#Nowhere to run, nowhere to hide: carnivoran extinction in the Holocene

Orlin S. Todorov, John Alroy, Who else?

Affiliations:

Macquarie University, School of Natural Sciences

# Abstract

# Introduction

Carnivoran evolution and extinction - general

Mammalian extinction in the Holocene zoning into carnivorans

Why carnivores are a good model?

Variables used in the context of carn extinction (diet vs locomotion, brain, body size, longevity, abundance, home range etc)

Hypotheses

# Materials and Methods

## Data availability

All data, code and phylogenetic trees used for the analyses in this manuscript are available as electronic supplement and on Github at <https://github.com/orlinst/Carnivores-ext> [still not public]

## Sample

We collated a sample of 114 terrestrial carnivoran species above XX kg (excluding pinnipeds) available in the Ecoregister (http://ecoregister.org). Four variables of interest were obtained by averaging the observed values per species for 934 observations over 195 sites from all continents - body mass, home range, basal metabolic rate (BMR), diet. All sites were dated from Xya to Xya (covering the late Pleistocene and Holocene) and the final sample contained 11 extinct species (9.7%), and 103 extant species (90.3%) – a coverage of ~42% of all extant species in the order (excluding Pinnipedia). Additionally, we used information on which biogeographic region the species were found from the Ecoregister.

Moreover, we collated data on brain size, EQ, locomotor type, activity period, social complexity, group size, gestation length, interbirth interval, weaning age, and litter size from literature sources (see TABLE X for details and sources) and calculated abundance, and Shannon and Simpson indices of diet diversity based on primary data (scat and camera-traps) from the Ecoregister. Scat categories in the Ecoregister include traces of vertebrates (mammals, birds, reptiles), invertebrates (arthropods, arachnids, insects), fruit, plant, aquatic (annelid, mollusc) and fungi. They were obtained from 134 observations from 103 different sites, based on the composition of between 12 and 3878 scats per site. Scat composition was subsequently recategorized using k-means clustering, resulting in 6 different diet categories: 1-6…[describe]

Shannon and Simpson indices of diet (diversity of food sources in the scat) were calculated using the package vegan (see Table X). Species’ abundance was calculated using data from camera-traps available in the Ecoregister using the following formula:

(count – number of sites) / (count / count per day) [CHECK THIS]

Data are censored at 1 photograph per study – in order to remove censorship we use a standard approach of ….. . This makes it possible for an abundance to approach 0.

<the number of sites is the min number of photos needed>

Furthermore, we obtained the first principal component (PC) derived from a principal component analyses (PCA) based on the following variables obtained from Noonan et al. (2015): social class, natal den, hibernaculum, predation, food storage, residence, fossorial propensity, primary, secondary, occupant, and burrowing class (for detailed description of these variables see Noonan et al. (2015)). The so obtained PC is related to burrowing behaviour and fossoriality as indicated by the highest positively loaded variables - fossorial propensity (0.68) and burrowing class (0.53). We chose only the first PC explaining ~53% of the variance, as there was a steep decline in explained variance into the subsequent principal components (PC2 – 23% proportion of variance explained, PC3 - 12%, and PC4 – 5%).

Due to high collinearity we used only EQ and not brain volume, and for the same reason BMR was transformed to ‘residual BMR’ after a phylogenetic regression (pgls) on body size, and only the residuals were used in further analyses. A correlation table of all variables can be found in the supplement (Table XXX).

All continuous variables were natural-log transformed and min-max normalised before imputation. Abundance was Yeo-Johnson transformed due to many 0 values.

The resulting dataset contained 25 variables in total. Detailed description of all variables used, including sources, distribution (along with distribution of the imputed datasets) is available in Table X.

Table X – data sources and detailed description of variables, including histograms of the distribution of the original dataset, and density plots of the 20 imputed datasets (red lines) and the observed original data (blue line). NM = number of missing species (out of 114)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | Units | Description | Distribution | Imputation | Source | NM |
|  |  |  |  |  |  |  |
| Body mass | kg | Body mass |  |  | Ecoregister | 10 |
| Home Range | km2 | ? |  |  | Ecoregister | 45 |
| Brain | mm3 | Endocranial volume |  |  | (Michaud et al., 2022) | 28 |
| EQ | Index | Derived using CEQ formula |  |  | (Chambers et al., 2021) | 49 |
| PC | PC score | See methods for details |  |  | (Noonan et al., 2015) | 46 |
| BMR | ml O2 hr-1 |  |  |  | Ecoregister | 0 |
| Litter size | number | Number of offspring per litter |  |  | (Noonan et al., 2015) | 34 |
| Shannon index of diet | Index |  |  |  |  | 46 |
| Simpson index of diet | Index |  |  |  |  | 46 |
| Abundance | Frequency of camera trap observation | See methods for details |  |  |  | 34 |
| Longevity | years | Maximal longevity |  |  | (Noonan et al., 2015) | 49 |
|  |  |  |  |  |  |  |
| **CATEGORICAL** |  |  |  |  |  |  |
| Extinct or Extant | 0 – extinct, 1 - extant |  |  |  | Ecoregister | 0 |
| Locomotor mode | 6 categoris | A – arboreal  R – semiarboreal  SA – semiaquatic  SC – scansorial  SF – semifossorial  T - terrestrial |  |  | (Gálvez-López & Casinos, 2022) | 0 |
| Diet | 6 Categories | Based on k-means clustering of scat data |  |  | Ecoregister | 46 |

## Missingness and Multiple imputation

We imputed 20 datasets following a recommendation for imputation of as many datasets as the percentage of missing data in the original dataset (~19% in our case) (White et al., 2011). The variable with the highest number of missingness was longevity (43% or 49 missing values). We analysed the phylogenetic signal in the missingness using the D statistic and additionally, the overall pattern of missingness was shown to not be missing completely at random (MCAR) (Little’s MCAR test = 1190, df=1022, p<0.001 (Little, 1988)). All continuous variables (natural log transformed and min-max normalised) were imputed using Predictive Mean Matching with distance aided selection of donors (PMM with midas touch algorithm), all categorical variables with two levels were imputed using logistic regression, and all categorical variables with more than two levels were imputed using Polytomous logistic regression (Polyreg) (Buuren & Groothuis-Oudshoorn, 2011; Rubin, 1987). Imputation was run over 100 iterations using the package mice (Buuren & Groothuis-Oudshoorn, 2011). Detailed description of the missingness pattern and imputation analyses can be found in the Supplementary material. The distribution of the 20 imputed datasets is shown in Table X.

## Phylogeny

We used the phylogeny provided by PHYLACINE (version 1.2.1) comprising 1000 trees of 111 of the 114 carnivoran species in our sample (Faurby et al., 2018). The 3 missing species were manually added to all 1000 trees as follows: *Martes caurina* as sister species to *Martes americana*, *Conepatus robustus* as sister to both *Conepatus chinga* *and Conepatus leuconotus*, and *Felis lybica* as sister to *Felis silvestris*.

## Statistical analysis

All statistical analyses were performed in R 4.2.2 (R Core Team, 2021) using RStudio 2022.12.0 Build 353. All packages used are listed in TABLE X. All phylogenetic analyses were conducted on the 20 imputed datasets and were run across all 1000 trees.

Table X Packages used in the analyses and data processing.

|  |  |  |
| --- | --- | --- |
| **Package name** | **Version** | **Use** |
| ape (Paradis & Schliep, 2019) | 5.6.2 | Phylogenetic analyses |
| BAMMtools (Rabosky et al., 2014) | 2.1.10 | Tree manipulation |
| caper (Orme, 2012) | 1.01 | Estimation on D phylogenetic signal |
| dplyr (Wickham, 2020) | 1.0.10 | Various data manipulation |
| easystats (Lüdecke, 2022) | 0.5.2 | Descriptive statistics and plotting |
| ggplot2 (Wickham, 2016) | 3.3.6 | Plotting |
| ggtree (Yu et al., 2017) | 3.4.4 | Phylogenetic tree visualisations |
| lattice (Sarkar et al., 2015) | 0.20.45 | Plotting |
| lavaan (Rosseel et al., 2017) | 0.6.12 | Structural Equation Modelling |
| mice (Buuren & Groothuis-Oudshoorn, 2011) | 3.14.0 | Multiple imputation |
| MCMCglmm (Hadfield, 2010) | 2.34 | MCMCglmm |
| mulTree (Guillerme & Healy, 2014) | 1.3.7 | Incorporating multiple trees in MCMCglmm |
| naniar (Tierney et al., 2019) | 0.6 | Analyses of missing data |
| phylolm (Ho et al., 2016) | 2.6.2 | Phylogenetic logistic regressions |
| phytools (Revell, 2012) | 1.2.0 | Various phylogenetic data manipulation |
| psych (Revelle, 2015) | 2.2.9 | Descriptive statistics and plotting |
| semTools (Jorgensen et al., 2016) | 0.5.6 | Plotting SEM |
| vegan (Oksanen et al., 2013) | 2.6.4 | Calculating of Shannon and Simpson index |

## Analyses

## Initially we performed liner discriminant analyses in order to identify the orthogonal variables in our dataset, which we tested in a separate model. Additionally, we tested X other models, as follows:

1.

2.

3.

## We ran a custom function incorporating phylogenetic logistic regressions using phylolm on each model on all 20 imputed datasets and 100 randomly selected trees from the 1000 available. The results were pooled and averaged, and the resulting p-values and estimates were plotted.

# Results

# Discussion

# References

Buuren, S. v., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations inR. *Journal of Statistical Software*, *45*(3), 1-67. <https://doi.org/10.18637/jss.v045.i03>

Chambers, H. R., Heldstab, S. A., & O’Hara, S. J. (2021). Why big brains? A comparison of models for both primate and carnivore brain size evolution. *PLoS ONE*, *16*(12), e0261185. <https://doi.org/10.1371/journal.pone.0261185>

Faurby, S., Davis, M., Pedersen, R. Ø., Schowanek, S. D., Antonelli1, A., & Svenning, J.-C. (2018). PHYLACINE 1.2: The Phylogenetic Atlas of Mammal Macroecology. *Ecology*, *99*(11), 2626-2626. <https://doi.org/https://doi.org/10.1002/ecy.2443>

Gálvez-López, E., & Casinos, A. (2022). Scaling pattern of the carnivoran forelimb: Locomotor types, differential scaling and thoughts on a dying similarity. *bioRxiv*, 2022.2006.2029.498091. <https://doi.org/10.1101/2022.06.29.498091>

Guillerme, T., & Healy, K. (2014). mulTree: a package for running MCMCglmm analysis on multiple trees. *Zonodo*. <https://doi.org/10.5281/zenodo>. 12902

Hadfield, J. D. (2010). MCMC Methods for Multi-Response Generalized Linear Mixed Models: TheMCMCglmmRPackage. *Journal of Statistical Software*, *33*(2), 1-22. <https://doi.org/10.18637/jss.v033.i02>

Ho, L. S. T., Ane, C., Lachlan, R., Tarpinian, K., Feldman, R., Yu, Q., van der Bijl, W., Maspons, J., Vos, R., & Ho, M. L. S. T. (2016). Package ‘phylolm’. *See* [*http://cran*](http://cran)*. r-project. org/web/packages/phylolm/index. html (accessed February 2018)*.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., Rosseel, Y., Miller, P., Quick, C., Garnier-Villarreal, M., Selig, J., Boulton, A., & Preacher, K. (2016). Package ‘semTools’. In.

Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, *83*(404), 1198-1202. <https://doi.org/10.1080/01621459.1988.10478722>

Lüdecke, P., Ben-Shachar, Wiernik, & Makowski. (2022). easystats: Framework for Easy Statistical Modeling, Visualization, and Reporting. *CRAN. Available from* [*https://easystats.github.io/easystats/*](https://easystats.github.io/easystats/).

Michaud, M., Toussaint, S. L. D., & Gilissen, E. (2022). The impact of environmental factors on the evolution of brain size in carnivorans. *Communications Biology*, *5*(1), 998. <https://doi.org/10.1038/s42003-022-03748-4>

Noonan, M. J., Newman, C., Buesching, C. D., & Macdonald, D. W. (2015). Evolution and function of fossoriality in the Carnivora: implications for group-living [Original Research]. *Frontiers in Ecology and Evolution*, *3*. <https://doi.org/10.3389/fevo.2015.00116>

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O’hara, R., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Wagner, H. (2013). Package ‘vegan’. *Community ecology package, version*, *2*(9), 1-295.

Orme, C. D. L. (2012). The caper package: comparative analyses in phylogenetics and evolution in R. 1-36. <https://cran.r-project.org/web/packages/caper/vignettes/caper.pdf>

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526-528. <https://doi.org/10.1093/bioinformatics/bty633>

R Core Team. (2021). R: A language and environment for statistical computing. In: Vienna, Austria: R Foundation for Statistical Computing.

Rabosky, D. L., Grundler, M., Anderson, C., Title, P., Shi, J. J., Brown, J. W., Huang, H., & Larson, J. G. (2014). BAMM tools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution*, *5*(7), 701-707.

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217-223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>

Revelle, W. (2015). Package ‘psych’. *The comprehensive R archive network*, *337*, 338.

Rosseel, Y., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., & Barendse, M. (2017). Package ‘lavaan’. *Retrieved June*, *17*(2017), 1.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. <https://doi.org/10.1002/9780470316696>

Sarkar, D., Sarkar, M. D., & KernSmooth, S. (2015). Package ‘lattice’. *Version 0.20*, *33*.

Tierney, N., Cook, D., McBain, M., Fay, C., O'Hara-Wild, M., & Hester, J. (2019). Naniar: Data structures, summaries, and visualisations for missing data. *R Package*.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, *30*(4), 377-399. <https://doi.org/10.1002/sim.4067>

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. Springer.

Wickham, H. (2020). Package ‘plyr’. *Obtenido Httpscran Rproject Orgwebpackagesdplyrdplyr Pdf*.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28-36.

Supplement

Table XXX Phylogenetic signal in the missingness pattern (D statistic)

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **D** | **Prob random (D ≠ 1)** | **Prob Brownian (D ≠ 0)** |
| Body mass | 0.2903328 | 0 | 0.236 |
| Brain size | 0.932837 | 0.267 | 0 |
| Home range | 1.047108 | 0.685 | 0 |
| PC | 0.6307042 | 0 | 0.001 |
| Litter size | 0.7311753 | 0.01 | 0 |
| Shannon | 0.8283136 | 0.044 | 0 |
| Simpson | 0.8282997 | 0.045 | 0 |
| Abundance | 0.9189907 | 0.231 | 0 |
| Longevity | 0.932138 | 0.253 | 0 |
| Diet Category | 0.8314431 | 0.052 | 0 |

The D statistic for all variables except for body mass is significantly different from 0 (0 indicating strong phylogenetic signal) and for body mass, PC and litter size it is also significantly different from 1 (random), assuming an alpha level of 0.01. This suggests, that there is no strong phylogenetic signal in the missingness pattern, except for body mass, where D is significantly < 1 but also not significantly different from the Brownian expectation (D = 0).