002) Divergence time and evolutionary rate estimation with multilocus data. Systematic Biology, 51, 689–702.
- Tominaga, O., Su, Z.H., Kim, C.G., Okamoto, M., Imura, Y. & Osawa, S. (2000) Formation of the Japanese Carabina fauna inferred from a phylogenetic tree of mitochondrial ND5 gene sequences (Coleoptera, Carabidae). *Journal of Molecular Evolution*, **50**, 541–549.
- Wolfe, K.H., Sharp, P.M. & Li, W.H. (1989) Mutation-rates differ among regions of the mammalian genome. *Nature*, 337, 283–285.
- Xie, W., Lewis, P.O., Fan, Y., Kuo, L. & Chen, M.H. (2011) Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology*, **60**, 150–160.
- Yang, Z. & Rannala, B. (2006) Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution*, 23, 212–226.
- Zuckerkandl, E. & Pauling, L. (1965) Evolutionary divergence and convergence in proteins. *Evolving Genes and Proteins* (eds V. Bryson & H.J. Vogel), pp. 97–166. Academic Press, New York.

Received 12 June 2013; accepted 4 December 2013 Handling Editor: Robert Freckleton

#### **Supporting Information**

Additional Supporting Information may be found in the online version of this article.

**Figure S1.** Histogram with the lnBF values obtained for the comparisons of the two independent runs conducted on each of the 13 068 the simulation analyses in BEAST (a) and MCMCTTRE (b).

Figure S2. Results of the initial evaluation of incongruence limits in BEAST.

Figure S3. Results of the initial evaluation of incongruence limits in MCMCTTRE.

 
 Table S1. Prior distributions used in the initial evaluation of incongruence limits in BEAST.

 
 Table S2. Range of age values used for the initial evaluation of incongruence limits in BEAST.

 
 Table S3. Prior distributions used in the initial evaluation of incongruence limits in MCMCTree.

 
 Table S4. Range of age values used for the initial evaluation of incongruence limits in MCMCTree.

 Table S5. Data about parameters used in the simulated calibration analyses conducted in BEAST.

 
 Table S6. Data about parameters used in the simulated calibration analyses conducted in MCMCTree.

Table S7. Data about specimen and nd5 sequences employed to test the performance of BFCA on a real DNA data set in the genus *Carabus*.

 
 Table S8. Data about prior parameters used in the BFCA calibration analyses of the nd5 *Carabus* data set conducted in BEAST.

 Table S9. Results of the initial evaluation of incongruence limits in BEAST.

 Table S10. Results of the initial evaluation of incongruence limits in MCMCTTRE.

 Table S11. Evolutionary rates and tree root ages estimated for the Carabus nd5 data set.

Data S1. DNA extraction, purification and PCR reaction.

**Data S2.** Information on phylogenetic analyses conducted for the genus *Carabus* nd5 data set.

Data S3. Calibration hypotheses used for dating of genus Carabus.

### **Author Query Form**

| Journal: | MEE3  |
|----------|-------|
| Article: | 12151 |

Dear Author,

During the copy-editing of your paper, the following queries arose. Please respond to these by marking up your proofs with the necessary changes/additions. Please write your answers on the query sheet if there is insufficient space on the page proofs. Please write clearly and follow the conventions shown on the attached corrections sheet. If returning the proof by fax do not write too close to the paper's edge. Please remember that illegible mark-ups may delay publication.

Many thanks for your assistance.

| Query reference | Query   | Remarks |
|-----------------|---|---------|
| 1               | AUTHOR: Please provide significant of bold value given in Table 6.  |         |
| 2               | AUTHOR: Ho & Philips 2009 has been changed to Ho and Phillips 2009 so that this cita-<br>tion matches the Reference List. Please confirm that this is correct.  |         |
| 3               | AUTHOR: Yang & Rannala (2003) has not been included in the Reference List, please sup-<br>ply full publication details.   |         |
| 4               | AUTHOR: Please provide the city name for publisher for reference Ho (2005).   |         |
| 5               | AUTHOR: Please check this website address and confirm that it is correct. (Please note that it is the responsibility of the author(s) to ensure that all URLs given in this article are correct and useable.) |         |
| 6               | AUTHOR: Text S1, S2 and S3 have been changed to Data S1, S2 and S3. Please amend similar changes in the supporting information folder too.  |         |

#### USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION



| Once you have Acrobat Reader open on your computer, click   | c on the Comment tab at the right of the toolbar:   |
|---|---|
| This will open up a panel down the right side of the document tools you will use for annotating your proof will be in the Anno pictured opposite. We've picked out some of these tools below  | t. The majority of trations section, w:<br>$ \begin{array}{c c} \bullet & \bullet \\ \bullet &$ |
| <ol> <li>Replace (Ins) Tool – for replacing text.</li> <li>Strikes a line through text and opens up a text box where replacement text can be entered.</li> <li>How to use it</li> <li>Highlight a word or sentence.</li> <li>Click on the Replace (Ins) icon in the Annotations section.</li> <li>Type the replacement text into the blue box that appears</li> </ol> | <ul> <li>2. Strikethrough (Del) Tool – for deleting text.</li> <li>Strikes a red line through text that is to be deleted.</li> <li>How to use it <ul> <li>Highlight a word or sentence.</li> <li>Click on the Strikethrough (Del) icon in the Annotations section.</li> </ul> </li> </ul>   |
| ndard framework for the analysis of m<br>icy. Nevertheless, it also led to exoge<br>ble of strateg •• dthreshe<br>nber of comp<br>: is that the si<br>nain compo<br>level, are exe<br>important works on entry by on ire  | there is no room for extra profits at<br>c ups are zero and the number of<br>cet) values are not determined by<br>Blanchard and Kiyotaki (1987),<br>erfect competition in general equilil<br>its of aggregate demand and supply<br>classical framework assuming monop<br>cen an exogenous number of firms   |
| <ul> <li>3. Add note to text Tool – for highlighting a section to be changed to bold or italic.</li> <li>Image: Highlights text in yellow and opens up a text box where comments can be entered.</li> <li>How to use it</li> <li>Highlight the relevant section of text.</li> <li>Click on the Add note to text icon in the Annotations section.</li> </ul>           | <ul> <li>4. Add sticky note Tool – for making notes at specific points in the text.</li> <li>Marks a point in the proof where a comment needs to be highlighted.</li> <li>How to use it <ul> <li>Click on the Add sticky note icon in the Annotations section.</li> <li>Click at the point in the proof where the comment should be inserted</li> </ul> </li> </ul>   |
| <ul> <li>Type instruction on what should be changed regarding the text into the yellow box that appears.</li> <li>namic responses of mark ups cent with the VAR evidence</li> <li>sation</li> <li>sation</li> <li>og/od/2011 15:31:38</li> <li>inthe ell ed</li> </ul>  | Type the comment into the yellow box that<br>appears.     and and suppry shocks, tyrost of<br>a minetic suppry shocks, tyrost of<br>a dard fr:<br>.cy. Nev<br>ble of sti  |

to a on stent also with the demand-

ber

iber of competitors and the map

is that the structure of the secto

on n

#### USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

## WILEY

