Resolution of basal wolf-like canid divergence using ancient DNA and fossil data

EEOB 563 Project

Bruno do Rosario Petrucci

April 27, 2021

1 Introduction

1.1 Background

The study of phylogenetics has been invaluable for the understanding of evolutionary relationships of both extinct and extant species. Unfortunately for paleobiologists, an overwhelmingly proportion of studies involving extinct taxa are forced to trust solely morphological characters, the only data available for fossils, which leads to a number of biases related to the subjective nature of morphological scoring [1]. This is a source of uncertainty for a number of studies looking into evolutionary relationships involving extinct species. One such example is the placement of the dire wolf (*Canis dirus*) on the canid phylogeny. High morphological similarity between the dire wolf and the popular gray wolf (*Canis lupus*) has led to the former's placement as a sister species or sampled ancestor of the latter in most studies using morphological data to study these relationships (e.g. [2]). Still, some have argued the high variability between individuals and the actual common set of characters might make a case for the placement of *dirus* in its own genus, *Aenocyon* [3]. This debate boils down to whether morphological similarity between *dirus* and *lupus* is the result of true evolutionary similarity or convergent evolution. As such, it is representative of many other issues on the phylogenetic placement of extinct species.

A recent study has helped shed light on this issue by sequencing DNA from dire wolf individuals [4]. This allows us to look for answers on these complex evolutionary relationships using a source of data that is less biased by convergence and subjective scoring. This study's results seem to support the view that *dirus* should be placed on its own genus, especially given that it was placed farther from *lupus* than some non-*Canis* species, namely the African wild dog (*Lycaon pictus*) and the dhole (*Cuon alpinus*). While this is good progress on the resolving of canid evolutionary relationships, the paper's analysis failed to resolve the order of divergence of basal canina (wolf-like canids) species - whether *dirus* or the African jackals (*Canis mesomelas* and *adustus*) diverged from the rest of the canina first remains an open question. I believe total-evidence analysis, using the study's nuclear data allied with morphological and fossil ages, can help resolve these intricacies involving specific divergence times of canids.

Total-evidence analysis has been developed as a framework for using all available data in the construction of phylogenetic trees [5]. The use of both morphological and DNA data, when available, is often accompanied by the consideration of fossil data, which can help reduce uncertainty on node age estimations done from molecular and morphological data [6]. The Fossilized-Birth-Death process is one of the models that allows for these analysis by explicitly modelling species diversification and sampling as part of the construction of tree topology [7]. These analyses can account for morphological and molecular evolution while using fossil ages to improve divergence time estimations. In this study, I will collect nuclear DNA, morphological data, and fossil ages for the *Canis*, *Lycaon*, *Cuon*, and *Urocyon* (a fox, as an outgroup) genera, and use an FBD model as a framework for a total-evidence analysis to attempt and reduce uncertainty on the divergence times of basal canina. Even if it does not resolve the relationships we are interested in, this might provide clues on how combined-evidence analyses can help reconcile conflicting results when using different sources of data.

1.2 Data

The nuclear data for this project will come directly from the study. In correspondence with the authors, I received the raw SNP data from 5 dire wolf, 3 gray wolf, 2 African wild dog, and one each for 7 other canina species, plus two foxes as outgroups. This accounted for around 12 million DNA sites. It should be noted that this data contained 1. transversions only, since ancient DNA damage (such as for the dire wolf) could show up as

transitions so that these are eliminated, and 2. a high amount of missing data ("N" base), especially for the ancient DNA samples. To treat the data, I first eliminated 3 dire wolf samples completely, since they were damaged enough that keeping only the sites with no missing data would result in an alarmingly low number of samples. I also cut one of the outgroup species, since the morphological data (see below) did not include data for this group. Then I eliminated any site with missing (again, "N") data. Finally, I grouped specimens by their species (the 2 remaining dire wolf samples, 3 gray wolf samples, and 2 African wild dogs) and kept only sites where there was complete among-species agreement between specimens. In this step of treatment, I considered the African golden wolf (*Canis lupaster*) and the golden jackal (*Canis aureus*) as the same species, since this used to be the case before, and I only have morphological data for *aureus*. This is unlikely to change any of the analysis, since their proximity means their positions within the tree containing all other canina should be interchangeable. This left me with a total of 569583 bases for the data set, which still seems pretty high. Since I do not intend to do parsimony analysis, I kept both constant (around 70k bases) and singleton (around 300k) sites. After confusing results during my initial (draft) work, I also ran a topology-only analysis of nuclear data without cutting any specimens besides the 3 dire wolves with a lot of missing data. This led to 621403 bases from 16 specimens spanning 12 species (two foxes and ten canina).

The morphological data will be a subset of the morphological matrix in a previous study of the evolutionary relationships of new world extinct and extant canids [2]. The complete matrix contains 123 scored characters for 132 canids plus an outgroup. This included 9 of our extant (+ *dirus*, henceforth referred to as extant for simplicity) canids plus one of the outgroups (*Urocyon cinereoargenteus*). I also kept the data for 13 extinct species for which we had fossil samples (see below). This left us with 123 characters for 23 species. After removing all characters that had missing data for all of our remaining species, we were left with 104 characters. I again kept singleton and constant characters since I do not intend to use a parsimony analysis. Since I need to partition the data by the maximum state in each character to set up a morphological evolution model, I changed the constant character's score to their maximum score in the original matrix so that they can be considered in the correct "category".

Fossil samples were collected from the paleobiology database by selecting fossils with genus name *Canis*, *Lycaon*, *Cuon* or *Urocyon*. The proposed genera for the dire wolf (*Aenocyon*) and the African jackals (*Lupulella*) were also considered, but contained no samples. I then cut all samples with undefined species names, and all samples with species that had no morphological or nuclear data. This is not necessary, but for the sake of simplicity we will keep it as such for this project. Keeping only one representative of each unique (taxon name, min age, max age) combination. This left me with 119 fossils spanning 22 species (our 23 extant and extinct species, minus *Canis simensis*, the Ethiopian wolf). While we could run FBD with all these samples (remembering to copy the morphological data for samples of a given species), I chose to run a simpler version by keeping only min and max age for each species, as the complete version was taking impractically long to run. Our final fossil data set, then, will be a tsv file representing a table with taxon names, min ages (0 for extant species, minimum fossil age otherwise), and max ages (maximum fossil ages). I changed the minimum age of *Canis mosbachensis* because, due to the way in which stratigraphic uncertainty is recorded, it was set to 0 which could confuse the FBD model into thinking it is an extant species.

2 Methods

Since RevBayes models are constructed under the framework of a graphical model for intuition and modularity, I will show the corresponding graphical model for the model considered at each step when relevant.

2.1 Model selection

To do model selection, I ran model averaging in RevBayes drawing heavily from the model averaging RB tutorial. For the substitution model, I set up five options of Q matrix: Q_{JC} (Jukes-Cantor), Q_{K80} (Kimura 80), Q_{F81} (Felsenstein 81), Q_{HKY} (Hasegawa-Kishino-Yano), and Q_{GTR} (General Time Reversible). I set up a categorical variable Q_i which allowed me to set the model such that $Q = Q_v[Q_i]$, where Q_v is the vector with the aforementioned matrices as elements, and Q_i was a parameter to be estimated (so it would move and therefore change the matrix being used in each generation). I set the transversion rate κ, base frequencies π and exchangeability rates ε by using priors we frequently use, namely letting the latter two always sum to 1 since they represent relative rates (in RevBayes, all Q matrices have relative rates as arguments, e.g. for a JC matrix the non-diagonal row elements are all $1/3$). I also set $\varepsilon_{AG} = \varepsilon_{CT} = 0$, since these data represent transversions only. I then set up model selection on Among-Site Rate Variation (henceforth ASRV) by setting α to an RJMCMC routine jumping between 10^8 (representing no ASRV, since $Gamma(\alpha, \alpha)$ would always yield the same number approximately) and a uniform distribution. I did the same for p_{inv}, the probability of invariant sites. The latter gave a bug on RevBayes after around 10k generations, so I took it out of the model. It seems like a fine choice, however, given this is a SNPs

data set and therefore should not contain invariant sites. The rest of the model was standard - a uniform topology prior and exponential prior for the branch lengths. The following graphical model represents our complete model selection setup (where, here and henceforth in this paper, Δ represents the sequence data, τ a tree topology, β_i the branch length for branch i, N the number of taxa, and $\mathbf{T}$ the phylogenetic tree). Note that the absence of an overall substitution rate parameter (frequently referred to as a "molecular clock") means this tree's branch lengths will not be in units of time, but this has no influence on the selection of best model.

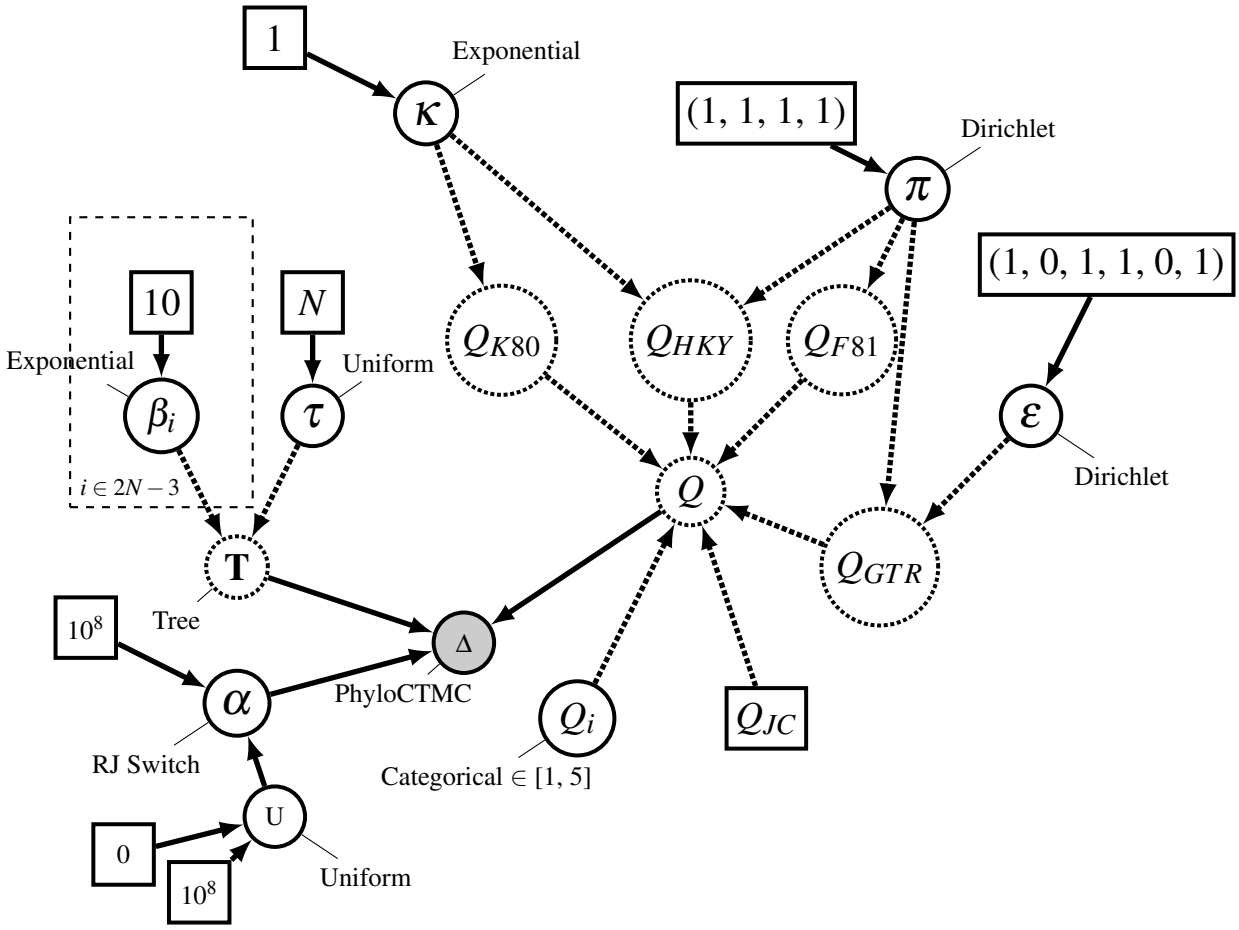


Figure 1: Graphical model for averaging over nuclear substitution models

Here, solid circles are stochastic nodes (parameters to be estimated), solid squares are constant nodes (numbers, matrices etc. that do not change), dashed circles are deterministic nodes (quantities defined deterministically by other nodes), and shaded circles are observed nodes (nodes for which we have data).

For morphological data, my model averaging ended up being full of bugs. I will therefore explain the subset of it that actually worked, though the complete script can be found at the Git repo.

Most of the model was the same, with the topology, branch lengths, and α ASRV setups being equal. Then, for each state sample size (defined here as the maximum state value observed in a given character plus one) S, I 1. Selected the subset of the morphological sites that had maximum state $S-1$; 2. Define ε_s and π_s with an RJ switch similar to α, going between 62000 and a uniform distribution; 3. Defined multipliers m_ε and m_π with a uniform; 4. Got 4 values representing rate and 4 representing frequency categories with a discretized beta using ε_s and π_s as scale and shape; 5. For each category between 1 and 4, defined π and ε as a normalized vector of the discretized beta with shape and scale equal to the corresponding category value from 4 (times the corresponding multiplier m if the scale equals 62000), and then used them to define Q_{JC}, Q_{F81} and Q_{GTR}. In this way we are testing 1. ASRV, 2. JC, F81 and GTR models of rate variation, and 3. Whether separating sites into 4 categories improves the model. When ε_s is 62000, the discretized beta returns approximately equal numbers for the four categories, which is why we use m_ε to account for the scale in that case (and same for π). The separation into categories requires the definition of a vector M of size 4 defining the proportion of sites in each category. We also define Q_i as before to keep track of which Q we are operating under. I did the analysis for each S separately so as to improve efficiency (since we can run all of them in parallel on the cluster). So the following graphical model represents the model for model averaging under each S.

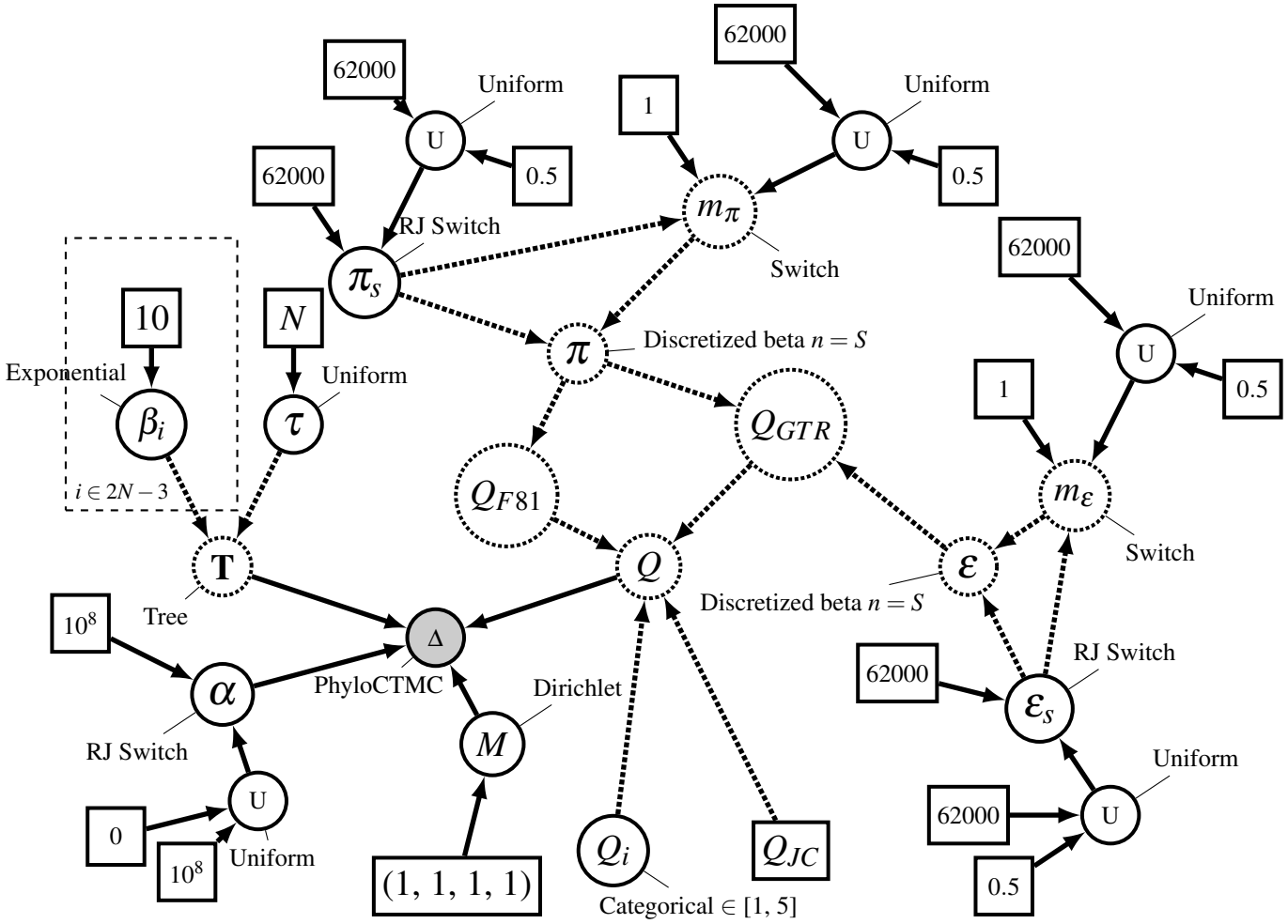


Figure 2: Graphical model for averaging over morphological substitution models with data of equal state sample size.

where "Switch" is how I chose to represent a deterministic switch - when π_s equals 62000 (because of the RJ switch), m_π comes from the distribution; otherwise, it is 1. This is such so that when the scale is high enough that there is no differentiation between categories, leading to all the parameters being equal to 0.5 (the mean of the discretized beta with shape equal to scale), m_π acts as a parameter such that twice the estimate of m_π would be the estimate of scale when categories are not there. Note that the correct thing in this graph would be to have a rectangular box such as the one around branch lengths also enclosing the π and ε parts, since there will be one for each category. But I judged this was already complex enough for one figure, so we assume here one can think of π and ε as SxS matrices such that the rows sum to 1 and the parameters for category i will come from row i of these matrices.

I ran the nuclear model averaging analysis for 5×10^4 generations with 2 runs, sampling for the tree and the numerical parameters every 10 generations. The morphological analysis was the same, except it ran for 10^5 generations. In both cases I set a tuning interval - the interval of generations between tuning of the parameter moves - of 200. All parameters of interest for model selection had high ESS, and visual inspection of the traces indicate well mixed runs.

2.2 Main analysis

From model selection, we conclude that a GTR model with no ASRV was the best model for molecular evolution (see **Results** below). For morphological evolution, things were a bit confusing. The details are again below, but we will proceed with the description of the final analysis taking into account that the results were weird and therefore I elected to restrict my morphological analysis to only binary characters (leaving us with 72 characters still). For this subset, the best model was GTR (which is equivalent to F81 for binary characters). So we will use GTR for molecular evolution, and F81 for morphological evolution.

For the main analysis, I will use a Fossilized-Birth-Death model [7] as the tree prior, with a topological constraint to force the foxes as an outgroup, a GTR model for the molecular evolution, and F81 model for morphological evolution. I will also set overall substitution rates ("clocks") for the molecular and morphological evolution models. They will be branch-specific so as to not force the branches to be connected. Let us go deeper into this complex model.

First, we look over the phylogenetic tree itself. The Fossilized-Birth-Death model is a model for calculating likelihoods of metric phylogenetic trees using the birth-death model with sampling. We set a birth rate λ (the speciation rate), a death rate μ (the extinction rates), and a sampling rate ψ (the fossilization rate). These all have exponential priors with means coming from previous papers or my own analysis. We also need to set a time for the origination of the tree, the root time ϕ, set with a uniform prior between 6 and 30 (the minimum and maximum fossil ages for the whole clade, with some extra). Finally, we set ρ, the sampling probability of an extant lineage, to $10/38$ since we have 10 species out of the 38 extant (+ recently extinct). The FBD tree allows for the setting of topological constraints - which is the only way to set an outgroup since it estimates rooted trees. I make use of this by enforcing all non-fox species to always be a clade, so that we guarantee the foxes are the outgroup. Finally, we set the likelihood of the fossil ages by, for each of the M fossils in the data set, 1. Taking the node age of this taxon, t_i; 2. Taking the minimum and maximum age of this taxon from the data set; 3. Creating a uniform stochastic node F_i with minimum $t_i - b_i$ and maximum $t_i - a_i$; and 4. Clamping F_i as observed at 0. This means F_i will have likelihood 1 when the node age is within the min and max ages for that fossil, and 0 when it is not. The following graph illustrates this model.

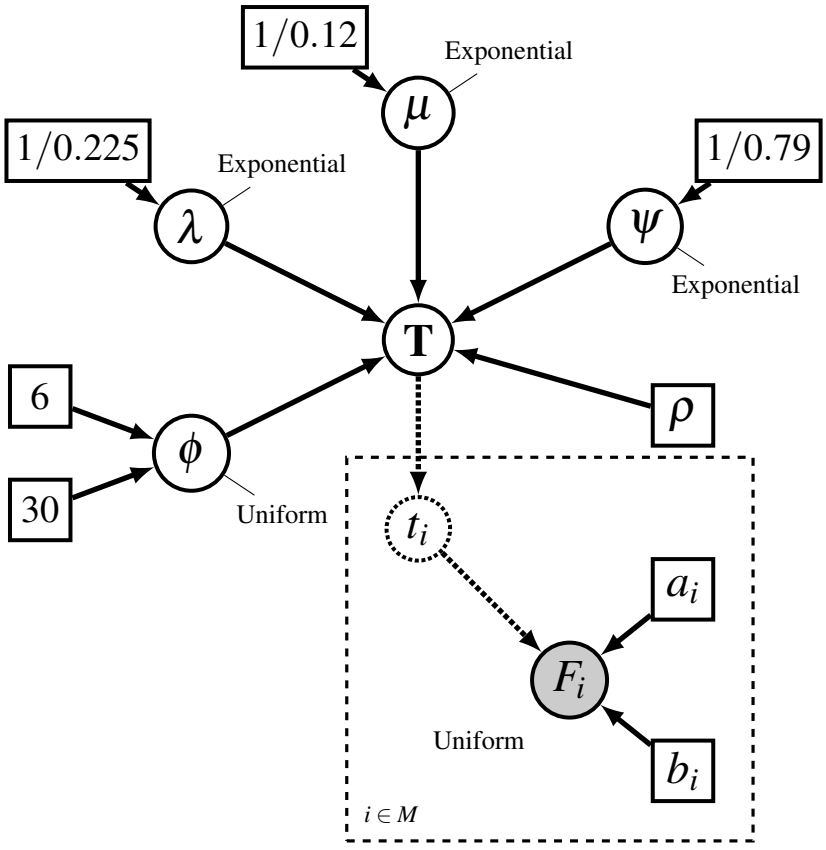


Figure 3: Graphical model for the Fossilized-Birth-Death model.

Note that this model will not always lead to a tree with N tips, since it allows for fossil species to be not only fossil tips (representing the terminal sampling event of an extinct species), but also sampled ancestors (representing a direct ancestor of a sampled species). The model estimates the position of each fossil as a tip or a sampled ancestor jointly with everything else.

Having set the parameters and moves for the phylogenetic tree, we proceed to the molecular evolution model. This will be a simple GTR model with Dirichlet priors on π and ε, since we set them to sum to 1, with the caveat again that the AG and CT rates are 0. We also set branch-specific clock rates by setting a mean $\bar{r}$ to an exponential distribution with mean 0.1, and then setting each branch-specific rate to an exponential with mean $\bar{r}$ as a hierarchical model. This can be represented in the following graphical model.

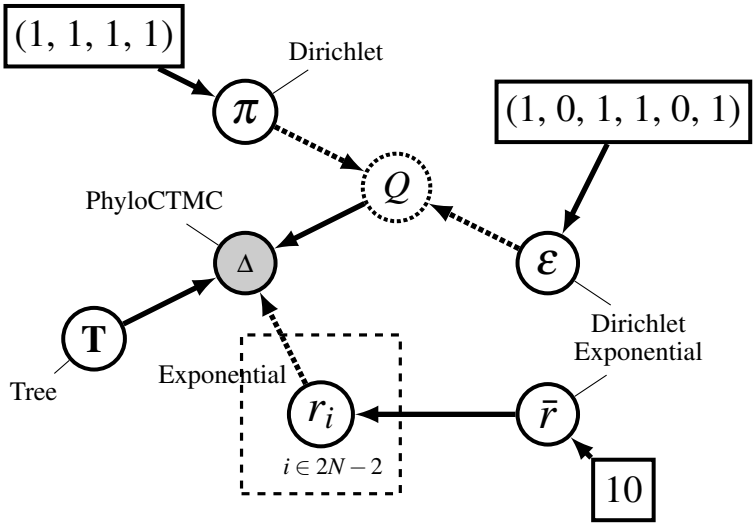


Figure 4: Graphical model for nuclear substitution.

where I omitted the details on the distribution of **T** since those are already present in Figure 3.

The morphological evolution model will be remarkably similar, with the exception that we have no ε parameter (since F81 and GTR are equivalent models for 2-state characters), and our π parameter is defined from a normalized discretized beta with an estimated scale (a fancy way of saying it will be $(x, 1-x)$ for some $x \in (0,1)$). This is equivalent, of course, to using the Dirichlet since in two dimensions the Dirichlet prior will simply choose one rate and its complement with regards to 1. This seems more elegant. I set a lognormal prior (mean 0 standard deviation 1.18) for this scale, for which I have no real justification except it was the prior set on the RevBayes tutorial for a similar analysis. The following graph describes this model.

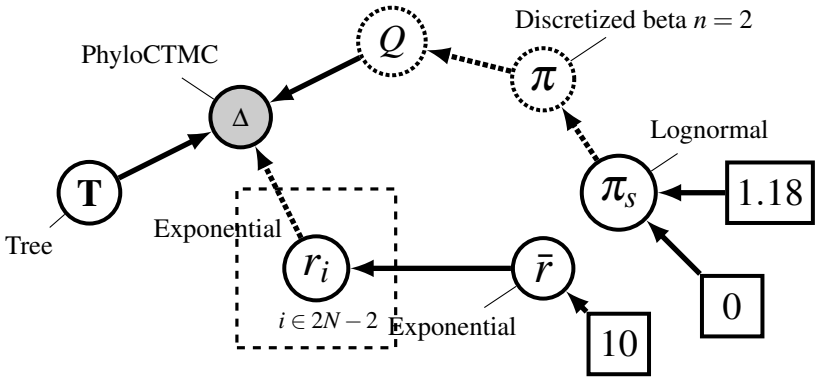


Figure 5: Graphical model for morphological substitution.

which completes the model. The three aspects - fossils, morphology, and DNA - define how the likelihood for a given topology and branch lengths is calculated, and therefore allows for joint estimation of all parameters involved. For improved mixing, I set up more than one move for each parameter, making it so there are low, mid and high window moves. I then set the MCMC to run for 100000 generations, sampling every 10 for trees (including the extant tree, for which I took **T** and pruned the extinct species out) and numerical parameters, and tuning every 200.

For comparisons on how morphological or fossil data could be influencing the results, I also ran a simple nuclear and a simple molecular analysis. These had a uniform prior on the topology, exponential priors on the branch lengths, and the corresponding evolution model described above. The results of these analysis will be useful only for topological comparisons, since they do not estimate any timing-related parameter. The nuclear analysis had good convergence measures after running for 50000 generations, but the morphological one required 1500000, since the morphological data is smaller and less informative. After results found for the draft were counterintuitive, I re-ran the nuclear analysis including all specimens in the molecular data with high enough coverage. I used the exact same model.

3 Results

3.1 Model selection

The results for model selection of molecular evolution models were fairly unambiguous. The MCMC converged really well, and we got high ESS (> 1500, most often >5000) for all parameters of interest. The traces also seem to imply high exploration of parameter space (I thought it would be boring to put a bunch of images of traces here, but one can easily explore them by throwing the log at output_nuclear_avg/nuclear_avg.log on Tracer). The 95% HPD interval for α was $[10^7, 10^8]$, strongly implying that a model without ASRV was the best one. Furthermore, Q_i did not move after burnin, implying the selected model (GTR) was strong enough to keep any model move from being accepted. So we conclude that GTR with no +G is the best model for molecular evolution.

For morphological evolution, the results were unfortunately much more ambiguous and confusing. ESS for some tree branch lengths were low, which might indicate bad topology mixing, but for our model parameters the ESS was always above 5000. For 2-state characters, Q_i seems to strongly favor GTR, which is surprising given GTR and F81 are the same model in 2-state characters. Furthermore, the indicator variables for α and the categories seem to indicate that a model including ASRV and categories would be the best one, but the 95% HPD interval for both π_s and ε_s indicate very small differences between categories in the best model, the interval for α was again $[10^7, 10^8]$, which indicates no ASRV. I am unsure how to interpret these results, as they were very counterintuitive and seem to look weird for the model at hand. While I initially thought this might be due to my unorthodox setup for model averaging in RevBayes, later results of the analysis (see below) made me doubt the validity of the application of this morphological data set in general (see **Discussion** for more details). Given all these caveats, I interpreted the results as best as I could and chose GTR with no ASRV or categories as the best model for the data.

3.2 Main analysis

The main FBD model unfortunately does not seem to have converged after 100000 generations, so I will discuss the results given this but we must keep that in mind for now. The ESS for some branch lengths, node ages, and molecular clocks are under 100, and traces seem to not have adequately explored parameter space, including the posterior. I also ran a 300k and 1.2M generations versions, but the uncertainty remains high which might imply we do not have enough data to lower this uncertainty.

All this said, we can evaluate the results we have. The following is the Maximum Clade Credibility (MCC) tree extracted from treeAnnotator

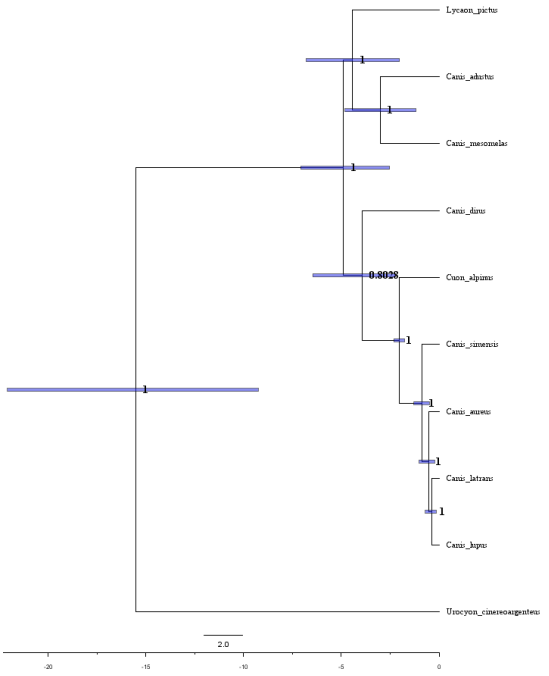
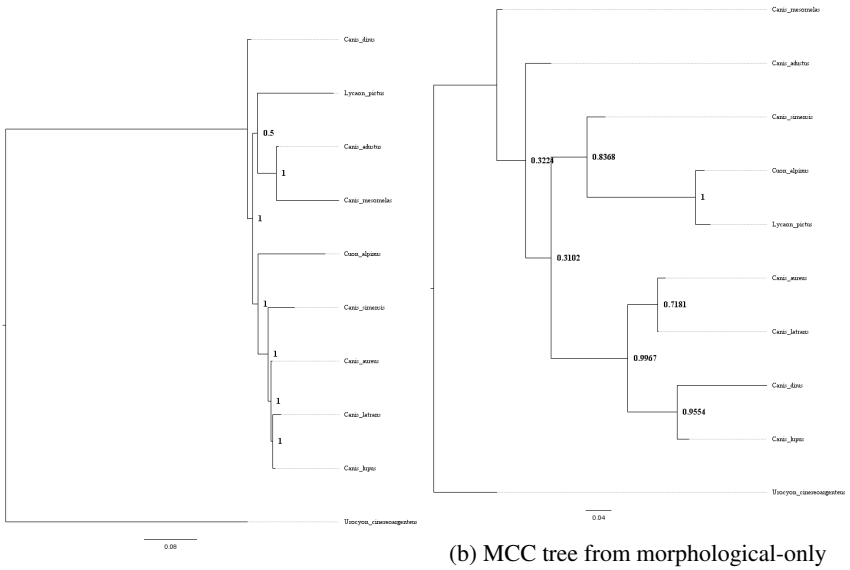


Figure 6: MCC tree from FBD combined-evidence analysis. Node bars represent node age uncertainty (my), and node labels represent posterior probability of nodes.

The trees from the nuclear- and morphology-only analyses might help shed light on some interesting differences between this tree and the one on the original paper.



(a) MCC tree from nuclear-only analysis.

(b) MCC tree from morphological-only analysis

Figure 7: MCC trees from simpler analyses. Node bars represent node age uncertainty (my), and node labels represent posterior probability of nodes.

and, maybe even more interesting, the tree from the nuclear-only analysis containing all specimens can further aid in those considerations.

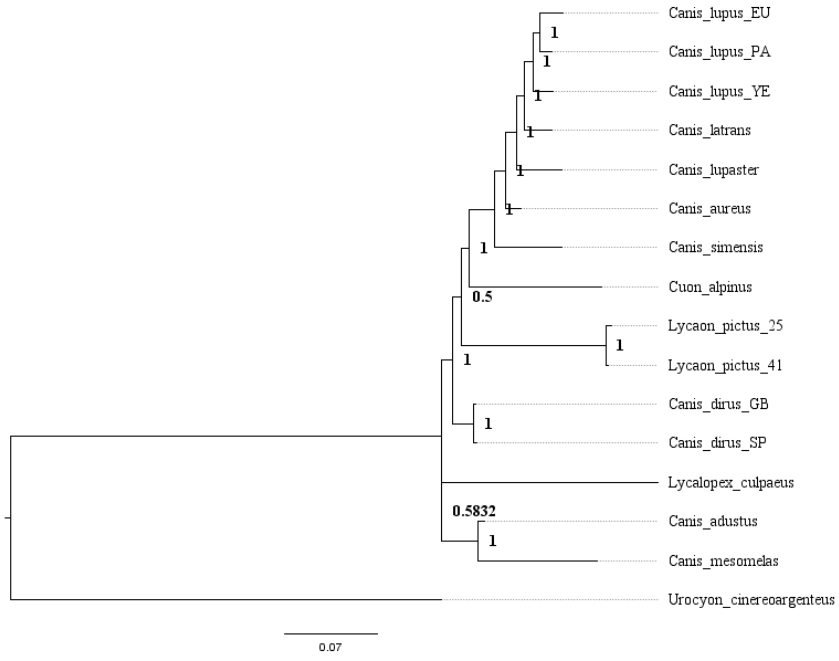


Figure 8: MCC tree from nuclear-only analysis with all 16 specimens from 12 species.

Note that the *culpaeus*-(*adustus*, *mesomelas*) clade is not a polytomy, rather a clade with very small branch length between its divergence and the divergence of *culpaeus*. We will discuss more on this later.

Let C_J be the clade of all extant (+ dire wolf) canina present in the tree minus the dire wolf, and C_D as the clade of all these plus the dire wolf, minus the jackals. Then for the combined-evidence, $P(C_D) = 0.2$ and $P(C_J) = 0.25$, so that the jackals have a higher posterior probability of having diverged before. For the nuclear analysis, $P(C_D) = 1$ and $P(C_J) = 0$, so that after burnin all topologies had the dire wolf diverging first. Finally,

for the morphological-only model, $P(C_D) = 0$ and $P(C_J) = 0.31$, so while less certain it is clear that the posterior topologies favor jackals diverging before the dire wolf. For the FBD analysis, we also kept track of the divergence time of dire wolfs and jackals, showing slightly higher preference for the later divergence of the dire wolf, though they are still very similar (same 95% HPD interval, means differing by only 0.035my).

4 Discussion

Figure 3 tells a sad story indeed, when combined with figure 7 - in my attempt to reduce uncertainty in the topology by combining morphological and molecular data, I instead increased it. There are a couple of interesting aspects to explore regarding that, so let us go through them.

The first and most glaring discrepancy between the nuclear and morphological trees is the position of *dirus*. We discussed above how the dire wolf is morphologically incredibly similar to the gray wolf, having been considered by some a sampled ancestor or sister species. Morphological analyses, then, can be expected to pair them together. Nuclear data seems to disagree strongly, putting *dirus* as the farthest or second farthest group from the gray wolf. We could interpret the combined-evidence tree's placement of *dirus*, then, as a "compromise" between the two, yielding a tree that does not assist in reducing the uncertainty in its placement. The competing likelihoods of our nuclear and morphological data might be pushing the high posterior probability topologies to weird spots in tree space, leading to relationships that are supported by neither parts of the data alone. Note that the full nuclear analysis complicates this even further, by agreeing with the combined-evidence analysis on the placement of *dirus* but adding further confusion on the placement of the jackals. Furthermore, the low ESS in many topology-related parameters (some branch lengths and node ages) are further clues as to the difficulty of these data in resolving these relationships. While some low ESS are expected (such as the branch lengths, since FBD analyses collapse and expand branches frequently to account for sampled ancestors, leading to complex jumps in tree space), other aspects of tree summaries can be interpreted as dangerously high uncertainty - the MCC tree from 1.2 million generations still contains negative branch lengths when nodes ages are decided by average node age, for example. This is a marker of high uncertainty in the topology. Combined-evidence analysis may be a powerful tool, but confounding processes such as convergent evolution might drive this analysis to yield higher uncertainty and incorrect topologies. See **Conclusions** for my thoughts on how to proceed with this knowledge.

Some words are necessary to address the topological differences between my nuclear trees and all trees in the original paper. First, the displacement of *Lycaon pictus* from a sister group of the following canina, to a sister group of the jackals. It is not actually a huge leap, since most other phylogenies I found in the literature either place it as the group diverging immediately before the jackals, or immediately after (such as the dire wolf sequencing paper). It is something interesting to investigate, though, since it indicates that either my model or treatment of the data led to differing results from the original paper, which could be influencing my final data as well. As seen in figure 8, this discrepancy was resolved with the inclusion of the missing specimens, which probably means that keeping only bases where the two *pictus* specimens differed originally might have erased important variation. The low posterior probability of the node leading to the clade of *pictus* and the others also implies it might have jumped around a lot in the analysis, showing that this data might indeed have trouble with its placement. Note that the full nuclear-only analysis has a discrepancy with the original tree itself - the placement of *Lycalopex culpaeus* in a monophyletic clade with the jackals. This is a grave mistake, given we know that *culpaeus* is a fox and canina is a monophyletic clade. The low posterior probability and branch length of that node shows how this was a difficult part of the tree for the analysis to assess, however, which leads me to believe that including more bases in the analysis (which would imply including missing data) might resolve it. Another option would be to enforce clade constraints on the non-foxes being a monophyletic clade, but it is usually good form to avoid enforcing constraints as much as necessary.

The second consideration I wanted to discuss is my current belief, after spending around 30 hours staring at/writing scripts for/reading about this data, that the morphological data set I chose does not work for these analyses. The first clue of this is the insistence of some branch lengths on the morphological tree not to converge, even after running the analysis for 10000000 iterations. Some attempts of morphological-only and combined-evidence analysis also yielded negative branch lengths on the MCC tree, a sign that some "probable" relationships actually had high uncertainty. A more critical appraisal of the data set post-hoc may explain this - from my 72 binary characters, only 41 are non-constant, and only 12 of those are non-singleton. Furthermore, these characters were scored for a data set containing 132 extant and extinct canids, so they are on average reasonably general. This might muddle phylogenetic signal even further. The unsatisfying conclusion here is that this data set might simply not be enough to resolve some relationships in this tree, leading to the unconventional (and just plain wrong in some cases) relationships in the morphological-only tree (such as the *latrans-aureus* clade).

5 Conclusion

The most important takeaway from this project is that more data does not always equal better signal. While combined-evidence analysis might be music to a researcher's ear when trying to lower topological uncertainties, confounding processes such as convergent evolution might lead to compromise trees that are not more informative to the individual's question. This is similar to the interaction between different evolving genes that leads to the necessity of the coalescent model or partitioned analysis - forcing genes to evolve together might pull the likelihood in separate directions, leading to a compromise. Furthermore, the analysis is full of clues that these data are not enough to reduce uncertainty in the relationships I am interested in, but rather they increase uncertainty. There are a couple steps I would like to take - for when I expand this project into a publication, hopefully - to resolve these and the other issues listed above.

First, I want to investigate the discrepancies between nuclear-only analysis between my models and the paper's. If I can not find a way for them to agree it is possible one of the models is simply worse, but I suspect the differing data treatments actually led to these discrepancies. Furthermore, adding non-SNP nuclear data to my analysis might lower uncertainty given the known issues of using SNP data for phylogenetic inference [8]. Then, I want to look for a more informative morphological data set. The paper in question collected a series of morphometric landmark data on mandibule and molar morphology for the taxa in question that might help me with that, since these are more specific characteristics that could vary more strongly among closely related taxa. Finally, even with all these issues solved I do not believe a direct combined-evidence analysis is the way to go for these taxa. While my analyses show deeper issues, the contradictions in topology between nuclear and morphological trees will always exist, particularly when it comes to the dire wolf (and, to a lesser extent, to the African wild dog). As such, a more nuanced approach might be necessary to use the morphological data on an FBD model with nuclear data. One possibility is a gene tree-species tree framework, where the morphological and nuclear trees are considered gene trees (with differing weights), allowing for inference while explicitly considering their contradictions. Another idea is to use the nuclear data - which we can be more confident on due to its more objective nature, though ascertainment bias could still be an issue with SNP data - to create a backbone tree, and then use morphological data and fossils to lower uncertainty on specific relationships and divergence times. I suspect this will always bias the final tree on the direction of the dire wolf getting closer to the wolf, but not sure how to escape that.

In any case, one thing is clear: while this project was not successful in its end goal, it was a great example of the shortcomings of combined-evidence analysis, and allowed for interesting discussion on the flaws of common phylogenetic inference methods in general. I believe this made the effort worthwhile, even if it will make my life harder when I want to publish. Some of the most interesting discoveries in the history of our wondrous exploration of the tree of life started out by making someone's life harder, after all.

References

[1] I. V. Caldas and C. G. Schrago, "Data partitioning and correction for ascertainment bias reduce the uncertainty of placental mammal divergence times inferred from the morphological clock," *Ecology and Evolution*, vol. 9, no. 4, pp. 2255–2262, 2019. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ece3.4921.

[2] G. J. Slater, "Iterative adaptive radiations of fossil canids show no evidence for diversity-dependent trait evolution," *Proceedings of the National Academy of Sciences*, vol. 112, pp. 4897–4902, Apr. 2015. Publisher: National Academy of Sciences Section: Biological Sciences.

[3] J. C. Merriam, "Note on the systematic position of the wolves of the Canis dirus group," *Bull. Dept. Geol. Univ. California*, vol. 10, pp. 531–533, 1918.

[4] A. R. Perri, K. J. Mitchell, A. Mouton, S. Álvarez Carretero, A. Hulme-Beaman, J. Haile, A. Jamieson, J. Meachen, A. T. Lin, B. W. Schubert, C. Ameen, E. E. Antipina, P. Bover, S. Brace, A. Carmagnini, C. Carøe, J. A. Samaniego Castruita, J. C. Chatters, K. Dobney, M. dos Reis, A. Evin, P. Gaubert, S. Gopalakrishnan, G. Gower, H. Heiniger, K. M. Helgen, J. Kapp, P. A. Kosintsev, A. Linderholm, A. T. Ozga, S. Presslee, A. T. Salis, N. F. Saremi, C. Shew, K. Skerry, D. E. Taranenko, M. Thompson, M. V. Sablin, Y. V. Kuzmin, M. J. Collins, M.-H. S. Sinding, M. T. P. Gilbert, A. C. Stone, B. Shapiro, B. Van Valkenburgh, R. K. Wayne, G. Larson, A. Cooper, and L. A. F. Frantz, "Dire wolves were the last of an ancient New World canid lineage," *Nature*, vol. 591, pp. 87–91, Mar. 2021. Number: 7848 Publisher: Nature Publishing Group.

[5] C. Zhang, T. Stadler, S. Klopfstein, T. A. Heath, and F. Ronquist, "Total-Evidence Dating under the Fossilized Birth–Death Process," *Systematic Biology*, vol. 65, pp. 228–249, Mar. 2016.

[6] F. Ronquist, S. Klopfstein, L. Vilhelmsen, S. Schulmeister, D. L. Murray, and A. P. Rasnitsyn, "A Total-Evidence Approach to Dating with Fossils, Applied to the Early Radiation of the Hymenoptera," *Systematic Biology*, vol. 61, pp. 973–999, Dec. 2012.

[7] T. A. Heath, J. P. Huelsenbeck, and T. Stadler, "The fossilized birth-death process for coherent calibration of divergence-time estimates," *Proceedings of the National Academy of Sciences*, vol. 111, pp. E2957–E2966, July 2014.

[8] E. J. McTavish and D. M. Hillis, "How do SNP ascertainment schemes and population demographics affect inferences about population history?," *BMC Genomics*, vol. 16, p. 266, Apr. 2015.