

RESEARCH ARTICLE

Published
2021-11-29

Cite as

Emeric Figuet, Marion Ballenghien, Nicolas Lartillot and Nicolas Galtier (2021)
Reconstruction of body mass evolution in the Cetartiodactyla and mammals using phylogenomic data, Peer Community Journal, 1: e42.

Correspondence

nicolas.galtier@umontpellier.fr

Peer-review

Peer reviewed and recommended by
PCI Evolutionary Biology,
<https://doi.org/10.24072/pci.evolbiol.100042>



This article is licensed under the Creative Commons Attribution 4.0 License.

Reconstruction of body mass evolution in the Cetartiodactyla and mammals using phylogenomic data

Emeric Figuet¹, Marion Ballenghien², Nicolas Lartillot³, and Nicolas Galtier¹

Volume 1 (2021), article e42

<https://doi.org/10.24072/pcjournal.55>

Abstract

Reconstructing ancestral characters on a phylogeny is an arduous task because the observed states at the tips of the tree correspond to a single realization of the underlying evolutionary process. Recently, it was proposed that ancestral traits can be indirectly estimated with the help of molecular data, based on the fact that life history traits influence substitution rates. Here we challenge these new approaches in the Cetartiodactyla, a clade of large mammals which, according to paleontology, derive from small ancestors. Analysing transcriptome data in 41 species, of which 22 were newly sequenced, we provide a dated phylogeny of the Cetartiodactyla and report a significant effect of body mass on the overall substitution rate, the synonymous vs. non-synonymous substitution rate and the dynamics of GC-content. Our molecular comparative analysis points toward relatively small Cetartiodactyla ancestors, in agreement with the fossil record, even though our data set almost exclusively consists of large species. This analysis demonstrates the potential of phylogenomic methods for ancestral trait reconstruction and gives credit to recent suggestions that the ancestor to placental mammals was a relatively large and long-lived animal.

¹ISEM, Univ Montpellier, CNRS, IRD, Montpellier, France, ²Adaptation et Diversité en Milieu Marin, UMR7144, CNRS, Université Pierre et Marie Curie, Sorbonne Universités, Station Biologique de Roscoff, France, ³Laboratoire de Biométrie et Biologie Evolutive, UMR 5558, CNRS, Université Claude Bernard Lyon 1, Villeurbanne, France

Contents

Introduction	2
Methods	4
Results	6
Discussion	9
Data availability	11
Acknowledgments	11
Conflict of interest disclosure	11
References	11

Introduction

Accessing the ancestral states of a trait is of primary importance to study the processes having shaped its evolution (Cunningham et al., 1998; M Donoghue, 1989). Yet in most cases, ancient species and their characteristics are no longer observable and can only be indirectly estimated. In some cases, information on ancestral morphology or anatomy can be obtained through paleontological data. For instance, the analysis of a large fossil data set from marine environments was consistent with so-called Cope's or Depéret's rule, which postulates a tendency for lineages to evolve towards larger body mass (Bokma et al., 2016; Heim et al., 2015). Fossils, however, are scarce and their interpretation is often controversial, so that paleontology can but rarely provide us with a complete and detailed picture of the evolution of a trait in a clade.

To fill the gaps of direct evidence, evolutionary biologists have for long turned their attention onto the development of estimation methods using the comparative approach. This is typically achieved by fitting a model of trait evolution on a phylogeny to the observed data at the tips of the tree. For continuous character, such as body mass, the most basic model assumes an unidimensional Brownian motion with constant rate across branches (Felsenstein, 1985). Since then, more complex models have been proposed, such as models assuming an Ornstein–Uhlenbeck process, which constantly pulls the trait toward an optimum value (Beaulieu et al., 2012), or models seeking for a preferential trajectory in the random walk (Hunt, 2006). The major impediment of these approaches is that they rely on a unique realization of the modelled process. Consequently, they lack statistical power and are sensitive to model misspecification (Oakley and Cunningham, 2000).

The quickly accumulating molecular data might convey a new source of information about ancestral traits. Molecular evolution is influenced by a number of factors related to species biology and ecology, such as mutation rate, effective population size and mating systems (Lynch, 2007), which in turn are presumably influenced by species life-history traits. Genomic sequences are expected to keep track of the various selective or neutral pressures that species have undergone through time. The idea, therefore, would be to indirectly estimate ancestral traits based on ancestral molecular reconstructions. The strength of this approach comes from the fact that molecular characters are numerous, which considerably increases the power and accuracy of phylogenetic inference methods.

In mammals, various aspects of the molecular evolutionary process have been linked to species life-history traits (LHT), including nucleotide and amino acid substitution rates (Bromham et al., 1996; Martin and Palumbi, 1993; Nabholz et al., 2008) and genomic GC content (Lartillot, 2013; Romiguier, Ranwez, et al., 2013). The nonsynonymous (= amino acid changing, d_N) to synonymous (d_S) substitution rate ratio, d_N/d_S , is of specific interest, as it has been found to be positively correlated with species body size, longevity and generation time in mammals (Nikolaev et al., 2007; Popadin et al., 2007; Romiguier, Figuet, et al., 2012). These relationships were interpreted as reflecting a decreased effective population size (N_e) in large mammals (Damuth, 1981, 1987; White et al., 2007), compared to small-sized ones. According to the nearly neutral theory of molecular evolution (Ohta, 1987), purifying selection against slightly deleterious mutations is

less efficient in small populations, in which the random effects of genetic drift dominate. Assuming that a majority of nonsynonymous changes are deleterious, an increased proportion of nonsynonymous substitutions is therefore expected in (large-bodied) low- N_e species (Figuet, Nabolz, et al., 2016). Besides d_N/d_S , an increased effective population size in small-sized species is also expected to impact the efficiency of biased gene conversion (gBGC), a recombination-driven process that favors the spread of GC relatively to AT alleles, mimicking a selective advantage for G and C (Duret and Galtier, 2009; Galtier, Piganeau, et al., 2001; Webster et al., 2005). Consistent with theoretical predictions, it was shown that small-sized, large- N_e mammals tend to exhibit a high GC-content compared to large-sized species, particularly at third codon positions (Romiguier, Ranwez, et al., 2010).

Recently, a few studies took advantage of these correlations between species life-history traits and molecular evolutionary processes to estimate ancestral character states in mammals using phylogenetic approaches. Using hundreds of nuclear genes, Romiguier, Ranwez, et al. (2013) reported that the estimated d_N/d_S ratio and GC content dynamics during early divergence in placental mammals were comparable to those of current long-lived mammals, and very different from those of current short-lived mammals. They concluded that placental ancestors were probably long-lived (>20 years) animals, in sharp contrast with the classical view of tiny early mammals diversifying after the massive extinction of large reptiles (e.g. Feldhamer et al., 2007). Lartillot and Delsuc (2012) analysed a smaller data set using a more sophisticated method in which the correlation between molecular rates and LHT was explicitly modelled (Lartillot and Poujol, 2011). Their results were largely consistent with those of Romiguier, Ranwez, et al. (2013), rejecting the hypothesis of mouse-like placental ancestors, even though point estimates for ancestral traits differed between the two studies.

These estimates of ancestral LHT, which concern early placental ancestors, can hardly be compared to paleontological data owing to the uncertainty regarding placental origins (Archibald et al., 2001; Asher et al., 2005). Molecular dating, for instance, places the placental root in the Cretaceous (dos Reis, P Donoghue, et al., 2014; dos Reis, Inoue, et al., 2012), but no fossils from this epoch (Goswami et al., 2011; Wible et al., 2007, 2009) are unambiguously assigned to Placentalia. Perhaps for this reason, the results of Lartillot and Delsuc (2012) and Romiguier, Ranwez, et al. (2013), which rely on novel, untested methods, so far have had a limited impact on the paleontological literature. To assess the reliability of molecular-aided reconstructions of past life-history traits, one idea would be to focus on clades for which the fossil record provides a well-supported and consensual evolutionary scenario.

The clade of cetartiodactyls appears as an interesting candidate in this respect. Cetartiodactyls are a group of 332 recognized species, classified in 22 families (Hassanin et al., 2012), that emerged around the Cretaceous-Paleogene crisis according to molecular dating (Meredith et al., 2011; Psouni et al., 2012). It comprises, among others, clades such as suines (pig and peccaries), tylopodes (camels), true ruminants (giraffe, cow, antelopes) as well as hippos and cetaceans. Extant species of cetartiodactyls exhibit a wide spectrum of body mass, ranging from a few kilograms (mouse-deer, dwarf antelopes, pygmy hog) to tons (hippos, giraffes) and even hundreds of tons (whales). Furthermore, with the exception of cetaceans, they share a number of ecological similarities, such as a terrestrial life style and a herbivorous diet often associated with rumination, making their LHT perhaps more comparable than in Placentalia-scale analyses. The fossil record of Cetartiodactyla is quite furnished with a mean of six extinct genera per extant genus (McKenna and Bell 1997). Despite the majority of extant representatives of this clade exceeding a hundred of kilograms, the early fossil record consists in small-sized animals, such as the emblematic *Diacodexis*, the oldest known cetartiodactyl, which only weighted a few kilograms (Rose 1982). Besides *Diacodexis*, the Eocene record of cetartiodactyls, with families such as Heloyidae, Raoellidae, which are close relative to whales, or Choeropotamidae of which the genus *Amphirhagatherium* only includes relatively small-sized species (Hooker and Thomas, 2001; Orliac and Ducrocq, 2011). It is therefore widely admitted that body mass in cetartiodactyls has been convergently increasing in a number of independent lineages during the evolution of this group.

This peculiar history of body mass evolution creates an engaging challenge for molecular data. Corroborating the existence of small-sized ancestral cetartiodactyls based on a sample of

large-sized extant species would be quite reassuring regarding the reliability of these approaches. An analysis of complete mitochondrial genomes in 201 species of cetartiodactyls suggested that DNA-aided reconstruction can estimate ancestral LHTs more accurately than classical methods (Figuet, Romiguier, et al., 2014). Mitochondrial DNA, however, is limited in size. In addition, it experiences a high rate of substitution and is thus potentially subject to mutational saturation, making ancestral reconstructions more challenging. Nuclear DNA, in contrast, shows a relatively low rate of evolution but is subject to recombination and GC-biased gene conversion. Although making the analysis potentially more complicated, these additional evolutionary forces also represent genuine sources of signal about ancestral traits through among-lineage variation in GC content.

In this study, we aim to evaluate the potential and robustness of genomic data for ancestral life-history trait reconstruction in the Cetartiodactyla. To this end, we generated a dataset of thousands of genes in 41 species, of which 22 were newly sequenced, and conducted analyses inspired by Romiguier, Ranwez, et al. (2013) and Lartillot and Delsuc (2012). Firstly, we approached ancestral body mass, longevity and age of sexual maturity based on the observed correlation between life-history and molecular traits across extant species (Figuet, Romiguier, et al., 2014; Romiguier, Ranwez, et al., 2013). Secondly, we used the integrated approach developed by Lartillot and Poujol (2011), in which the correlated evolution of life-history traits and molecular parameters across the phylogenetic tree is modelled as a multivariate Brownian motion, so as to obtain quantitative estimates of ancestral body masses and credibility intervals. The strength of this approach is to account for phylogenetic inertia and various sources of uncertainty by combining all variables in a single Bayesian analysis.

Methods

Sample collection, RNA extraction and sequencing.

Twenty one Cetartiodactyla species were sampled and analyzed in this study, namely the Chacoan peccary, common warthog, hippo, pygmy hippo, okapi, giraffe, reindeer, tufted deer, fallow deer, red deer, sika deer, Siberian musk deer, nilgai, nyala, common eland, Kirk's dik-dik, dama gazelle, Nile lechwe, bontebok, addax, and white oryx. One species of Perissodactyla, the white rhinoceros, was also newly sequenced here. All reads are available from the NCBI-SRA database, BioProject PRJNA388863. Coding sequences from twelve additional Cetartiodactyla species and seven outgroups (two Eulypotyphla, two Chiroptera, two Carnivora, and one Perissodactyla) were retrieved from publicly available genomic or RNA-Seq data bases (see electronic supplementary material, appendix A, table S1).

Orthologous genes and alignments.

De novo transcriptome assembly was performed with a combination of the Abyss and Cap3 programs following strategy B in Cahais et al. (2012). The longest ORF of each predicted cDNA was kept, contigs not including any ORF longer than 200bp being discarded. Orthology was predicted using the OrthoMCL software (Li et al., 2003, see electronic supplementary material, appendix A for details). Predicted orthologous sequences were aligned using the MACSE program (Ranwez et al., 2011), which relies on amino acid translated sequences while allowing for frameshifts at nucleotide level. Obviously mis-aligned regions were semi-automatically cleaned out using home-made scripts. Briefly, this was done by scanning each sequence for small regions flanked by undetermined base pairs, and removing them if they shared less than 70% similarity with any other sequence. The 3,022 resulting alignments were restricted to positions having less than 30% missing data, returning a total length of 3.4Mb of coding sequences in 41 species, with 17% missing data. Because the Bayesian integrated method developed by Lartillot and Poujol (2011) can hardly handle a data set of that size, we created a reduced data set in which we only kept genes at which less than two Cetartiodactyla species missing. The criterion was satisfied by 1,223 alignments, leading to a concatenation of 1.5Mb with 12% missing data.

Life-history traits.

Life-history traits values for body mass, maximum longevity and age of sexual maturity were retrieved both from the AnAge database build 13 (de Magalhães and Costa, 2009, male and female values for sexual maturity) and the PANTHERIA database (Jones et al., 2009). When several estimates per species were available we calculated the mean (see electronic supplementary material, table S1). All values were log10-transformed prior to perform any analyses. Maximum longevity and age of sexual maturity in *Pantholops hodgsonii* as well as age of sexual maturity in *Pteropus vampyrus*, which were missing from AnAge and PANTHERIA, were obtained from the Animal Diversity Web resource (<http://animaldiversity.org/>).

Phylogenetic reconstruction, substitution mapping, branch-specific d_N/d_S .

A phylogenetic tree was estimated using RaxML version 8 (Stamatakis, 2006) with a GTR+GAMMA model. We trusted the reconstructed phylogeny as far as within-Cetartiodactyl branching orders were concerned. Regarding the inter-ordinal relationships, which are controversial, we used two alternative topologies taken from the literature, i.e., (Eulipotyphla, ((Chiroptera, (Carnivora, Perissodactyla)), Cetartiodactyla)) (Nishihara et al., 2006) and (Eulipotyphla, ((Carnivora, Perissodactyla), (Chiroptera, Cetartiodactyla))) (Nery et al., 2012). We also considered two alternative branching orders for the Nile lechwe Kobus (see electronic supplementary material, appendix B). The substitution mapping procedure (JY Dutheil et al., 2012; Romiguier, Figuet, et al., 2012) was used to estimate the mean d_N/d_S in each branch of the uncalibrated tree across the 3,022 alignments, following Figuet, Nabholz, et al. (2016) - see electronic supplementary material, appendix A for details about the methodology. This approach is similar but faster than, e.g., full likelihood branch models such as implemented in codeml (Yang, 2007).

Clustering procedure.

To assess the potential noise introduced by recent variation in effective population size or life-history traits in terminal branches that could impede their relation with long-term d_N/d_S estimated by substitution mapping, we also combined the signal from closely related terminal branches into a single data point. To achieve this, we defined monophyletic clusters of species and calculated one d_N/d_S ratio per cluster. Clusters were defined at the level of tribes in Ruminantia, at the family level in Suina and Tylopoda, and at the parvorder level in Cetacea. The d_N/d_S value of a cluster was obtained by summing the nonsynonymous and synonymous substitution counts across its terminal branches and calculating the ratio. Life-history trait values of a cluster were defined as the mean of the log10-transformed values of its representatives, either from our sampling or from all available species at this taxonomic level in the AnAge database build 13. This clustering procedure resulted in the creation of 19 clusters in Cetartiodactyla, of which eleven contained two or three species and eight contained a single species (see electronic supplementary material, table S2).

Bayesian analysis of the correlated evolution between substitution rates and body mass.

Alternatively, we used the integrated COEVOL method (version 1.5, Lartillot and Poujol, 2011) to jointly assess the correlations between life-history traits, d_S and d_N/d_S and estimate ancestral states, based on the reduced concatenated data set. The method models the correlated evolution of the analyzed variables by assuming a multivariate Brownian diffusion process parameterized by a covariance matrix, along a time tree constrained by fossil calibrations. All parameters are estimated in a Bayesian framework using Monte Carlo Markov Chains. The strength of the coupling between a molecular variable and a LHT was measured by the correlation coefficient (posterior mean estimate) between the variables, and the statistical support was evaluated based on the posterior probability (pp) of a positive (pp close to 1) or negative covariation (pp close to 0). d_S and d_N/d_S were analyzed as dependent variables, either jointly or separately. Ten calibration points were used (see electronic supplementary material, table S3, appendix A for details).

GC3 analysis.

Ancestral GC-content at third codon positions (GC3) was estimated for all nodes of the tree separately for each of the 3,022 genes using the approach of (Galtier and Gouy, 1998), implemented in the bpp-ML suite (Dutheil and Boussau, 2008). This method uses a nonhomogeneous and nonstationary Markov model of nucleotide evolution to estimate branch-specific equilibrium GC-content in a maximum-likelihood framework on the uncalibrated tree. We then employed the procedure used in (Romiguier, Ranwez, et al., 2013) to correlate GC3 evolution to life-history traits. This means calculating, for each of eleven selected pairs of extant species (see electronic supplementary material, appendix A for details about the methodology, an index of GC3 conservation, defined as $\gamma = -t/\log(\tau)$, where t is the divergence time, and τ is Kendall rank correlation coefficient of GC3 across genes between the two considered species. This index was also calculated for pairs of ancestral taxa and used for ancestral body mass estimation based on the correlation with body mass in extant pairs.

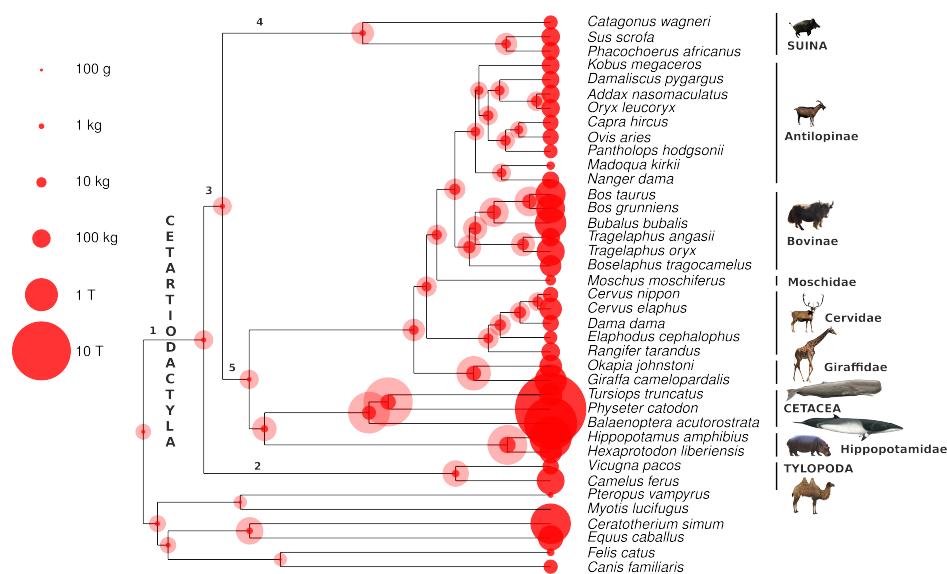


Figure 1 – Reconstruction of body mass in the Cetartiodactyla using the integrated approach. All three LHTs were included in the analysis as well as both d_S and d_N/d_S as dependent variables. Dark- and light-shaded disks indicate 95% credibility intervals for ancestral body masses (in grams). A uniform prior distribution was used to calibrate divergence times; see electronic supplementary material, figure S1. Numbers are labels for internal branches.

Results

Phylogeny and molecular dating.

We took benefit of our extensive genomic data set to produce a phylogeny of the Cetartiodactyla clade. Analysed with RaxML, the reduced and complete data set supported topologies that were identical (figure 1) and in very good agreement with previously published phylogenies (Agnarsson and May-Collado, 2008; Hassanin et al., 2012; Price et al., 2005). Of note, Tylopoda branch out as the sister group to all other Cetartiodactyla in this tree, contradicting the species-rich mitochondrial-based analysis of Hassanin et al. (2012), in which Suina was the earliest diverging group. Bootstrap support was 100% for all nodes, with the exception of the position of the Nile lechwe, for which 88% and 78% bootstrap support were obtained from the complete and reduced data set, respectively. The estimated divergence times estimated by coevol on the

reduced data set appeared to be robust to the choice of the prior on divergence times (uniform versus birth death) (see electronic supplementary material, figure S1). In particular, within the Cetartiodactyla clade, estimated dates differ by no more than 2 Myr between the analyses under the two alternative priors.

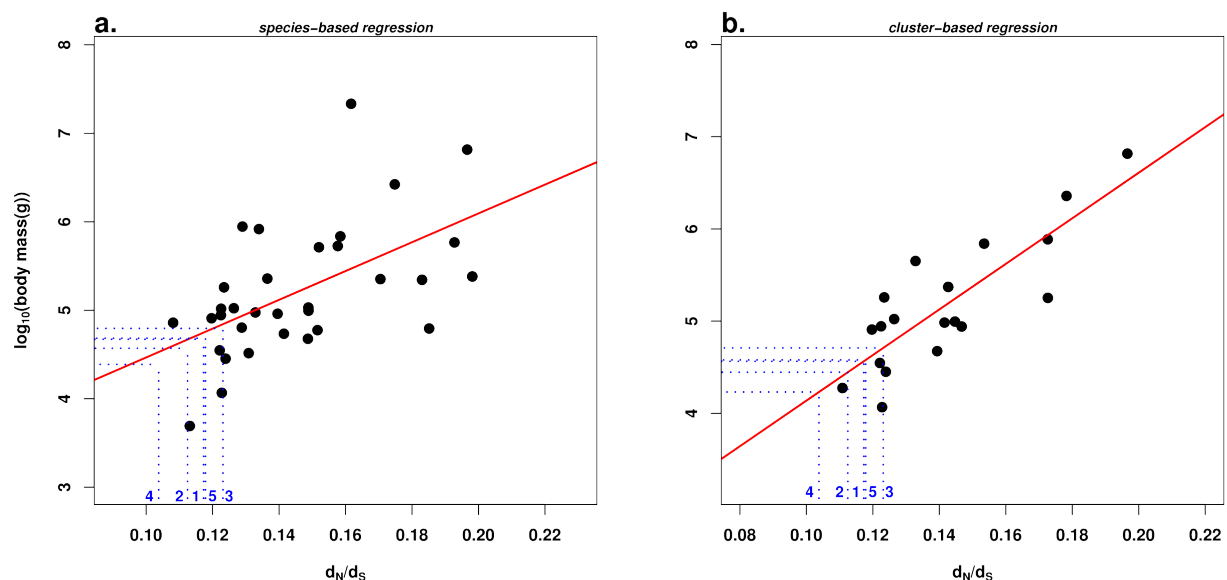


Figure 2 – Relationship between $\log_{10}(\text{body mass})$ and d_N/d_S in Cetartiodactyla. a. Each dot represents a species. d_N/d_S values were obtained using substitution mapping. The regression line of $\log_{10}(\text{body mass})$ on terminal branch d_N/d_S is shown in red. The blue dashed lines show ancestral body mass predictions for five basal branches of Cetartiodactyla. Branch labels: see figure 1. b. Each dot represents a cluster of monophyletic species. Body mass value of a cluster corresponds to the mean of the \log_{10} -transformed values of its representatives.

d_N/d_S vs. life-history traits correlations in the Cetartiodactyla.

Using a data set of 3,022 genes and 3.4 millions of positions, we estimated d_N/d_S in terminal branches of the tree and correlated it to three life-history traits. Significant correlations were detected between d_N/d_S and body mass ($r = 0.56$, $P < 10^{-3}$, figure 2a), longevity ($r = 0.59$, $P < 10^{-3}$) and sexual maturity ($r = 0.56$, $P < 10^{-3}$), thus confirming in the Cetartiodactyla the influence of these traits on coding sequence evolutionary processes. However, no significant correlation was found between K_r/K_c (ie, ratio of radical over conservative amino-acid substitution rate) and any of the three life-history traits. Because current life-history traits might not accurately predict the long-term effective population size, we constructed clusters of monophyletic species by combining the substitution information across all its terminal branches. This procedure led to stronger correlations for body mass ($r = 0.83$, $P < 10^{-4}$, figure 2b), longevity ($r = 0.78$, $P < 10^{-4}$), and age of sexual maturity ($r = 0.72$, $P < 10^{-3}$) when LHTs were averaged across our sampled species or across all species available in the AnAge database (body mass: $r = 0.82$, $P < 10^{-4}$; longevity: $r = 0.81$, $P < 10^{-4}$; age of sexual maturity: $r = 0.85$, $P < 10^{-5}$).

We estimated ancestral d_N/d_S values in five basal branches of the Cetartiodactyla tree (figure 1) via substitution mapping. These fell in the lowest range of extant species d_N/d_S , and are therefore each suggestive of relatively small body-mass – 10-100kg – in the corresponding ancestor (blue dotted lines in figure 2). The strength and slope of the d_N/d_S - LHTs correlations and the estimated ancestral d_N/d_S values were not sensitive to the use of alternative topologies in which the branching order among outgroups or the position of the Nile lechwe differed (see

electronic supplementary material, Appendix B). Using the reduced data set of 1,223 genes with less missing data also returned similar results (see electronic supplementary material, table S5).

Modelling the correlated evolution between substitution parameters and LHTs.

In order to properly account for the covariation between LHTs and substitution rates, we used the *coevol* program on the reduced data set, which models the evolution of molecular and life-history variables as a multivariate Brownian motion. The analysis under *coevol* indicates a negative correlation of the synonymous substitution rate d_S , and a positive correlation of d_N/d_S , with the three LHTs (table 1, correlation coefficients vary by no more than 0.03 between both chains). Concerning d_S , the correlation is essentially driven by sexual maturity ($r = -0.61$), which suggests a generation-time effect on the mutation rate. In contrast, for d_N/d_S , the correlation is primarily with body-mass ($r = 0.52$). Overall, the correlation estimated by this integrative modeling strategy are consistent with the mapping-based methods (see above).

	body mass	longevity	sexual maturity
d_S	-0.56**	-0.53**	-0.61**
d_N/d_S	0.52**	0.44**	0.23

Table 1 – Relationships between life-history traits and molecular substitution rates in the Cetartiodactyla using the integrated approach. Values are given for chain 1. **: $pp > 0.975$ or < 0.025 ; *: $pp > 0.95$ or < 0.05

The body mass inferred by *coevol* in the early lineages of the group suggest a relatively small ancestor (95% credible interval between 500g to 115kg). This should be contrasted with what would be obtained without any information coming from molecular sequences – thus, based on a univariate log-Brownian model for body mass (credible interval between 10kg and 770kg). Analyses using either d_S or d_N/d_S as the single dependent variable against all three LHTs suggest that the correlation between LHTs and d_S is the primary driver of the inference of smaller ancestral body mass, although d_N/d_S also appears to contribute a weak signal in favor of a small-bodied ancestor (table 2).

	$d_S, d_N/d_S$ - traits	d_S - traits	d_N/d_S - traits	traits
chain1	530g - 89kg	570g - 97kg	7.6kg - 498kg	12 kg - 760 kg
chain2	640g - 115kg	590g - 111kg	8.5kg - 607kg	10 kg - 770 kg

Table 2 – Estimated body mass (95% credibility interval) at the root of the Cetartiodactyla. All *coevol* and the *ancov* analysis, that does not use molecular data, are shown.

GC3-based ancestral reconstruction.

We analyzed the variation of GC3, a variable known to be influenced by LHT in mammals, to approach ancestral life-history traits. We retrieved strong correlations in the reduced data set between the across-genes mean GC3 and LHT in Cetartiodactyls despite the low range of observed mean GC3 (49.5%-51.5%): $r = -0.65$, $P < 10^{-4}$ for body mass, $r = -0.69$, $P < 10^{-5}$ for maximum longevity and $r = -0.67$, $P < 10^{-4}$ for age of sexual maturity (see electronic supplementary material, figure S3), in agreement with a more pronounced action of gBGC in small-sized cetartiodactyls presumably having larger effective population sizes (Lartillot, 2013; Romiguier, Ranwez, et al., 2010).

GC3 cannot be directly used to estimate LHTs because of phylogenetic inertia. Following Romiguier, Ranwez, et al. (2013), we identified eleven pairs of closely related species and calculated for each the correlation coefficient of GC3 across genes, and its divergence time-corrected version, γ . We observed a strong correlation between body mass and γ across the eleven pairs of

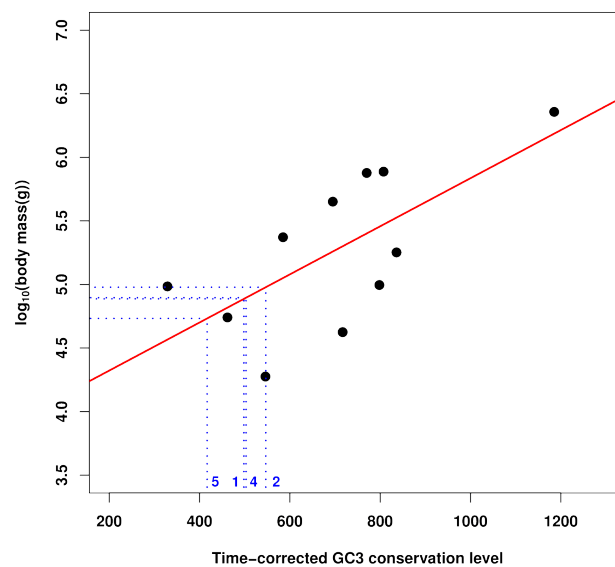


Figure 3 – Relationship between $\log_{10}(\text{body mass})$ and time-corrected GC3 conservation level. Each dot represents a pair of cetartiodactyl species. The x axis corresponds to $\gamma = -t/\log(\tau)$, where t is the divergence time taken from TimeTree and τ the Kendall rank correlation coefficient. The regression line of $\log_{10}(\text{body mass})$ on time-corrected GC3 conservation level is shown in red. The blue dashed lines show the ancestral body mass predictions for four of the five most basal branches of Cetartiodactyla (branch 3 is not displayed because of its particularly low γ value: 98). Branch labels: see figure 1.

extant species ($\tau = 0.68$, $P < 0.02$, figure 3). Then we calculated γ in five pairs of ancestral Cetartiodactyl species (figure 1), ancestral GC3 being estimated based on a non-homogeneous, non-stationary model of sequence evolution. Similarly to the d_N/d_S analysis, we observed that pairs of basal Cetartiodactyl species all exhibited a low level of GC3 conservation, again suggesting a relatively small body size for these ancestors. Similar results were obtained when divergence times were taken from our coevol analysis rather than TimeTree (see electronic supplementary material, figure S4), the correlation between body mass and γ being slightly stronger in this analysis ($\tau = 0.79$, $P < 0.004$). Two of the five ancestral branches were suggestive of medium-size ancestors, not small-size ancestors. These are the Tylopoda and Suina ancestral branches, each corresponding to a period of time longer than 25My.

Discussion

Our goal in this study was to test the ability of molecular-aided method for ancestral trait reconstruction in a difficult case study – the Cetartiodactyla. We show that between lineages patterns of variation in synonymous and non-synonymous substitution rates and in GC-content all point towards relatively small ancestors for this group. Our analyses indicate that the per year synonymous substitution rate, the d_N/d_S ratio and GC-content dynamics all carry some signal regarding ancestral traits in cetartiodactyls. In terms of the underlying molecular evolutionary mechanisms, this signal is possibly mediated by two distinct effects, on generation-time (for the variation in absolute synonymous substitution rate) and on effective population size (for the variation in d_N/d_S and GC content), both of which are known to correlate with body-mass in vertebrates (Bromham, 2002; Nabholz et al., 2008; Woolfit and Bromham, 2005). Thus, in at least two respects, molecular evolution in ancestral cetartiodactyls was similar to molecular evolution in currently small-sized taxa and different from molecular evolution in currently large-sized taxa.

The inference of a small cetartiodactyl ancestor is in agreement with the fossil record – and this, even though the vast majority of current species are large in size. Molecular-aided point estimates of ancestral body mass are about one order of magnitude below those of DNA-unaware methods (between 0.5 and 115 kg with molecular information, versus 10kg to 770kg without this information). This illustrates a point already observed in another phylogenetic context (evolution of the optimal growth temperature in Archaea (Groussin and Gouy, 2011; Horvilleur and Lartillot, 2014), namely, that molecular information makes it possible to infer global trends in trait evolution along a tree, which classical comparative approaches are not able to identify, at least without fossil information. Our analysis therefore illustrates the potential of molecular data to solve difficult problems regarding ancestral state reconstruction, in situations where classical comparative methods have inherent limitations.

The most ancient fossils of Cetartiodactyla are found in the Eocene, i.e., at least ten million years after the origin of the group according to molecular dating (Meredith et al., 2011; Murphy et al., 2007, this study). Fossils such as *Diacodexis* are difficult to pinpoint precisely onto the phylogeny of extant taxa, and can only be used as a raw guide for comparison with our estimated body masses in basal branches. Still, our reconstruction suggests that cetartiodactyl ancestors have remained small-sized for an extended period, from the Paleocene to the end of the Eocene, during which paleontology indeed only witnesses small-sized species. We infer independent evolution toward larger sizes in the major Cetartiodactyla lineages (figure 1), namely Tylopoda, Suina, Ruminantia and Cetacea. This is again consistent with the fossil record of these groups, particularly Tylopoda, in which families such as Oromerycidae consisted of small-sized species, and the well documented Cetacea (Montgomery et al., 2013).

Our approach to rate analysis relies on a single, supposedly known species tree and concatenated sequences. It was recently suggested that such an approach could be biased in case of incomplete lineage sorting, when gene trees differ from the species tree (Hahn and Nakhleh, 2016; Mendes et al., 2016). Incomplete lineage sorting, however, can only be effective if internal branch length is of the order of magnitude of, or lower than, within species polymorphism (Scornavacca and Galtier, 2017). Published population genomic data in Cetartiodactyla indicate that coding sequence polymorphism is of the order of 0.001-0.1% (Choi et al., 2015; Kardos et al., 2015; H Wu et al., 2014; Yim et al., 2014). Only two internal branches in our reference RaxML tree are that short (see electronic supplementary material, figure S2, appendix B), which represent a tiny fraction of the whole tree length. We conclude, following Scornavacca and Galtier (2017), that incomplete lineage sorting is probably a negligible issue in this analysis. Please also note that our results are robust to changes in the reference phylogeny (Table S5).

Interestingly, the Bayesian analysis conducted here is mainly driven by d_S , with d_N/d_S contributing only a minor amount of information, whereas substitution mapping suggests quite a strong link between d_N/d_S and LHTs. The reason for such a discrepancy between the two approaches is yet unclear. Please note that a time tree is used in the Bayesian analysis, whereas d_N/d_S were estimated using a non-ultrametric tree with substitution mapping. This could explain in part the discrepancies between the two analyses. In addition, clustering species by clade and analysing cluster averages yielded a substantial increase in correlation coefficient, compared to the basic analysis (figure 2). This would not be expected if the logarithm of rates and traits were evolving according to a multivariate Brownian diffusion process, as assumed by the co-evol model, and suggests that not all the information contained in the data was captured by our analysis. Modelling the correlated evolution of molecular and life-history traits is a complex, computationally demanding task, but a necessary one if one wants to properly account for phylogenetic inertia (Lartillot and Poujol, 2011). This study confirms the potential of an integrated, Bayesian approach while offering suggestions for improvement of existing models. Besides these methodological aspects, it should be noted that the correlation between rates and traits is probably imperfect in nature – an intrinsic limitation to our approach, which probably determines in the first place the width of credibility intervals.

The success of DNA-aided methods in reconstructing small cetartiodactyl ancestors, such as independently indicated by fossil evidence, gives credit to other analyses for which fossil data

are lacking — in particular, the suggestion by Lartillot and Delsuc (2012) and Romiguier, Ranwez, et al. (2013) of a relatively large and long-lived placental ancestor. Given the controversy on the age of this ancestor and the uncertain interpretation of mammalian fossils from the Cretaceous (Ronquist et al., 2016; Springer et al., 2017), relying on the correlations between life history traits and substitution processes is probably the optimal way to currently approach ancestral states in mammals. Time-independent molecular variables, such as the d_N/d_S ratio, are particularly appropriate when dates are controversial, as in the case of early placental evolution. Another promising approach involves analysing the variance, not just the mean, of substitution rates across genes (J Wu et al., 2017).

Our approach, however, is only applicable when strong correlations exist between rates and traits, which is the case in some but not in every taxonomic group (Bromham, 2002; Figuet, Nabholz, et al., 2016; Lourenço et al., 2013; Rolland et al., 2016). Such correlations exist because life-history traits influence substitution rates through the mutation rate, effective population size, recombination rate, or other evolutionary forces. Further understanding the link between rates and traits, e.g. via population genomic studies, should enhance our ability to model these relationships and improve ancestral reconstructions.

Data availability

All reads are available from the NCBI-SRA database, BioProject PRJNA388863. Supplementary material is available from <https://www.biorxiv.org/content/10.1101/139147v3>.

Acknowledgments

The authors are grateful to the Lunaret zoo, Servion zoo, Pescheray zoo, Gramat zoo, Cédric Libert, Elodie Trunet, Eugène Chabloz, Jean-Marc Charpentier, and Pierre Delrieu for their great help with sampling. They are particularly grateful to the San Diego Frozen Zoo and Cynthia Steiner for providing RNA data from species that were particularly difficult to sample. They also thank Laurent Duret and Jonathan Romiguier for helpful discussion and the Montpellier Bioinformatics & Biodiversity platform for support regarding computational aspects. This work was supported by European Research Council advanced grant 232971 (PopPhyl) and Agence Nationale de la Recherche grant ANR-10-BINF-01-01 (Ancestrôme).

Conflict of interest disclosure

The authors declare that they have no financial conflict of interest with the content of this article. Nicolas Galtier is a recommender for PCI Evolutionary Biology and PCI Genomics. Nicolas Lartillot is a recommender for PCI Evolutionary Biology.

References

- Agnarsson I, May-Collado L (2008). *The phylogeny of Cetartiodactyla: the importance of dense taxon sampling, missing data, and the remarkable promise of cytochrome b to provide reliable species-level phylogenies*. *Mol Phylogenet Evol* **48**, 964. <https://doi.org/10.1016/j.ympev.2008.05.046>.
- Archibald J et al. (2001). *Late Cretaceous relatives of rabbits, rodents, and other extant eutherian mammals*. *Nature* **414**, 62. <https://doi.org/10.1038/35102048>.
- Asher R et al. (2005). *Stem Lagomorpha and the Antiquity of Glires*. *Science* **307**, 1091. <https://doi.org/10.1126/science.1107808>.
- Beaulieu J et al. (2012). *Modeling stabilizing selection: Expanding the Ornstein-Uhlenbeck model of adaptive evolution*. *Evolution* **66**, 2369. <https://doi.org/10.1111/j.1558-5646.2012.01619.x>.
- Bokma F et al. (2016). *Testing for Depéret's Rule (Body Size Increase) in Mammals using Combined Extinct and Extant Data*. *Syst Biol* **65**, 98. <https://doi.org/10.1093/sysbio/syv075>.
- Bromham L (2002). *Molecular clocks in reptiles: life history influences rate of molecular evolution*. *Mol Biol Evol* **19**, 302. <https://doi.org/10.1093/oxfordjournals.molbev.a004083>.

- Bromham L et al. (1996). Determinants of rate variation in mammalian DNA sequence evolution. *J Mol Evol* **43**, 610. <https://doi.org/10.1007/BF02202109>.
- Cahais V et al. (2012). Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Mol Ecol Resour* **12**, 834. <https://doi.org/10.1111/j.1755-0998.2012.03148.x>.
- Choi J et al. (2015). Whole-genome resequencing analyses of five pig breeds, including Korean wild and native, and three European origin breeds. *DNA Res* **22**, 259. <https://doi.org/10.1093/dnares/dsv011>.
- Cunningham C et al. (1998). Reconstructing ancestral character states: a critical reappraisal. *Trends Ecol Evol* **13**, 361. [https://doi.org/10.1016/s0169-5347\(98\)01382-2](https://doi.org/10.1016/s0169-5347(98)01382-2).
- Damuth J (1981). Population density and body size in mammals. *Nature* **290**, 699. <https://doi.org/10.1038/290699a0>.
- Damuth J (1987). Interspecific allometry of population density in mammals and other animals: the independence of body mass and population energy-use. *Biol J Linn Soc* **31**, 193. <https://doi.org/10.1111/j.1095-8312.1987.tb01990.x>.
- de Magalhães J, Costa J (2009). A database of vertebrate longevity records and their relation to other life-history traits. *J Evol Biol* **22**, 1770. <https://doi.org/10.1111/j.1420-9101.2009.01783.x>.
- Donoghue M (1989). Phylogenies and the analysis of evolutionary sequences, with examples from seed plants. *Evolution* **43**, 1137. <https://doi.org/10.1111/j.1558-5646.1989.tb02565.x>.
- dos Reis M, Donoghue P, et al. (2014). Neither phylogenomic nor palaeontological data support a Palaeogene origin of placental mammals. *Biol Lett* **10**, 20131003. <https://doi.org/10.1098/rsbl.2013.1003>.
- dos Reis M, Inoue J, et al. (2012). Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc R Soc B Biol Sci* **279**, 3491. <https://doi.org/10.1098/rspb.2012.0683>.
- Duret L, Galtier N (2009). Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet* **10**, 285. <https://doi.org/10.1146/annurev-genom-082908-150001>.
- Dutheil J, Boussau B (2008). Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs. *BMC Evol Biol* **8**, 255. <https://doi.org/10.1186/1471-2148-8-255>.
- Dutheil J et al. (2012). Efficient selection of branch-specific models of sequence evolution. *Mol Biol Evol* **29**, 1861. <https://doi.org/10.1093/molbev/mss059>.
- Feldhamer G et al. (2007). *Mammalogy: Adaptation, Diversity, Ecology*. Baltimore: Johns Hopkins University Press.
- Felsenstein J (1985). *Phylogenies and the Comparative Method*. *Am Nat* **125**, 1. <https://doi.org/10.1086/284325>.
- Figuet E, Nabholz B, et al. (2016). Life History Traits, Protein Evolution, and the Nearly Neutral Theory in Amniotes. *Mol Biol Evol* **33**, 1517–1527. <https://doi.org/10.1093/molbev/msw033>.
- Figuet E, Romiguier J, et al. (2014). Mitochondrial DNA as a tool for reconstructing past life-history traits in mammals. *J Evol Biol* **27**, 899. <https://doi.org/10.1111/jeb.12361>.
- Galtier N, Gouy M (1998). Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Mol Biol Evol* **15**, 871. <https://doi.org/10.1093/oxfordjournals.molbev.a025991>.
- Galtier N, Piganeau G, et al. (2001). GC-Content Evolution in Mammalian Genomes: The Biased Gene Conversion Hypothesis. *Genetics* **159**, 907. <https://doi.org/10.1093/genetics/159.2.907>.
- Goswami A et al. (2011). A radiation of arboreal basal eutherian mammals beginning in the Late Cretaceous of India. *Proc Natl Acad Sci USA* **108**, 16333. <https://doi.org/10.1073/pnas.1108723108>.
- Groussin M, Gouy M (2011). Adaptation to environmental temperature is a major determinant of molecular evolutionary rates in archaea. *Mol Biol Evol* **28**, 2661. <https://doi.org/10.1093/molbev/msr098>.

- Hahn M, Nakhleh L (2016). Irrational exuberance for resolved species trees. *Evolution* **70**, 7. <https://doi.org/10.1111/evo.12832>.
- Hassanin A et al. (2012). Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *C R Biol* **335**, 32. <https://doi.org/10.1016/j.crvi.2011.11.002>.
- Heim N et al. (2015). Animal Evolution. Cope's rule in the evolution of marine animals. *Science* **347**, 867–870. <https://doi.org/10.1126/science.1260065>.
- Hooker J, Thomas K (2001). A new species of Amphirhagatherium (Choeropotamidae, Artiodactyla, Mammalia) from the late eocen headon hill formation of southern england and phylogeny of endemic european "Anthracotherioids". *Palaeontology* **44**, 827. <https://doi.org/10.1111/1475-4983.00203>.
- Horvillour B, Lartillot N (2014). Monte Carlo algorithms for Brownian phylogenetic models. *Bioinformatics* **30**, 3020. <https://doi.org/10.1093/bioinformatics/btu485>.
- Hunt G (2006). Fitting and Comparing Models of Phyletic Evolution: Random Walks and beyond. *Paleobiology* **32**, 578. <https://doi.org/10.1666/05070.1>.
- Jones K et al. (2009). PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology* **90**, 2648. <https://doi.org/10.1890/08-1494.1>.
- Kardos M et al. (2015). Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. *Mol Ecol* **24**, 5616. <https://doi.org/10.1111/mec.13415>.
- Lartillot N (2013). Phylogenetic Patterns of GC-Biased Gene Conversion in Placental Mammals and the Evolutionary Dynamics of Recombination Landscapes. *Mol Biol Evol* **30**, 489. <https://doi.org/10.1093/molbev/mss239>.
- Lartillot N, Delsuc F (2012). Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic model. *Evolution* **66**, 1773. <https://doi.org/10.1111/j.1558-5646.2011.01558.x>.
- Lartillot N, Poujol R (2011). A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Mol Biol Evol* **28**, 729. <https://doi.org/10.1093/molbev/msq244>.
- Li L et al. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* **13**, 2178. <https://doi.org/10.1101/gr.1224503>.
- Lourenço J et al. (2013). The determinants of the molecular substitution process in turtles. *J Evol Biol* **26**, 38. <https://doi.org/doi:10.1111/jeb.12031>.
- Lynch M (2007). *The origins of genome architecture*. Sunderland: Sinauer Associates Inc.
- Martin A, Palumbi S (1993). Body size, metabolic rate, generation time, and the molecular clock. *Proc Natl Acad Sci USA* **90**, 4087. <https://doi.org/10.1073/pnas.90.9.4087>.
- McKenna M, Bell S (2007). *Classification of Mammals above the Species Level*. New-York: Columbia University Press.
- Mendes F et al. (2016). Gene tree discordance can generate patterns of diminishing convergence over time. *Mol Biol Evol* **33**, 3299. <https://doi.org/10.1093/molbev/msw197>.
- Meredith R et al. (2011). Impacts of the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification. *Science* **334**, 521. <https://doi.org/10.1126/science.1211028>.
- Montgomery S et al. (2013). The evolutionary history of cetacean brain and body size. *Evolution* **67**, 3339. <https://doi.org/10.1111/evo.12197>.
- Murphy W et al. (2007). Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Res* **17**, 413. <https://doi.org/10.1101/gr.5918807>.
- Nabholz B et al. (2008). Strong variations of mitochondrial mutation rate across mammals-the longevity hypothesis. *Mol Biol Evol* **25**, 120. <https://doi.org/10.1093/molbev/msm248>.
- Nery M et al. (2012). Resolution of the laurasiatherian phylogeny: evidence from genomic data. *Mol Phylogenet Evol* **64**, 685. <https://doi.org/10.1016/j.ympev.2012.04.012>.
- Nikolaev S et al. (2007). Life-history traits drive the evolutionary rates of mammalian coding and noncoding genomic elements. *Proc Natl Acad Sci USA* **104**, 20443. <https://doi.org/10.1073/pnas.0705658104>.

- Nishihara H et al. (2006). *Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions*. *Proc Natl Acad Sci USA* **103**, 9929. <https://doi.org/10.1073/pnas.0603797103>.
- Oakley T, Cunningham C (2000). *Independent contrasts succeed where ancestor reconstruction fails in a known bacteriophage phylogeny*. *Evolution* **54**, 397. <https://doi.org/10.1111/j.0014-3820.2000.tb00042.x>.
- Ohta T (1987). *Very slightly deleterious mutations and the molecular clock*. *J Mol Evol* **26**, 1. <https://doi.org/10.1007/BF02111276>.
- Orliac M, Ducrocq S (2011). *Eocene raoellids (Mammalia, Cetartiodactyla) outside the Indian Subcontinent: palaeogeographical implications*. *Geol Mag* **149**, 80. <https://doi.org/10.1017/S0016756811000586>.
- Popadin K et al. (2007). *Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals*. *Proc Natl Acad Sci USA* **104**, 13390. <https://doi.org/10.1073/pnas.0701256104>.
- Price S et al. (2005). *A complete phylogeny of the whales, dolphins and even-toed hoofed mammals (Cetartiodactyla)*. *Biol Rev* **80**, 445. <https://doi.org/10.1017/s1464793105006743>.
- Psouni E et al. (2012). *Impact of carnivory on human development and evolution revealed by a new unifying model of weaning in mammals*. *PLoS One* **7**, e32452. <https://doi.org/10.1371/journal.pone.0032452>.
- Ranwez V et al. (2011). *MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons*. *PLoS One* **6**, e22594. <https://doi.org/10.1093/molbev/msy159>.
- Rolland J et al. (2016). *Molecular evolutionary rates are not correlated with temperature and latitude in Squamata: an exception to the metabolic theory of ecology?* *BMC Evol Biol* **16**, 95. <https://doi.org/10.1186/s12862-016-0666-4>.
- Romiguier J, Figuet E, et al. (2012). *Fast and Robust Characterization of Time-Heterogeneous Sequence Evolutionary Processes Using Substitution Mapping*. *PLoS One* **7**, e33852. <https://doi.org/10.1371/journal.pone.0033852>.
- Romiguier J, Ranwez V, et al. (2010). *Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes*. *Genome Res* **20**, 1001. <https://doi.org/10.1101/gr.104372.109>.
- Romiguier J, Ranwez V, et al. (2013). *Genomic evidence for large, long-lived ancestors to placental mammals*. *Mol Biol Evol* **30**, 5. <https://doi.org/10.1093/molbev/mss211>.
- Ronquist F et al. (2016). *Closing the gap between rocks and clocks using total-evidence dating*. *Philos Trans R Soc London B* **371**, 20150136. <https://doi.org/10.1098/rstb.2015.0136>.
- Rose K (1982). *Skeleton of diacodexis, oldest known artiodactyl*. *Science* **216**, 621. <https://doi.org/10.1126/science.216.4546.621>.
- Scornavacca C, Galtier N (2017). *Incomplete lineage sorting in mammalian phylogenomics*. *Syst Biol* **66**, 112. <https://doi.org/10.1093/sysbio/syw082>.
- Springer M et al. (2017). *Waking the undead: Implications of a soft explosive model for the timing of placental mammal diversification*. *Mol Phylogenet Evol* **106**, 86. <https://doi.org/10.1016/j.ympev.2016.09.017>.
- Stamatakis A (2006). *RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models*. *Bioinformatics* **22**, 2688. <https://doi.org/10.1093/bioinformatics/btl1446>.
- Webster M et al. (2005). *Male-driven biased gene conversion governs the evolution of base composition in human alu repeats*. *Mol Biol Evol* **22**, 1468. <https://doi.org/10.1093/molbev/msi136>.
- White E et al. (2007). *Relationships between body size and abundance in ecology*. *Trends Ecol Evol* **22**, 323. <https://doi.org/10.1016/j.tree.2007.03.007>.
- Wible J et al. (2007). *Cretaceous eutherians and Laurasian origin for placental mammals near the K/T boundary*. *Nature* **447**, 1003. <https://doi.org/10.1038/nature05854>.
- Wible J et al. (2009). *The Eutherian Mammal Maelestes gobiensis from the Late Cretaceous of Mongolia and the phylogeny of cretaceous eutheria*. *Bull Am Museum Nat Hist* **327**, 1–123. <https://doi.org/10.1206/623.1>.

- Woolfit M, Bromham L (2005). *Population size and molecular evolution on islands*. *Proc Biol Sci* **272**, 2277. <https://doi.org/10.1098/rspb.2005.3217>.
- Wu H et al. (2014). *Camelid genomes reveal evolution and adaptation to desert environments*. *Nat Commun* **5**, 5188. <https://doi.org/10.1038/ncomms6188>.
- Wu J et al. (2017). *Rates of Molecular Evolution Suggest Natural History of Life History Traits and a Post-K-Pg Nocturnal Bottleneck of Placentals*. *Curr Biol* **27**, 3025–3033. <https://doi.org/10.1016/j.cub.2017.08.043>.
- Yang Z (2007). *PAML 4: Phylogenetic Analysis by Maximum Likelihood*. *Mol Biol Evol* **24**, 1586. <https://doi.org/10.1093/molbev/msm088>.
- Yim H et al. (2014). *Minke whale genome and aquatic adaptation in cetaceans*. *Nat Genet* **46**, 88. <https://doi.org/10.1038/ng.2835>.