Supplemental Material for Matzke and Wright (2016), “Inferring node dates from tip dates in fossil Canidae: the importance of tree priors”

Supplemental Introduction

Terminology. The methods we refer to as “tip-dating” methods are often called “total evidence” dating methods. As originally devised, these methods combined molecular and morphological data with tip-dates. However, in this study, we use a morphology-only dataset, but the models and methods are otherwise the same. So, we refer to the methods as “tip-dating” rather than “total evidence” in this paper.

Brief review of tip-dating studies. Major papers have introduced tip-dating methods and models [1-6]. A number of tip-dating studies have been published at the time of writing [4, 5, 7-23].

A review of this literature, while generally approving, shows that some studies consider some of their tip-dating results implausible (e.g. [9, 22, 24]), and some infer dates that are wildly uncertain [12, 17, 25]. Evaluation of the methods against each other, or against expectations based on the fossil record, is hampered by the complexity of Bayesian analyses: differences in results might be produced by differences in clock models, tree models, site models, priors (user-set or default) on any of the parameters used in these models, issues in implementation (bugs in the code, decisions about defaults, MCMC operators, etc.), user error in setting up the analysis or post-analysis processing, and/or issues with the data itself.

Canidae background. The bulk of canid evolution occurred in North America from the Eocene to present, and their fossil record is approximately continuous, with fossil diversity greater than extant diversity (approximately 35 living species, at least 123 well-described fossil taxa). In addition, the group has been thoroughly revised in three major monographs on the three subfamilies of Canidae: the extinct Hesperocyoninae (~27 species, 40-15 Ma; [26]), the extinct Borophaginae (~66 species, 34-2 Ma; [27]), and the extinct and extant Caninae (>40 fossil species, 34 Ma-present; [28]). All living dogs thus represent a small surviving branch, originating 10-12 Ma, of a much more massive tree of fossil Canidae. Thus, apart from utility for methods testing, the Canidae serve as a useful group for comparing trait evolution inferences made with living-only versus living+fossil datasets [19, 29, 30].

Further caveats on the Wang/Tedford Canidae tree.

An early version of this manuscript suggested that the Canidae tree might serve as a “ground truth” dataset for phylogenetic dating methods. (Therefore, various filenames in Supplemental Data reflect this language, although we have changed it in the manuscript.) “Ground truth” is a term taken from remote sensing [31], where researchers visit points on the ground to measure the accuracy of statistical classifications of landscape features from satellite imagery (e.g. primary forest versus secondary forest versus grassland). Subsequent discussion indicated that using the term “truth” raised philosophical questions about what “the truth” means, whether or not “good approximations” can count as “ground truth”, whether or not approximate truth about key focal dates can be valid even while fine-scale topological issues are unresolved, etc. While the term “ground truth” is a commonplace in geography, and does not carry the connotation of absolute precision or perfect knowledge of the truth, introducing it to phylogenetics would require a larger discussion than is possible here. Therefore, this study focuses on comparing Bayesian inference to previous expert opinion.

Several points should be considered by researchers making use of the Wang/Tedford expert tree as a test dataset for phylogenetic methods. While Wang and Tedford are indeed extremely accomplished experts, their dated trees are still subjective to a degree, constituting hand-drawings built on a maximum-parsimony tree, using stratigraphic consistency to resolve some polytomies, and then using stratigraphy, continuous traits and additional data and expert judgment to subjectively place uncoded specimens and to depict ancestral-descendant relationships. In addition, the Wang/Tedford tree does not perfectly match the molecular tree for living Caninae, and the reciprocal monophyly of the subfamilies may not be as secure as Wang/Tedford assumed (Graham Slater, personal communication). These points are significant for various individual nodes, but we believe that the three main node dates we use for Figure 1 would not be affected by these issues.

A second potential criticism that could be raised is that paleontological timetrees, as a general rule, are not hypotheses about divergence times: they depict taxon stratigraphic ranges and cladistic relationships, but the node spacing below the stratigraphic ranges is arbitrary. This is indeed usually the case, but Tedford and Wang actually went a fair bit beyond usual practice. For example, they actually did not adhere to common paleontological practices such as (a) always forcing nodes below stratigraphic ranges according to something like a 1-Ma minimum branch length (MBL) approach; (b) forcing sister taxa with overlapping stratigraphic ranges to have their ancestor node below the first occurrence of both species; or (c) relying on stratigraphic ranges of higher taxa (genera etc.) made up of occurrences of specimens that may not be identified to species.

Instead, Tedford and Wang (for the most part) depict species-level stratigraphic ranges. They depict these time ranges at sub-million-year resolution, and they allow direct ancestors and paraphyletic species rather than forcing cladistic sister taxa to be reciprocally monophyletic in the time tree. Furthermore, the fossil species diversity is often equal to or greater than the living species diversity, and the sampling of species through time is unusually through (meaning that dramatic range extensions of most of these species are not likely -- absence in sampling at a time point is likely reasonably good evidence of true absence at that time point, particularly past a few million years).

In this case, therefore, we think we are more justified than one would usually be in taking the expert paleontological tree as a source of comparison dates. Another way to consider the question is the following: in the case of Canidae, what is likely in terms of future discoveries about modifications and extensions of species stratigraphic ranges? Are we likely to see *Canis lupus* (current time range ~0-1 Ma) range-extended back to 15 Ma? Are we likely to find a crown Caninae (crown age ~12 Ma) older than all of the stem Caninae *Leptocyon* species, of which 11 or so are known that range from 7-34 Ma?

Such issues are indeed likely common in many clades studied by palaeontologists, but are much less of an issue in Canidae. We believe that a reasonable summary of the situation is that Tedford and Wang's semi-subjective hand-drawn plots are likely to be better estimates of a dated history than almost any dating analysis for almost any other fossil group (foraminifera might be an exception). This statement is admittedly less true for the Hesperocyoninae part of the tree, as discussed elsewhere.

Supplemental Methods

Expert tree. The expert tree was digitized using TreeRogue [32], with judgment calls resolved in favour of preserving Wang and Tedford's depictions of divergence times. The source figures were, specifically, Figure 65 of Wang (1994) [26]; Figure 141 of Wang et al. (1999) [27]; and Figure 66 of Tedford et al. (2009) [28]. Digitization resolution was <1 my, undoubtedly more precise than either the expert estimate or any Bayesian inference.

A plot of the tree, and a lineages-through-time plot, are available in Supplemental Figure 1. The Newick file is Canidae_ground_truth.newick (Supplemental Data).

Data. Morphological characters and dates came from the published matrix of Slater (2015) [19], specifically the Dryad repository [33] containing a NEXUS file with both morphology and MrBayes commands. Slater's matrix synthesized and updated the matrices published in the monographs by Wang and Tedford.

The data (characters and tip-dates) were left unchanged, except that no non-North American species were removed. Slater used last-occurrence dates for the tip-date of each species; for the purposes of tip-dating, this decision might be suboptimal, because a specimen bearing characters may sample from anywhere in a species' time-range. However, the nature of OTUs in tip-dating analyses is a complex question not yet addressed in the literature (Matzke and Irmis, this volume). Therefore, Slater's tip dates were retained for purposes of simplicity and direct comparability. Slater also used extensive node calibrations to represent the stratigraphic first occurrences of many taxa; these are reasonable given Slater's goal (fitting models of trait evolution), but node calibrations obscure the differences between tip-dating methods and so they were deleted from most analyses here.

Terminology for analyses. There appears to be variation in the literature and in the program documentation of MrBayes 3.2.x and Beast2 regarding the exact models being referred with terms such as "birth-death serial sampling" and "fossilized birth-death" process. Therefore, we are adopting the terminology of

“SA-BDSS” and “noSA-BDSS,” following the usage of the Bapst et al. (in review, this issue).

"SA-BDSS" refers to:

- Beast2 "SABD" model
  (which approximate equals, or is supposed to be very similar to)
- MrBayes FBD with Sampled Ancestors
- Or, sometimes it is called just "Fossilized Birth-Death" (e.g. Gavryushkina et al 2014)
- MrBayes with these settings:
  - prset brlenspr=clock:fossilization;
    [BDSS instead of uniform node age prior]
  - prset samplestrat = random;
    [possible some tips are sampled ancestors; available starting with MrBayes 3.2.5]

"noSA-BDSS" refers to:

- Beast2 "BDSS"
- Beast1 "BDSS"
  (these approximate equal, or are supposed to be very similar to)
- MrBayes FBD without Sampled Ancestors
- Or sometimes it is called just "Transmission Birth-Death" (e.g. Gavryushkina et al 2014)
- MrBayes with
  - prset brlenspr=clock:fossilization
    [BDSS instead of uniform node age prior]
  - prset samplestrat = fossiltip;
    [no sampled ancestors, every tip goes extinct; available starting with MrBayes 3.2.2]

MrBayes analyses.

The 6 focal analyses presented in the main text are mb1_UC (Slater’s original uniform tree prior analysis including node date calibrations, with some corrections), mb8_UU (uniform node age prior, unconstrained node dates, flat priors on clock parameters, uniform(45,100) prior on the root age), mb9x_SA (mb8_UU but with SA-BDSS tree prior and flat priors on speciation, extinction, and sampling rate), and mb10_noSA (mb9_SA but noSA-BDSS, i.e. disallowing sampled ancestors via the command “prset samplestrat = fossiltips;”).

These focal analyses were selected out of a much larger collection of MrBayes analyses (58 total) that were run while experimenting with modifications of the original Slater (2015) NEXUS file. A summary of the variant inputs, and the results, is presented in Supplemental Table S2. Apart from the issues surrounding the uniform tree prior versus birth-death tree priors, which are the topic of the main text, Appendix 1 identifies other issues noticed in the interactions between in the Slater NEXUS file, the MrBayes implementation, and documentation.

All MrBayes analyses (following Slater's settings) ran for 5 million generations (sampling every 2500), with 2 runs, 4 chains. (The only exceptions were 3 analyses run with varying temperatures, and another 4 runs aimed at improving the unsatisfactory topological result of the original mb9_SA run, and 8 exploratory runs (conducted in the review stage) with MrBayes 3.2.6; see Supplemental Table S2). Typically this was sufficient for convergence as assessed in Tracer plots and MrBayes output statistics. For a few non-focal runs (Supplemental Table S2) it was not. Sometimes this was due to improper settings, although Table S2 shows some unusual behavior in some MrBayes SA-BDSS runs even after settings were corrected to the best of our ability. As the purpose of many of the runs was exploratory, to determine the effects of certain versions, models, or settings (i.e. including problems with convergence), no effort was made to improve sampling further (except for the mb9 variants mentioned above). The Perl scripts burntrees and catmb [34] were used to extract the last 50% of each tree sample and convert to dated trees. The mb9x_SA run, selected for Figure 1, is one of the mb9 variants run for much longer (100 million generations) and at the default temperature (0.1, instead of Slater's 0.5), in order to minimize the chance of displaying an uncharacteristic result.

Justification of using fixed tip-dates in this study. We deliberately decided to use Slater's (2015) fixed tip-dates (which were last-occurrence dates), even though they are not necessarily ideal from the perspective of obtaining a "best possible" dating analysis and maximum matching between dating analyses and the expert tree. This was done for several reasons:

1. Direct comparability between the Slater analysis, the variant analyses here, and the expert tree (where the tips terminate at the last occurrence of the lineage, a practice also followed by Slater).
2. Isolation and identification of the effect of methodological choices (meaning specifically: settings in the computer programs, such as choice of tree/node age priors) on dates. If the tip-dates are varying, this could be the cause of any difference observed, confounding the comparison.
3. The best practices for incorporating species' stratigraphic ranges in tip ages are, at the moment, unresolved in the literature. It appears to be becoming standard practice that, for species with stratigraphic ranges, the species' stratigraphic ranges are input as uniform priors on the tip dates. However, this decision assumes that the following two things are identical:
    a. A fossil specimen has a date that is uncertain, so the tip gets a Uniform prior across this date range.
    b. A fossil taxon has a stratigraphic range, so the tip gets a Uniform prior across this date range.

    While (a) is correct, (b) actually represents multiple specimens. It could be argued that, given knowledge of a fossil species, its stratigraphic range, and the assumption (perhaps a big one) that the character states do not change within the species over the stratigraphic range, then for all we know, those character states might have been sampled from any time point within the species' stratigraphic range.

However, this might or might not be a reasonable approximation of reality. More appropriate strategies might be:

c. The researcher codes many individual specimens and then puts them into the analysis as separate tips with separate specimen-specific date priors. This, of course, involves vastly more work and perhaps much more wrestling with incomplete specimens and the like.
d. Take the character states for the OTU and create a number of duplicate OTUs with identical character states, perhaps one for each relevant stratigraphic unit where the species has been found. This would be much easier than (c), but involves creating data.

Which strategies are computationally feasible (creating many OTUs may fatally slow down MCMC searches) and/or acceptable approximations are basically unexplored in the literature at present.

4. Finally, a "best possible" dating analysis would include DNA (which exists for virtually all living dog species, and some extinct ones) in a total evidence analysis, which again introduces issues beyond those we choice to explore in this study.

Resolution of point #3 will take additional focused studies, which cannot be done here. And, the goal of the present study is to evaluate which tip-dating methods are "in the ballpark," which is a question that needs to be answered before much more elaborate and time-consuming analyses of canids are done.

Justification of a broad prior on the root age. For most of the dating analyses (the replications of Slater's intended analysis – mb1_orig and mb1_UC – are the exceptions), we set a broad, uninformative prior on the root age: Uniform(45, 100). The minimum age was set to match Slater's minimum age for Canidae+the outgroup, and the maximum was set to represent the situation where no useful information is supplied to the analysis about the age of the root. We did this because, in many real-life dating situations, the part of the point of the dating analysis is to *infer* the age of the root and nodes near the root, rather than have it be constrained *a priori* by an informative prior.

Readers who wish to see what a uniform node-age prior analysis would look like with a reasonable, informative prior on the root age may consult the mb1_UC analysis. This analysis has many additional node constraints, but a "no constraints except for an informative root age prior" analysis would fall between mb1_UC and mb8_UU.

Comparing key node dates between the expert tree and the Bayesian inferences. The three nodes chosen for comparison in Figure 1 represent the time of origin for major taxa in Canidae. The definitions used here are:

- Node 1. Crown *Canis*: Common ancestor of all living *Canis* in the tree (including *Cuon*, *Lycaon*, and *Xenocyon* included inside of *Canis*, as it is known to be paraphyletic with respect to these taxa; [28]).
- Node 2. Crown Caninae: The common ancestor of living *Canis* in the tree (*Canis, Cuon*, *Lycaon*, and *Xenocyon*) and living foxes (*Urocyon* and *Vulpes*). This node is the common ancestor of all living Canidae.
- Node 3. Total Group Canidae: the common ancestor of all living and fossil species in the dataset, excluding Slater's "outgroup" OTU.

These three nodes were used as the primary evaluation method because:

a. They represent "important" nodes – Nodes 1 and 2 would be important calibration points in node-dating studies, and Node 3 represents the origin of the group.
b. The Wang/Tedford expert tree and the Slater character matrix do not overlap perfectly in their OTUs. This can be solved by reducing both trees to the set of common OTUs, but such trees might shift the meaning of crown group nodes if key taxa are left out.
c. Node ages will be highly correlated with each other across the tree (older trees will tend to have most/all nodes older), so it is debatable if analysing dozens of node dates provides a great advantage over analysing a few key node dates.
d. When many nodes are analysed, for example with regressions, there is a tendency to interpret these as detailed statements about the biases of the models. We have included these regressions in Supplemental Data, but we warn that they cannot bear the weight of representing detailed statements about the biases of the models. For example, as discussed elsewhere, the fact that the BDSS-type analyses infer ages younger than the expert tree at the bottom of the tree is very likely due to the use of Slater's last-occurrence dates at the tips, using fixed dates, and the lower actual rate of sampling fossils for the Hesperocyoninae, when the analyses we have conducted assume constant sampling rates.

Attempting to improve MrBayes SA-BDSS inference. The topological result of the original mb9_SA SA-BDSS run was unsatisfactory in that wild dogs (*Lycaon pictus*) and the dhole (*Cuon javanicus*) came out as a clade within the extinct Borophagines. It was suggested that adding more runs and varying temperatures might help. We attempted this by changing the number of runs from the default 2 to 4, and then running three analyses with different temperatures: 9a (4 runs, temperature=0.05), 9b (4 runs, temperature=0.5, the default), and 9c (4 runs, temperature=1). These runs are numbered 36a, 36b, and 36c in Supplemental Table S2. Further runs consisted of a series of replicate runs for 100 million generations, varying (a) settings with nruns=1, nchains=4 versus nruns=2, nchains=4; (b) settings with temperature=0.5 versus temperature=0.1; and (c) MrBayes 3.2.5 versus MrBayes 3.2.6. The results of these runs are also listed in Table S2. Overall, the extra runs indicated that mb9_SA was an unusual result, and most runs on most settings retrieve a more conventional topology. An example of a run with a conventional topology is reported in Figure 1 as mb9x_SA.

Beast2 analyses. BEASTmasteR was used to construct the XML files for two Beast2 tip-dating analyses; the R scripts and Excel settings files are available in SM. The first analysis (r1_noSA) used a noSA-BDSS tree prior [35]; the second (r2_SA) used SA-BDSS [2, 3]. Flat priors were used for each major parameter (mean and SD of the lognormal relaxed clock; and birth, death, and serial sampling rates). Rho (proportion of living species sampled) was fixed to 1, as it is not statistically identifiable if left free in noSA-BDSS [36], and the dataset is relatively complete, at least for widespread species in North America (where the vast majority of the fossil record is located). Although rho is identifiable in SA-BDSS [2], we kept the same setting to ensure direct comparability across analyses. All runs in both programs used a single morphology partition, with an M*k*v model correcting for the ascertainment bias against invariant characters [37, 38] and gamma-distributed rate variation with 4 rate categories.

The Beast2 analyses were run for 50 million generations (sampling every 25000). For all runs, TreeAnnotator was used to choose the Maximum Clade Credibility (MCC) tree and calculate node-date HPDs (95% highest posterior densities) and bipartition posterior probabilities (PP). Burntrees [34] was used to process MrBayes outputs for input into TreeAnnotator.

BEASTmasteR [39] and custom R scripts (Supplemental Data) were used to plot all MrBayes and Beast2 MCC trees, as well as the trace plots for all key parameters (SI), and to extract parameter and node-date estimates of interest. To assess the overall estimate of topology in each analysis, the topological symmetric distance (treedist function; phangorn R package; [40]) between the dated MCC tree and the MCC tree from the undated (mb2_undated) analysis was calculated and compared to the distribution of distances between trees in the mb2_undated post-burnin treecloud [41].

For the six focal analyses, an additional test was run to see how well the methods could predict selected tip dates [20]. Four tips were selected from across the tree (*Canis ferox*, 3.5 Ma; *Epicyon haydeni*, 5.3 Ma; *Leptocyon gregorii*, 23 Ma; *Hesperocyon gregarious*, 30.8 Ma). For each tip and focal analysis, the settings file was modified to change the tip date to a uniform(0,100) prior. After the MCMC run, the sampled tip date was extracted from the post-burnin tree sample and plotted as a histogram.

Supplemental Results

Summary results of all 40 variant analyses are presented in Supplemental Table S2 (supplemental Excel file). Trace plots of key variables for all 40 analyses are available in Supplemental Data (file *Canidae_traceLogs.pdf*). Plots of the MCC trees for all 40 analyses are also available (*Canidae_treeLogs.pdf*).

Rate parameters. Estimates of rate parameters in the focal analyses are consistent with the dating results, in that analyses with the youngest node age estimates have the highest clock, speciation, and sampling rates. The parameter describing the mean of relaxed clock branchwise rate variation (IGRvar for MrBayes, clockSD for Beast2) was inferred with similar precision across all analyses, despite uninformative priors, suggesting no special effort is needed to determine the prior for clock relaxation (clock models with autocorrelated rates may be different; [5]). The uncorrected Slater analysis does show the effect of the IGRvar prior used there (the intention was a diffuse prior, but the setting used forced a close-to-strict clock instead; see Appendix 1); however, the downstream effect on the analysis was minimal.

Topological distances between Bayesian posterior tree samples and expert tree. Comparing topological distances (Table S1) provides a more systematic assessment of topology differences between analyses. Randomly chosen trees in the post-burnin posterior distribution of Slater's undated MrBayes analysis (mb2_undated; SM) have a mean symmetric distance of 31.0% (95% C.I.=±8.3%). The dated MCC trees of all 8 focal analyses fall within this range, except for the MrBayes-uniform and -SA-BDSS analyses (which both have significantly higher topological distances).

Correlation between estimated and true node dates. A heuristic view of the correlation between date estimates and expert tree dates is shown in linear regression plots (Supplemental Data, file *expert_vs_estimated_node_ages.pdf*) comparing the ages of nodes that are shared between the expert tree and estimated tree (i.e., have the same descendant OTUs after removing OTUs not shared by both trees; 127 OTUs were shared). These regressions should be taken as heuristic exploration of the inference biases resulting from the of the combination of data, priors, and methods used in this study, rather than definitive statements about the biases of the methods in all situations.

While all analyses have statistically significant correlation to expert tree ages, mb8_UU has a lower R^2 (0.71) than the others (>0.9), and also has a systematic bias towards older ages (intercept=6.67±2.94; for other analyses intercept is ~0-1). All of the BD analyses have a bias towards underestimating dates near the base of the tree, where the expert tree is dominated by Hesperocyoninae, with a lower density of fossil OTUs and thus longer branches; the BD methods tend to infer shorter branches here, leading to younger ages. This bias leads to an underestimation of ages by about 2 my at age 20 Ma, to 4-5 my at 30 Ma.

Posterior prediction of tip dates. Prediction of tip dates (Supplemental Data) generally succeeded in overlapping the true value, although uncertainty is usually high (5+ my) and extreme in the case of mb8_UU (17+ my). SA-BDSS analyses consistently produce younger date estimates than noSA-BDSS analyses, but the effect is small (~1 my).

MrBayes SA-BDSS topology and convergence issues. The disagreement between the original MrBayes noSA-BDSS and SA-BDSS (mb10_noSA and mb9_SA) analyses about the position of the wild dogs (*Lycaon pictus*) and dhole (*Cuon javanicus*) was puzzling, because both analyses place this clade with a sister with posterior probability 1.0; they just disagree on whether that sister is *Xenocyon texanus* (in Caninae) or *Desmocyon thomsoni* (in the extinct Borophaginae). In undated analyses (mb2_undated), the clade has a relatively long morphological branch, and this may make it more difficult to place; however, further MrBayes runs (see Supplemental Methods) tended to place these taxa within Canini, so we report one of those longer runs in Figure 1 (mb9x_SA; see Supplemental Table S2).

Supplemental Discussion

Uniform tree prior and spacing of node dates. We have observed that even well-constrained analyses using the uniform tree prior appear to exhibit a tendency to have "unrealistically even" spacing of node ages between calibrations and tip dates, regardless of morphological branch lengths. This can be seen in our mb1_UC analysis, for example in the old age for crown *Canis*. We have also observed the phenomenon with other datasets, and it seems to be a feature of published uniform tree prior analyses as well. Admittedly we have not devised a way to quantify the observation of "unrealistically even" node date estimates, but the phenomenon does appear to be real, at least when uniform tree prior results can be compared side-by-side to the expert tree and noSA-BDSS/SA-BDSS trees (e.g., mb1_UC vs. mb10_noSA, r1_noSA, r2_SA), or to undated trees (e.g. mb2_undated).

Caveats about the accuracy of inferred ancestors. The only point we would make regarding inferring ancestors is that our Sampled-Ancestor analyses tend to retrieve some of the ancestors that were postulated by Wang and Tedford – most prominently, *Leptocyon* species as ancestral to Caninae. We note that this is encouraging, but we caution that our study is not intended as a detailed examination of the accuracy of inferring direct ancestors, although the Canidae dataset might prove to be useful for such a study in the future.

Informally, in experimenting with SA-BDSS analyses on various datasets, we have found that they tend to infer a fair number of direct ancestors both on datasets where the sampling of fossil species diversity is likely to be reasonably thorough (as in Canidae), but also in situations where it is not (for example, dinosaurs). Accurately inferring direct ancestors is much more likely to be plausible in the former situation than the latter.

Future work on this point should carefully examine the issue of how much of true fossil diversity is being sampled and whether or not this is being thoroughly taken into account; a key question may be the interpretation of the BDSS rho parameter for fossils-only datasets, as rho typically means "proportion of extant species sampled in the present." For example, if the group is extinct, rho should be 0, and sampling rate inferred – but we may have some prior belief about true fossil diversity versus the fossil diversity captured in the data matrix.

Captions for Supplemental Figures, Tables, and Data

Supplemental Figure S1. Top: plot of the expert tree, derived from digitization of the phylogenies of Canidae published in the monographs of Wang and Tedford, using TreeRogue. Bottom: Lineages through time plot of the expert tree.

Supplemental Table S1. Five Bayesian tip-dating analyses are compared to (column 1) the conclusions of Tedford & Wang. The two Slater analyses (original, and a modification repairing some unintended issues; cols. 2-3) represent MrBayes analyses under a uniform node age prior, constrained by both tip dates and many node-date priors. The third run (col. 4) shows the effect of removing the node-age calibrations, and putting flat priors on the parameters for clock rate and variation. The fourth run (col. 5) shows the drastic effect of switching to a fossilized birth-death prior. Columns 5 and 6 show Beast2 analyses with flat clock priors and noSA-BDSS and SA-BDSS tree priors, respectively. These five analyses are drawn from the 60 analyses shown in Supplemental Table S2.

Supplemental Table S2. Summary settings and results of all 60 tip-dating analyses. As the table is large, it is presented as an Excel file.

Supplemental Data Files

canidae_traceLogs_ALL_v2.pdf -- Trace plots of key variables for all 60 analyses.

canidae_treeLogs_ALL_v2.pdf -- Plots of the MCC trees for all 60 analyses.

expert_vs_estimated_node_ages.pdf -- Linear regressions showing the correlation between the expert tree and estimated node ages, for nodes shared between the expert tree and estimated trees. Caveats for these regressions are discussed in Supplemental Text.

Canidae_expert.newick – The "expert" tree, derived from digitization of the phylogenies of Canidae published in the monographs of Wang and Tedford, using TreeRogue.

Table_S2_TipDate_runs_v3.xlsx – Summary of all 60 variant analyses (contains Supplemental Table S2, and some associated notes and file locations)

Matzke_Wright_SuppData.zip – A zipfile of all inputs, outputs, and scripts for all analyses.

Supplemental References

[1] Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A. & Drummond, A.J. 2014 BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* **10**, e1003537. (doi:10.1371/journal.pcbi.1003537).

[2] Gavryushkina, A., Welch, D., Stadler, T. & Drummond, A.J. 2014 Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Comput Biol* **10**, e1003919. (doi:10.1371/journal.pcbi.1003919).

[3] Heath, T.A., Huelsenbeck, J.P. & Stadler, T. 2014 The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences* **111**, E2957-E2966. (doi:10.1073/pnas.1319091111).

[4] Pyron, R.A. 2011 Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Systematic Biology* **60**, 466-481. (doi:10.1093/sysbio/syr047).

[5] Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D.L. & Rasnitsyn, A.P. 2012 A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology* **61**, 973-999. (doi:10.1093/sysbio/sys058).

[6] Zhang, C., Stadler, T., Klopfstein, S., Heath, T.A. & Ronquist, F. 2016 Total-evidence dating under the fossilized birth-death process. *Systematic Biology* **65**, 228-249. (doi:10.1093/sysbio/syv080).

[7] Alexandrou, M.A., Swartz, B.A., Matzke, N.J. & Oakley, T.H. 2013 Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. *Molecular Phylogenetics and Evolution* **69**, 514-523. (doi:http://dx.doi.org/10.1016/j.ympev.2013.07.026).

[8] Arcila, D., Alexander Pyron, R., Tyler, J.C., Ortí, G. & Betancur-R, R. 2015 An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). *Molecular Phylogenetics and Evolution* **82, Part A**, 131-145. (doi:http://dx.doi.org/10.1016/j.ympev.2014.10.011).

[9] Beck, R.M.D. & Lee, M.S.Y. 2014 Ancient dates or accelerated rates? Morphological clocks and the antiquity of placental mammals. *Proceedings of the Royal Society of London B: Biological Sciences* **281**.

[10] Bracken-Grissom, H.D., Ahyong, S.T., Wilkinson, R.D., Feldmann, R.M., Schweitzer, C.E., Breinholt, J.W., Bendall, M., Palero, F., Chan, T.-Y., Felder, D.L., Robles, R., Chu, K.-H., Tsang, L.-M., Kim, D., Martin, J.W. & Crandall, K.A. 2014 The emergence of lobsters: phylogenetic relationships, morphological evolution and divergence time comparisons of an ancient group (Decapoda: Achelata, Astacidea, Glypheidea, Polychelida). *Systematic Biology* **63**, 457-479. (doi:10.1093/sysbio/syu008).

[11] Dornburg, A., Moore, J., Beaulieu, J.M., Eytan, R.I. & Near, T.J. 2015 The impact of shifts in marine biodiversity hotspots on patterns of range evolution: Evidence from the Holocentridae (squirrelfishes and soldierfishes). *Evolution* **69**, 146-161. (doi:10.1111/evo.12562).

[12] Grimm, G.W., Kapli, P., Bomfleur, B., McLoughlin, S. & Renner, S.S. 2015 Using more than the oldest fossils: dating Osmundaceae with three

Bayesian clock approaches. *Systematic Biology* **64**, 396-405. (doi:10.1093/sysbio/syu108).

[13] Huang, D., Goldberg, E.E. & Roy, K. 2015 Fossils, phylogenies, and the challenge of preserving evolutionary history in the face of anthropogenic extinctions. *Proceedings of the National Academy of Sciences* **112**, 4909-4914. (doi:10.1073/pnas.1409886112).

[14] Lee, M.S.Y., Cau, A., Naish, D. & Dyke, G.J. 2014 Morphological clocks in paleontology, and a mid-Cretaceous origin of crown Aves. *Systematic Biology* **63**, 442-449. (doi:10.1093/sysbio/syt110).

[15] Lee, M.S.Y., Cau, A., Naish, D. & Dyke, G.J. 2014 Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science* **345**, 562-566. (doi:10.1126/science.1252243).

[16] Marx, F.G. & Fordyce, R.E. 2015 Baleen boom and bust: a synthesis of mysticete phylogeny, diversity and disparity. *Royal Society Open Science* **2**.

[17] Sharma, P.P. & Giribet, G. 2014 A revised dated phylogeny of the arachnid order Opiliones. *Frontiers in Genetics* **5**, 255. (doi:10.3389/fgene.2014.00255).

[18] Slater, G.J. 2013 Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary. *Methods in Ecology and Evolution* **4**, 734-744. (doi:10.1111/2041-210X.12084).

[19] Slater, G.J. 2015 Iterative adaptive radiations of fossil canids show no evidence for diversity-dependent trait evolution. *Proceedings of the National Academy of Sciences* **112**, 4897-4902. (doi:10.1073/pnas.1403666111).

[20] Drummond, A.J. & Stadler, T. 2016 Bayesian phylogenetic estimation of fossil ages. *arXiv*, 1-23. (doi:arXiv:1601.07447v1).

[21] Matzke, N.J. 2016 The evolution of antievolution policies after *Kitzmiller versus Dover*. *Science* **351**, 28-30. (doi:10.1126/science.aad4057).

[22] Wood, H.M., Gillespie, R.G., Griswold, C.E. & Wainwright, P.C. 2015 Why is Madagascar special? The extraordinarily slow evolution of pelican spiders (Araneae, Archaeidae). *Evolution* **69**, 462-481. (doi:10.1111/evo.12578).

[23] Wood, H.M., Matzke, N.J., Gillespie, R.G. & Griswold, C.E. 2013 Treating fossils as terminal taxa in divergence time estimation reveals ancient vicariance patterns in the Palpimanoid spiders. *Systematic Biology* **62**, 264-284. (doi:10.1093/sysbio/sys092).

[24] Puttick, M.N., Thomas, G.H. & Benton, M.J. 2016 Dating placentalia: morphological clocks fail to close the molecular fossil gap. *Evolution*. (doi:10.1111/evo.12907).

[25] Tseng, Z.J., Wang, X., Slater, G.J., Takeuchi, G.T., Li, Q., Liu, J. & Xie, G. 2013 Himalayan fossils of the oldest known pantherine establish ancient origin of big cats. *Proceedings of the Royal Society of London B: Biological Sciences* **281**.

[26] Wang, X. 1994 Phylogenetic systematics of the Hesperocyoninae (Carnivora, Canidae). *Bulletin of the American Museum of Natural History* **221**, 1-207.

[27] Wang, X.T., Richard H.; Taylor, Beryl E. 1999 Phylogenetic systematics of the Borophaginae. *Bulletin of the American Museum of Natural History* **243**, 1-391.

[28] Tedford, R.H.W., Xiaoming; Taylor, Beryl E. 2009 Phylogenetic systematics of the North American fossil Caninae (Carnivora, Canidae). *Bulletin of the American Museum of Natural History* **325**, 1-218.

[29] Finarelli, J.A. & Flynn, J.J. 2006 Ancestral state reconstruction of body size in the Caniformia (Carnivora, Mammalia): the effects of incorporating data from the fossil record. *Systematic Biology* **55**, 301-313. (doi:10.1080/10635150500541698).

[30] Jablonski, D. & Shubin, N.H. 2015 The future of the fossil record: Paleontology in the 21st century. *Proceedings of the National Academy of Sciences* **112**, 4852-4858. (doi:10.1073/pnas.1505146112).

[31] Powell, R.L., Matzke, N., de Souza Jr, C., Clark, M., Numata, I., Hess, L.L. & Roberts, D.A. 2004 Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon. *Remote Sensing of Environment* **90**, 221-234. (doi:http://dx.doi.org/10.1016/j.rse.2003.12.007).

[32] Matzke, N.J. 2013 TreeRogue: R code for digitizing trees. https://stat.ethz.ch/pipermail/r-sig-phylo/2010-October/000816.html

[33] Slater, G.J. 2015. Data from: Iterative adaptive radiations of fossil canids show no evidence for diversity-dependent trait evolution. Dryad. Accessed May 1, 2015. http://dx.doi.org/10.5061/dryad.9qd51

[34] Nylander, J.A.A. 2014. Burntrees v.0.2.2.: Perl script for manipulating MrBayes tree and parameter files. Accessed May 1, 2015. https://github.com/nylander/Burntrees

[35] Stadler, T. 2010 Sampling-through-time in birth-death trees. *Journal of Theoretical Biology* **267**, 396-404. (doi:10.1016/j.jtbi.2010.09.010).

[36] Stadler, T. & Yang, Z. 2013 Dating phylogenies with sequentially sampled tips. *Systematic Biology* **62**, 674-688. (doi:10.1093/sysbio/syt030).

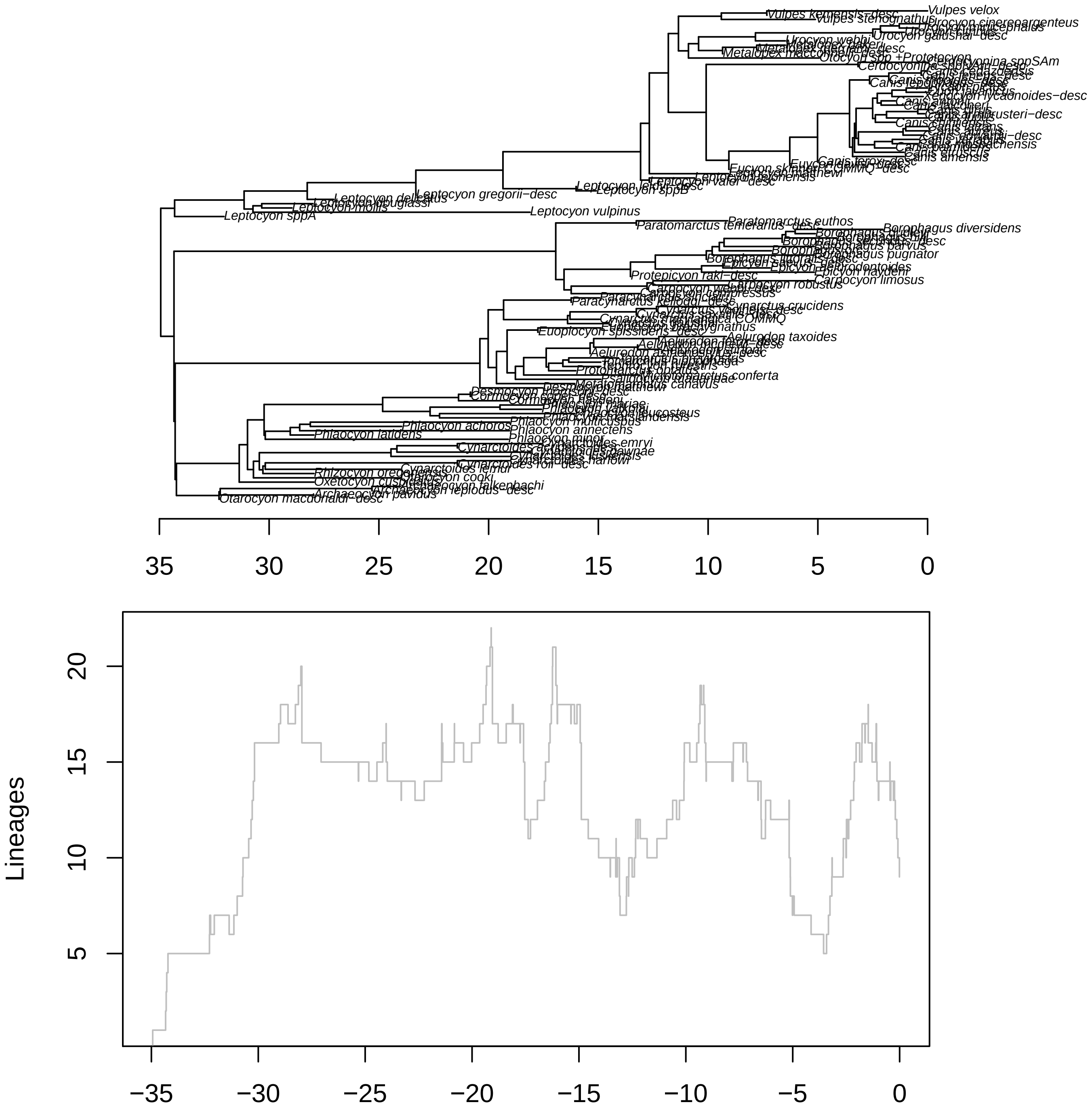
[37] Lewis, P.O. 2001 A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* **50**, 913-925. (doi:10.1080/106351501753462876).

[38] Felsenstein, J. 1992 Phylogenies from restriction sites: a maximum-likelihood approach. *Evolution* **46**, 159-173. (doi:10.2307/2409811).

[39] Matzke, N.J. 2015 BEASTmasteR: automated conversion of NEXUS data to BEAST2 XML format, for fossil tip-dating and other uses. PhyloWiki. http://phylo.wikidot.com/beastmaster

[40] Schliep, K.P. 2011 phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592-593. (doi:10.1093/bioinformatics/btq706).

[41] Matzke, N.J., Shih, P.M. & Kerfeld, C.A. 2014 Bayesian analysis of congruence of core genes in *Prochlorococcus* and *Synechococcus* and implications on horizontal gene transfer. *PLoS ONE* **9**, e85103. (doi:10.1371/journal.pone.0085103).



Supplemental Figure S1

Supplemental Table S1. Five Bayesian tip-dating runs are compared to (column 1) the conclusions of Tedford & Wang. The two Slater analyses (original, and a modification repairing some unintended issues; cols. 2-3) represent MrBayes analyses under a uniform node age prior, constrained by both tip dates and many node-date priors. The third run (col. 4) shows the effect of removing the node-age calibrations, and putting flat priors on the parameters for clock rate and variation. The fourth run (col. 5) shows the drastic effect of switching to a fossilized birth-death prior. Columns 5 and 6 shown Beast2 runs with flat clock priors and noSA-BDSS and SA-BDSS tree priors, respectively. These five runs are drawn from the 40 analyses shown in Supplemental Table 1.

Estimates	Analysis	Tedford & Wang, manually time-scaled cladogram	mb1_orig: Slater (2015) MrBayes analysis, original	mb1_UC: Slater (2015) MrBayes analysis, corrected	mb8_UU: MrBayes analysis, unconstrained, uniform node age prior	mb10_noSA: MrBayes analysis, unconstrained, noSA-BDSS tree prior	mb9x_SA: MrBayes analysis, unconstrained, fossilized-SA-BDSS tree prior	r1_noSA: Beast2, unconstrained, noSA-BDSS tree prior	r2_SA: Beast2, unconstrained, SA-BDSS tree prior	Notes
node ages	age(crown *Canis*)	3.2	7.3 (5.2,9.8)	7.4 (5.1,9.6)	27.5 (16.9,39.1)	4.4 (3.3,5.5)	2.8 (2.0,3.8)	3.7 (2.9,4.8)	2.7 (2.1,3.3)	4,6
	age(crown Caninae)	11.7	21.4 (17.9,25.8)	21.2 (17.8,25.5)	38.9 (29.0,54.0)	12.1 (10.7,13.7)	9.8 (8.9,10.2)	12.4 (10.9,14.1)	10.6 (10.3,12.0)	3,5
	age(Canidae)	40-36	41.0 (40.4,41.8)	41.0 (40.4,43.3)	49.0 (44.4,66.8)	38.3 (36.0,41.3)	36.3 (34.6,39.0)	36.8 (35.0,39.0)	36.1 (34.2,38.0)	2
	prior on root age	-	gamma(1,1) expect: 1 (0.024, 3.6)	offsetexp(45, 50) expect: 50 (45.1,63.6)	uniform(45,100)			none		
	age(root)	41.5	43.4 (41.2,45.5)	47.1 (45.0,52.0)	49.5 (45.0,67.7)	45.3 (45.0,46.4)	45.3 (45.0,46.4)	40.6 (40.0,42.1)	40.0 (40.0,41.2)	1
topology	Percent toplogical distance to mb2_undated, an undated MrBayes tree (mean betw. undated trees = 24.6%)		33.8%	32.3%	40.8%	33.1%	46.2%	36.2%	33.8%	
	crown Caninae monophyletic?	y	y	y	y	y	y	y	y	
	((Can., Boro.), Hesp.)?	y	n	n	y	y	n	y	y	
clock	model	-	IGR	IGR	IGR	IGR	IGR	ucld	ucld	
	prior	-	lognorm(-6,0.1) expectation: 0.0025 (0.002,0.003)		truncated normal(0.0025,0.1) expectation: 0.08 (0.003,0.23)			uniform(0,10)		
	clock rate	-	0.00309 (0.00201,0.00438)	0.0094 (0.0071,0.012)	0.0045 (0.0025,0.006)	0.019 (0.015,0.024)	0.027 (0.020,0.034)	0.038 (0.025,0.053)	0.052 (0.033,0.075)	
	variation prior	-	exp(126.887); expectation: 0.0079 (0.00019,0.029)	uniform(0.0001,200)				uniform(0,10)		
	among-branch variation parameter	-	0.00955 (0.00479,0.0154)	0.028 (0.018,0.039)	0.034 (0.021,0.046)	0.024 (0.015,0.033)	0.035 (0.023,0.049)	1.20 (0.98,1.42)	1.24 (1.03,1.47)	7
	among-site variation gamma parameter	-	1.03 (0.19, 1.92)	1.45 (0.87,2.02)	1.44 (0.86, 2.06)	1.38 (0.87,2.0)	1.270 (0.79,1.78)	1.16 (0.66,1.70)	1.14 (0.61,1.62)	
tip dates	*Canis ferox*	3.5	fixed(3.5)	4.4 (0.2,9.3)	10.4 (0.2,32.8)	2.8 (0.5,5.1)	(not run)	2.9 (0.5,5.2)	2.3 (0.4,4.2)	8,9
	Epicyon haydeni	10-5.3	fixed(5.332)	3.5 (0.4,7.9)	4.7 (0.2,17.1)	4.9 (1.3,7.4)	(not run)	5.4 (1.9,8.8)	5.0 (2.1,7.9)	
	Leptocyon gregorii	24.4-23	fixed(23)	23.3 (15.4,28.5)	32.3 (14.9,49.5)	21.8 (17.3,26.2)	(not run)	20.1 (14.3,26.7)	15.4 (12,20.4)	
	Hesperocyon gregarius	37.2-30.8	fixed(30.8)	35.7 (28,40.5)	32.2 (11.4,43.7)	33.7 (30.4,36.2)	(not run)	33.3 (30.1,35.7)	33.0 (30.3,34.8)	
tree	prior (all 3)	-	-	-	-	unif(0,10)	unif(0,10)	unif(0,10)	unif(0,10)	
	speciation	-	-	-	-	0.37 (0.041,0.79)	0.54 (0.44,0.64)	0.48 (0.29,0.73)	0.65 (0.36,1.08)	
	extinction	-	-	-	-	0.33 (0.037,0.71)	0.49 (0.44,0.53)	0.187 (0,0.53)	0.33 (0,0.90)	
	sampling	-	-	-	-	0.29 (0.033,0.63)	0.14 (0.096,0.21)	0.27 (0.16,0.38)	0.30 (0.15,0.44)	
Suppl. Table	run #	-	3	31	35	37	36x2	1	2	
	code	-	mb3.2.5, mb1_orig	mb3.2.5, mb1_UC	mb3.2.5, mb8_UU	mb3.2.5, mb10_noSA	mb3.2.5, mb9x_SA	r1_noSA	r2_SA	

Notes

1 Wang & Tedford (2008) date from Fig. 6.6, Arctoidea outgroup
40 Ma: Wang & Tedford (2008), Fig. 6.6, origin of Hesperocyoninae;

2 36 Ma: Tedford & Wang (2008), Fig. 7.1, divergence of Hesperocyoninae and Borophaginae+Caninae

3 Tedford et al. (2009), Fig. 66

4 Tedford et al. (2009), Fig. 66

5 For mb9, excluding Cuon/Lycaon, crown Caninae date is 11.6 (10.3,12.8)

6 For mb9, excluding Cuon/Lycaon, crown *Canis* date is 5.9 (5.0,7.1)

7 The branch-rate variation parameters are not equivalent between MrBayes IGR and Beast2 ucld relaxed clocks.

8 The time ranges in column 1 are the stratigraphic range of each species as indicated in Fig. 65 of Wang (1994; Hesperocyon), Fig. 141 of Wang et al. (1999; Epicyon), and Figure 66 of Tedford et al. (2009; Canis and Leptocyon).

9 The fixed() statements in column 2 are the tip ages used by Slater (2015) for these taxa. To estimate the tip date (next 4 columns), these were changed to uniform(0,100).