

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

EEOB 563 Project

Bruno do Rosario Petrucci

March 29, 2021

## 1 Background

The dire wolf (*Canis dirus*) has long been a source of wonder for researchers and the general public alike. While much work has been put into the study of its fossil remains, many have put into question whether its morphological similarities with the grey wolf (*Canis lupus*) are a result of close evolutionary relationships, or convergence. This is a difficult question to address, given extinct species rarely leave remains that allow for the sequencing of genetic data, which would be necessary for the construction of phylogenetic trees that could help untangle the dire wolf's relationship with the rest of the wolf-like canids (i.e. the genera *Canis*, *Cuon*, and *Lycaon*). Recent work has sequenced the dire wolf DNA for a number of well-preserved specimens, however, and put us one step closer to understanding these complex relationships [1].

In this study, it became clear the dire wolf was not phylogenetically close to the other representatives of the *Canis* genus, with the dhole (*Cuon alpinus*) and the African wild dog (*Lycaon pictus*) sharing a more recent common ancestor with the wolf than the dire wolf and the jackals clade (*Canis mesomelas* and *Canis adustus*). The question of whether the latter clade should be in the *Canis* genus is also a frequently debated one, with the genus *Lupulella* being increasingly adopted for that group instead. These results supported that conclusion, and furthermore provided support to the hypothesis that the dire wolf should not be on the *Canis* genus either. The molecular data in the study failed, however, to satisfactorily resolve the relationship between *dirus*, the (*mesomelas*, *adustus*) group and the rest of the wolf-like canids, and it is not clear whether the dire wolf or the jackals diverged earlier in the wolf-like canids group.

Previous work has shown integrating morphology and fossil sampling data into phylogenetic inference can improve divergence time estimations (e.g. [2]), so that total-evidence analysis is a possible path to untangle these relationships. In this project, I intend to make use of the Fossilized-Birth-Death process [3] in RevBayes [4] to attempt to resolve the divergence order of the basal wolf-like canids. I will utilize nuclear DNA sequence, morphology data, and fossil samples from extant and extinct canids from the *Canis*, *Cuon* and *Lycaon* genera.

## 2 Data

I intend to use the nuclear genetic data from the original dire wolf paper [1]. It required correspondence with the authors, since the data is not published, but I have acquired the final alignment used on the paper, with 5 specimens of dire wolf, 3 specimens of grey wolf, and 1 specimen of each other of the 11 species total used in the nuclear DNA analysis. This data contains transversions only, which must be taken into account when selecting a substitution model. This data contains around 12 million SNPs, and the dire wolf DNA is highly damaged, so that some treatment will be required. I will clean the data set by deleting every site that does not appear in all dire wolf specimens. If this leads to too low a number of SNPs (<10 thousand non-singleton sites), I will instead concatenate the dire wolf specimens all in one genome, resolving conflicts by random sampling. Either should work fine, given the study mentions the specimens of dire wolves formed a monophyletic lineage under all models, and this is unlikely to change when we include morphological data. Since this was also the case for the grey wolves, I will select one of the 3 specimen at random to serve as the representative in the analysis. As such, we will end up with a complete alignment for nuclear DNA data for the 11 species in the study, with no missing data. I might also delete singleton sites, given their pipeline (see SI for the study), but not sure yet since I am still in correspondence with the authors to understand the data better. Finally, I will cut one species (*Lycalopex culpaeus*) due to missing morphological data (see below).

For the morphological data, I intend to use the data matrix compiled by Graham Slater for a select group of extinct and extant canids [5]. This is a scoring of 123 characters for a group of canids that includes 10 species in our data set, the one left out being the andean fox (see above). For the identity of each morphological character, see the SI for the aforementioned study. They focus on mandible and teeth morphology, though span a reasonable range of characters that are important for canid phylogenetics. This comes with a small caveat - the data set includes *canis aureus*, not *Canis lupaster*. Since the latter used to be considered a subspecies of the former, and their position on the wolf-like canid phylogeny is interchangeable (i.e. a WLC phylogeny without *aureus* but with *lupaster* will have *lupaster* at the same place as *aureus* in the analogous tree without *lupaster* but with *aureus*), I judged it fair to use the *Canis aureus* morphological data with the *Canis lupaster* nuclear data. Similarly, I will use the *Cuon javanicus* morphological data as the data for *Cuon alpinus*, since the former is now considered a subspecies of the latter. Both of these would require more justification for a published study, of course. I further treated the data by deleting characters for which we had the same value for all species in the data set, since . This leaves us with 60 characters for the analysis.

For fossil sampled, I will use the paleobioDB package to find fossil occurrences from the paleobiology database. I will first find all samples with genus *Urocyon*, *Canis*, *Cuon*, and *Lycaon*. Then, I will only keep fossil samples for which the species name is accepted, to make sure we are not included misidentified genera in the analysis. Then, I will maintain only one fossil sample with a unique combination of taxon name, minimum age, and maximum age. This is due to the FBD model, where we cannot include more than one fossil sample with the same species and age range. I have not yet done this step, but from working previously with the canid fossil data, I predict we

will end with around 300 fossil samples to use for the data. The FBD model estimates, jointly with the topology, branch lengths and divergence times, the posterior probability that a fossil is a sampled ancestor (i.e. a fossil from a species that descended into one of the extant species) or an extinct tip.

### 3 Methods

As mentioned, I will use the FBD model in RevBayes for this analysis. This model jointly estimates the posterior probability of a set of topology, branch lengths, divergence times, sampled ancestor sampling times, and fossil tips, from a set of sequence, morphology, and fossil occurrences data. This is a highly complex model, leading to a high number of parameters in comparison to the data. As such, I plan to use simpler molecular evolution models than usual, with low/no partitioning. I am not set on the details of the analysis yet. I will run a simple analysis on the nuclear data alone to get some tree priors (e.g. trees with > 90% posterior probability). From this analysis, I will also test the tree's sensitivity to priors of diversification-sampling rates, molecular evolution, and topology. Given this knowledge, I will run a simple model selection analysis using BIC and/or posterior probabilities. The models and tree prior selected here will then be used on a full estimate, using the morphology and fossil data. The results of this analysis will be summarized with consensus trees, MAP and MCC trees, and credible intervals of divergence times for the speciation times relevant to our phylogenetic question. As mentioned, the details of this analysis might change as I proceed in the project, since FBD is a complex model and it might take some figuring out to set it up such that we get a tree with low enough uncertainty to answer our question.

### References

- [1] 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.
- [2] C. Zhang, T. Stadler, S. Klopstein, T. A. Heath, and F. Ronquist, "Total-Evidence Dating under the Fossilized Birth–Death Process," *Systematic Biology*, vol. 65, pp. 228–249, Mar. 2016.

- [3] 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.
- [4] S. Höhna, M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck, and F. Ronquist, “RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language,” *Systematic Biology*, vol. 65, pp. 726–736, July 2016.
- [5] 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.